

UFPE, Campus Recife (PE), 30/07 a 02/08/2018

Big Data: Desafios da Bioinformática



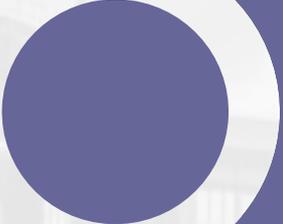
III Simpósio Norte e Nordeste de Bioinformática

ANAIS 2018

Organização



Recife-PE
2018



Anais 2018

ANAIS DO III SIMPÓSIO NORTE E NORDESTE DE BIOINFORMÁTICA

UFPE, RECIFE-PE | 30 de Julho a 02 de Agosto de 2018

Revisão Científica e Organização Técnica
Ana Maria Benko Iseppon

Editoração Eletrônica e Compilação
Alisson Amorim Siqueira

Todos os resumos neste livro foram reproduzidos de cópias fornecidas pelos autores e o conteúdo dos textos é de exclusiva responsabilidade dos mesmos. A Coordenação do evento não se responsabiliza por consequências decorrentes do uso de quaisquer dados, afirmações e/ou opiniões inexatas ou que conduzam a erros publicados neste livro de trabalhos.

Copyright © 2018 - III Simpósio Norte e Nordeste de Bioinformática

Todos os direitos reservados. Nenhuma parte desta obra pode ser reproduzida, arquivada ou transmitida, em qualquer forma ou por qualquer meio, sem permissão da organização do evento

Tema central

Big Data: Desafios da Bioinformática

Sessões Técnicas

Genes e genômica

RNA e transcriptômica

Proteínas e proteômica

Biologia de sistemas e integração de dados

Desenvolvimento de softwares e banco de dados

Apresentação

O Simpósio Norte e Nordeste de Bioinformática (SNNB) é um evento que reúne pesquisadores, professores e estudantes envolvidos com pesquisas na área de bioinformática, biologia da computação e biologia de sistemas, bem como suas interfaces no âmbito das ciências biológicas, da saúde e da agropecuária, entre outras.

O evento teve por objetivo integrar profissionais e estudantes das diversas áreas relacionadas à Bioinformática e Biologia de Sistemas, no âmbito das Ciências Biológicas, Ciências da Saúde, Ciências da Computação, Agropecuária e Veterinária (entre outros) promovendo a divulgação do conhecimento científico junto à comunidade acadêmica, científica ao setor de serviços e à população das regiões Norte e Nordeste, fomentando a integração dos grupos com interface nas áreas de ômicas, promovendo uma releitura e discussão dos temas concernentes à bioinformática e suas implicações.

A edição 2018 do SNNB aconteceu de 30 de Julho a 02 de Agosto de 2018 no Campus da UFPE, em Recife-PE e reuniu mais de 340 congressistas, 137 trabalhos, 9 minicursos e palestrantes nacionais e internacionais.

Organização e Apoio

Organização



Apoio Financeiro / Patrocínio



Organização

NOME	FUNÇÃO	INSTITUIÇÃO
Ana Maria Benko Iseppon	Presidente	UFPE
Ândrea Kely Campos Ribeiro dos Santos	Vice-Presidente (Região NE)	UFPA
Sandro José de Souza	Vice-Presidente (Região N)	BioME - UFRN
Reginaldo de Carvalho	Tesoureiro	UFRPE
Amaro de Castro Lira Neto	Comitê Gestor/Científico	IPA
Antonio Mauro Rezende	Comitê Gestor/Científico	IAM/FIOCRUZ
Flavia Figueira Aburjaile	Comitê Gestor/Científico	UFPE
Geyner Alves dos Santos Cruz	Comitê Gestor/Científico	UPE
João Pacifico Bezerra Neto	Comitê Gestor/Científico	UFPE
José Ribamar Costa Ferreira Neto	Comitê Gestor/Científico	UFPE
Roberta Lane de Oliveira Silva	Comitê Gestor/Científico	UFPE
Sérgio de Sá Leitão Paiva Junior	Comitê Gestor/Científico	UFRPE/UAST
Valdir de Queiroz Balbino	Comitê Gestor/Científico	UFPE
Valesca Pandolfi	Comitê Gestor/Científico	UFPE
Wilson José da Silva Júnior	Comitê Gestor/Científico	UFPE



Equipe de Apoio e Monitores

Artemisa Nazaré Costa Borges
Ayug Bezerra Lemos
Bruna Piereck Moura
Carlos André dos Santos Silva
Carolline de Jesús Pires
Elvson Wallacy da Silva
Gabriel Augusto Fabricio Bezerra
Gabriel Lucas Araújo Lima
José Bandeira do Nascimento Júnior

Lívia Maria Batista Vilela
Manuela Correia Dionísio
Marx Oliveira de Lima
Mireli de Santana Rêgo
Ricardo Severino Francelino
Rômulo da Fonsêca dos Santos
Valquíria da Silva
Vinicius Torres Guerra
Wilson Dias de Oliveira



Avaliadores *ad hoc*

Vasco Ariston de Carvalho Azevedo

Anne Cybelle Pinto

Tetsu Sakamoto

Rodrigo Bentes Kato

Armando de Menezes Neto

Marcus Vinicius Canário Viana

Valesca Pandolfi

Flávia Figueira Aburjaile

Antonio Mauro Rezende

Edson Luiz Folador

Eliseu Binneck

João Pacífico Bezerra Neto

Luciana Souto Mofatto

Marcelo Falsarella Carazzolle

Marcus Vinicius de Aragão Batista

Rodrigo Dias de Oliveira Carvalho

Sívio Torres Farias

Ândrea Ribeiro dos Santos

Gabriel da Luz Wallau

Gabriel da Rocha Fernandes

Guilherme Corrêa de Oliveira

João José de Simoni Gouveia

João Paulo Matos Santos Lima

Carolina V. Morgante

Veridiana Gomes Virginio

Wilson Junior

Fabício Martins Lopes

Jorge Estefano Santana de Souza

José Ribamar Costa Ferreira Neto

Rodrigo Juliani Siqueira Dalmolin

Sergio Sá Leião Paiva Jr.

Fabiana Cavalcante

Palestrantes

Dr. Richard Rötger

University of Southern Denmark (Dinamarca).

Dr. Alan Mitchell Durham

USP (São Paulo, SP).

Dra. Amanda Ferreira Vidal

UFPA (Belém, PA).

Dra. Ana Maria Benko-Iseppon

UFPE (Recife, PE).

Dra. Ândrea Kely C. Ribeiro dos Santos

UFPA (Belém, PA).

Dr. Antonio Mauro Rezende

Aggeu Magalhães – FIOCRUZ (Recife, PE).

Dr. Guilherme Corrêa de Oliveira

ITV-PA (Belém, PA).

Dr. João Paulo Matos Santos Lima

UFRN (Natal, RN).

Dr. Marcelo Falsarella Carazzolle

UNICAMP (Campinas, SP).

Dr. Rafael Melo Palhares

Myleus Biotecnologia (Belo Horizonte, MG)

Dr. Rodrigo Juliani Siqueira Dalmolin

UFRN (Natal, RN).

Dr. Sandro José de Souza

UFRN, BioME (Natal, RN).

Dr. Sávio Torres de Farias

UFPB (João Pessoa, PB).

Dr. Thiago Mafra Batista

UFMG (Belo Horizonte, MG).

M.Sc. Vitor Lima Coelho

UFRJ (Rio de Janeiro, RJ).

Marcel Pinheiro Caraciolo

Genomika Diagnósticos (Recife, PE)

Dr. William Farias Porto

Universidade Católica Dom Bosco (Brasília, DF).

Ministrantes de minicursos

João Pacífico Bezerra Neto

UFPE (Recife, PE)

Valdir de Queiroz Balbino

UFPE (Recife, PE)

George de Vasconcelos Carvalho Neto

Genomika Diagnósticos (Recife, PE)

Wilder Barbosa Galvão

Genomika Diagnósticos (Recife, PE)

Antonio Mauro Rezende

Aggeu Magalhães FIOCRUZ (Recife, PE)

Rodrigo Dias de Oliveira Carvalho

Aggeu Magalhães FIOCRUZ (Recife, PE)

Gabriel da Luz Wallau

Aggeu Magalhães FIOCRUZ (Recife, PE)

Valesca Pandolfi

UFPE (Recife, PE)

José Ribamar Costa Ferreira Neto (UFPE)

UFPE (Recife, PE)

Flávia Figueira Aburjaile

UFPE (Recife, PE)

Marx Oliveira de Lima

UFPE (Recife, PE)

Carlos André dos Santos Silva

UFPE (Recife, PE)

Bruna Piereck Moura

UFPE (Recife, PE)

Wilson J. da Silva Júnior

UFPE (Recife, PE)

Homenageado

Prof. Dr. Augusto Schrank

Por seu papel na indução de ações de ciência e tecnologia em bioinformática e biologia de sistemas.

Graduado em Farmácia e Bioquímica pela Faculdade de Farmácia, UFRGS (1977), Mestrado em Ciências (Microbiologia) pelo Instituto de Microbiologia Prof. Paulo de Góes, UFRJ (1981) e Doutorado em Biologia Molecular- University of Manchester Institute of Science and Technology (1988). É Professor Titular da Universidade Federal do Rio Grande do Sul no Departamento de Biologia Molecular e Biotecnologia. É orientador do programa de Pós-graduação em Biologia Celular e Molecular (PPGBCM) do Centro de Biotecnologia onde é líder de grupo. Foi Coordenador do PPGBCM e Chefe de Departamento. É Coordenador da área de Ciências Biológicas I da CAPES (2011-2016). Tem experiência na área de Genética Molecular de Microrganismos, com ênfase em Fungos, atuando principalmente com os fungos *Metarhizium* e *Cryptococcus* nos seguintes temas: determinantes de patogenicidade e mecanismos de infecção, quitinases, proteases e lipases. Tem experiência nas áreas de genômica, análise da expressão diferencial e em sistemas de transformação genética para a construção de mutantes funcionais em fungos. Tem experiência na formação de recursos humanos na graduação e pós-graduação



Agradecimentos

A ideia de realizar o III SNNB em Recife surgiu de um convite da Prof^a **Ândrea Ribeiros dos Santos (UFPA)** e do Prof. **Sandro de Souza (UFRN)** que foram os presidentes da segunda e da primeira edição do evento, respectivamente. A motivação do evento nos entusiasmou muito, pois deveria ser um evento com inscrições a preços módicos, destinado especialmente aos estudantes de graduação e pós-graduação. Além disso, o evento tinha a ambição de trazer a biologia computacional e a bioinformática a locais onde os eventos tradicionais não tem sido realizados por motivos financeiros e logísticos.

Ao participar do II SNNB, realizado em Belém em 2017, me senti ainda mais inspirada, pois vi muitos alunos entusiasmados com as palestras e cursos, ou ao explicar seus estudos em andamento. Além disso, o SNNB tem se destacado por dar os primeiros passos para consolidar a **Rede Nacional de Biologia Computacional (RENABIC)**, website <http://renabic.imd.ufrn.br/rede/>), integrando neste momento três redes apoiadas pelo Edital Biologia Computacional da **CAPES (Coordenação Aperfeiçoamento de Pessoal de Nível Superior)**, <http://www.capes.gov.br/>), a saber: (1) Rede InterSys de Interações Bióticas e Moleculares (coordenada pela UFPE); (2) Rede de Biologia Sistêmica do Câncer (Coordenada pela UFPA); (3) Rede de Pesquisas em Genômica Computacional Humana (Coordenada pela UFRN). Embora as redes tenham diferentes focos de estudo, o evento mostrou muitas convergências quanto às demandas e às abordagens computacionais, genéticas, ômicas, moleculares e metodológicas, apontando para uma necessidade de maior interação e sintonia não apenas entre esses três grupos pioneiros, mas entre todos os grupos participantes da RENABIC.

Tendo em vista o evento e suas motivações, não poderíamos deixar de mencionar o nosso Homenageado, o Prof. **Augusto Schrank**, não apenas pela sua ética, competência, seriedade e dedicação como pesquisador e como Coordenador da Área de Ciências Biológicas I na CAPES, onde o **Edital Biologia Computacional** teve sua gênese e implementação. O Prof. Augusto não atua como bioinformata ou como biólogo computacional, mas vislumbrou sua importância estratégica e se dedicou à sua indução no âmbito de sua gestão junto à CAPES. É de visionários e empreendedores desinteressados e isentos como o Prof. Augusto que o Brasil necessita. Nesse sentido, destacamos **nossa imensa gratidão à CAPES**, a seus gestores e técnicos, destacando a importância de que essa agência continue exercendo seu papel de formação de pessoal e indução de excelência científica e tecnológica. Um país sem ciência é um país escravizado e dependente, um país que não valoriza seu potencial e suas riquezas naturais. É em tempos de crise que necessitamos de governantes e gestores com visão estratégica.

Agradecimentos

Gostaria de agradecer a todos os **palestrantes, conferencistas e ministrantes de minicursos** que enriqueceram o evento compartilhando seu conhecimento e sua experiência com os participantes. Todos deixaram seus afazeres, interesses, sua família, sua rotina para nos agraciar com sua sabedoria e expertise.

Somos gratos também aos **avaliadores ad hoc** que nos ajudaram na avaliação dos **138 resumos**, propondo melhorias em quase um terço dos resumos submetidos, colaborando para a excelência dos anais. Nesse sentido, dirigimos um agradecimento especial aos coordenadores da comissão científica, Dr^a **Flávia F. Aburjaile**, Dr. **José Ribamar Ferreira Neto** e Dr^a **Valesca Pandolfi** do Depto. de Genética da UFPE.

Não poderíamos deixar de agradecer à comissão organizadora composta de membros de várias instituições pernambucanas, incluindo membros da **Universidade Federal de Pernambuco (UFPE)**, **Universidade de Pernambuco (UPE)**, **Instituto Agrônomo de Pernambuco (IPA)**, **Instituto de Pesquisas Aggeu Magalhães da Fundação Oswaldo Cruz (IPAM/FIOCRUZ)**, **Universidade Federal do Rio Grande do Norte (UFRN)**, **Universidade Federal do Pará (UFPA)** e da **Universidade Federal Rural de Pernambuco (UFRPE)**. Registramos aqui um agradecimento especial também à **Regional Nordeste da Sociedade Brasileira de Genética (SBG)** que nos apoiou com recursos angariados através de eventos regionais como palestras, colóquios, seminários e cursos. Uma mão lava a outra... muito obrigado!!

Agradecemos à **comissão organizadora** e aos **monitores do III SNNB**. Vocês se chamam “dedicação”!! Foram incansáveis e dedicados ao extremo. Desejamos agradecer a todos na figura de **João Pacífico Bezerra Neto**, **Carollina J. Pires** e **Artemisa J. C. Borges**. A equipe também agradece ao nosso web-designer **Alisson Amorim**, por sua atenção, presteza e especialmente pela criatividade e gosto apurado.

Também desejamos agradecer à **Associação Brasileira de Bioinformática e Biologia Computacional (AB3C)** e aos patrocinadores do evento, nominalmente às empresas **Genômika**, **Myleus**, **Merck**, **Eppendorf**, **Instituto Tecnológico Vale (ITV)** pelo apoio financeiro ou pelo financiamento de mobilidade e estadia de participantes.

Finalmente, mas não menos importante, agradecemos aos **congressistas** que se esforçaram para participar, inclusive se deslocando de locais distantes por via aérea e terrestre. Vocês são nossa motivação e nossa gratificação.

O próximo evento está planejado para ocorrer no segundo semestre de 2019 no estado do Pará. Contamos com todos vocês!!

Índice de áreas

Área	Páginas
Biologia de sistemas e integração de dados	9-26
Desenvolvimento de softwares e banco de dados	27-35
Genes e genômica	36-74
Proteínas e proteômica	75-105
RNA e transcriptômica	109-145



Biologia de sistemas e integração de dados

AN ASSOCIATION BETWEEN RNA-BINDING PROTEINS EXPRESSION AND DRUG RESPONSE IN TP53 GENOTYPES

Autores: Andre Luis Fonseca Faustino^{1,2}; Sandro jose de souza^{2,3}; Jean YJ Wang⁴;

E-mail para correspondência: andrefonseca@neuro.ufrn.br

Instituições: ¹Programa de Pos-graduação em Bioinformática, UFRN; ²Centro Multiusuário de Bioinformática (BioME), UFRN; ³Instituto do Cérebro, UFRN; ⁴University of California San Diego;

Palavras-chave: Cancer; Transcriptional regulation; Prognosis and treatment

Apoio: CAPES PPG-Bioinfo

RNA-binding proteins (RBPs) play essential roles in post-transcriptional regulation of mRNAs, such as splicing, polyadenylation, mRNA stabilization and translation. In the past few years, several studies have shown that deregulations in RBP-mediated mechanisms are associated with cancer initiation and progression. Regardless of their regulatory importance, RBP's biological background in cancer remains not completely understood. Actually, only a few studies cover systematics pan-cancer analyses and most of them do not address issues related to treatment and prognosis. To approach these issues, we carried out a large-scale analysis using public data from TCGA and GDSC (Genomics of Drug Sensitivity and Cancer). As guideline our workflow was divided into three main questions: i) How significantly are the RBPs associated with drug response? ii) Are the patient genotype associated with RBP regulation? iii) Finally, how the expression of RBPs affects the patient outcome? As result, we develop a framework to associate RBP expression, drug response and genotypes information. Firstly, we applied our strategy using cell line groups based on TP53 genotypes, due to the well-known association between TP53 mutations and drug response. Briefly, the cell lines were divided into TP53 wild-type and mutated groups. Next, each cell line had IC50 and expression data retrieved from GDSC database, used as input to calculate Pearson's correlations between gene expression and IC50. The process was applied for the whole genome using all drug types. In addition, we compared the correlations among all genes related to transcriptional regulation, such as TF, miRNA, and lncRNA. All the correlations were divided into distinct categories from lower to higher significance. As main findings, a large amount (201) of RBPs were shown to be highly associated with drug response in comparison with other genes, such as (15) TF. In addition, 49 RBPs are associated with three or more drugs, which suggest a preference among these RBPs to regulate similar biological process or pathways. Finally, a survival analysis was carried out using the TCGA data and an RBPs subset, where patient samples were grouped based on the individual RBPs expression level - lower, medium and high expression group. As result, were found dozen RBPs associated with prognosis status in several tumor types. In summary, we have found an association between RBPs expression, drug response and outcome for different cancer types.

Biologia de sistemas e integração de dados

ANÁLISE DO IMPACTO DA ESPARSIALIDADE EM REDES NEURAIS

Autores: André Luiz de Lucena Moreira¹; César Rennó Costa¹;

E-mail para correspondência: andrellm@ufrn.edu.br

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: Neurociência; Redes neurais; Tensorflow

Apoio: Instituto Metrópole Digital(IMD), Universidade Federal do Rio Grande do Norte(UFRN), BioME e CNPQ

Modelos computacionais de redes de neurônio são ferramentas importantes para investigar os mecanismos e características biológicas de diferentes processos cognitivos como sono e memória. Neste trabalho, vamos estudar o impacto da esparsialidade neural, a probabilidade de conexão entre neurônios, na capacidade de armazenamento de memórias em modelos de redes neurais artificiais. Para isto, utilizaremos o Tensorflow, uma biblioteca que permite a criação de redes neurais de forma simples, mas disponibilizando uma série de funções que permitem alterações e análises mais complexas dessas redes. Iremos implementar um modelo de rede neural já publicado na literatura e então analisar qual o impacto de variações da esparsialidade na capacidade da rede de armazenar memórias. Estudos semelhantes envolvendo apenas o aprendizado de máquina e desconsiderando características neurais, resultaram na melhora da capacidade de armazenamento. Dessa forma, esperava-se que a aplicação da esparsialidade nos modelos utilizados neste trabalho, resultaria na melhora do desempenho destes. Todavia, os resultados obtidos não corroboraram com os esperados, ocasionando na piora do aprendizado de máquina. Esse resultado pode ser justificado pela alta complexidade dos modelos utilizados. Estudos futuros devem aplicar a mesma metodologia a modelos menos robustos, podendo assim visualizar mais claramente as razões e implicações dos resultados obtidos. Este trabalho não contribui apenas para o estudo de sono e memória, mas também para o aprendizado de máquina, uma vez que algoritmos inspirados por mecanismos neurais podem melhorar o desempenho e armazenamento em técnicas de redes neurais artificiais já existentes.



Biologia de sistemas e integração de dados

ESTUDO SISTÊMICO DA EVOLUÇÃO DAS FLORES

Autores: Beatriz Moura Kfoury de Castro¹; Carlos Alberto Xavier Gonçalves¹; Lissur Azevedo Orsine¹; Tetsu Sakamoto¹; José Miguel Ortega¹;

E-mail para correspondência: beatrizkfoury@gmail.com

Instituições: ¹Universidade Federal de Minas Gerais;

Palavras-chave: Flores; Via biológica; Filogenia

Apoio: Capes Biologia Computacional rede BSC

As flores são inovações recentes em termos evolutivos, considerando o tempo geológico da história evolutiva das plantas. Elas são as estruturas reprodutivas das angiospermas (plantas com flores), da qual constituem o grupo das plantas mais diverso e cosmopolita. Para que a estrutura floral seja formada, é necessário que uma complexa rede regulatória de genes controle o seu desenvolvimento. Os objetivos deste trabalho foram (1) buscar na literatura o conhecimento sobre as bases moleculares do desenvolvimento floral, (2) descrever o desencadeamento dos processos bioquímicos e enzimáticos que ocorrem durante a formação da flor na forma de uma via metabólica e (3) realizar uma análise exaustiva sobre a origem evolutiva dos genes relacionados a floração. Para isso, foram coletados 1000 artigos científicos que descreviam a biologia molecular da floração e estes foram submetidos a ferramentas de mineração de texto, para ajudar na determinação dos genes e das biointerações descritos nos artigos. Para analisar a origem dos genes relacionados à floração, usamos a ferramenta de agrupamento de homólogos SeedServer e, em seguida, o algoritmo para a determinação do ancestral comum mais recente (LCA). Adicionalmente, realizamos uma curadoria manual da inferência do LCA, construindo árvores filogenéticas utilizando o TaxOnTree. Além disso, realizamos um estudo categórico sobre a história evolutiva das plantas terrestres utilizando abordagens filogenômicas. Através da análise de mineração de texto, 94 genes de *Arabidopsis thaliana* foram relacionadas com a floração e 67 delas foram organizadas em uma via metabólica. As inferências das origens dos genes demonstraram que os genes da via do desenvolvimento floral apareceram em diferentes clados durante a história evolutiva das plantas, mas tendo a maior parte dos seus genes surgido ou expandido junto com origem das plantas com flores ou no momento de sua diversificação no grupo de plantas com flores não basais. Nossos estudos comparativos de dados genômicos e proteômicos demonstraram a ocorrência de uma rápida divergência na linhagem das plantas com flores.



Biologia de sistemas e integração de dados

MODELAGEM DA DINÂMICA TUMORAL ANALISANDO A MOBILIDADE CÉLULAR EM DISTINTOS NÍVEIS DE HIPOXIA.

Autores: Dhiego Souto Andrade¹; César Renno Costa¹;

E-mail para correspondência: dsouto49@gmail.com

Instituições: ¹UFRN;

Palavras-chave: cancer modeling; tumor dynamics; cell motility

Apoio: Capes, Cnpq, BioME

We report the initial development of a 3D agent-based cancer model to support the simulation of cell behavior in different conditions of oxygen availability. We use agent-based modeling that explicitly models the interaction of an individual with its environment and neighbors and can give us insights about cellular behavior and its phenotype during tumor growth. Using 3D modeling, we can observe the spatial conformation of the cells. In our model, each cell state has a position, a volume, a cell cycle or death status, and mechanics parameters such as adhesion, deformation, and motility. Each cell interacts with the biochemical microenvironment through a vector of chemical substrates ruled by a system of reaction-diffusion Partial Differential Equations (PDEs). Our model begins with a single cancer cell with a high probability of duplication according to the number of existing neighbors, that is, the more existing neighbor cells, the bigger the probability of duplication, as long as the environmental conditions are right (in this model, satisfactory oxygen concentration). Initially, each cancer cell has no motility. This feature is only acquired when a defined concentration of oxygen is detected, and the cell needs to search for a favorable region of duplication, in other words, with high oxygen concentration. MmHg defines the oxygen value. We simulated three values for cell motility phenotype to be acquired: 15mmHg, a value in which hypoxic signaling starts; 8 mmHg, a trigger value for hypoxic responses; and 4 mmHg, a value in which hypoxic responses are at maximum. Each simulation performed was defined to last one day of wall time (time in real life), resulting thirty-four days of simulation time. At the end of the simulation, the number of cancer cells surpasses thirty thousand cells. We observed an exponential cellular growth in every simulation. However, the tumor growth becomes faster when cell motile phenotype appears before high rates of hypoxic responses appear. Necrosis rose after tumor acquires about twenty-five thousand cells. The tumor shows a spatially oval form during its dynamics. We used the modeling software "Physicell", an open source physics-based cell simulator for 3-D multicellular systems. The simulations were performed using NPAD-IMD-UFRN computer cluster. As future work, we aim to add metabolic pathways on the model to analyze cellular decision with different pathways topologies.



Biologia de sistemas e integração de dados

EVOLUÇÃO DA REDE DE GENES DE NEUROTRANSMISSÃO EM HUMANOS

Autores: Lucas Henriques Viscardi¹; Danilo Oliveira Imparato²; Maria Cátira Bortolini¹; Rodrigo Juliani Siqueira Dalmolin²;

E-mail para correspondência: imparatodanilo@gmail.com

Instituições: ¹Departamento de Genética, UFRGS, Porto Alegre, RS; ²Bioinformatics Multidisciplinary Environment (BioME), UFRN, Natal, RN;

Palavras-chave: biologia de sistemas; ppi; neurotransmissão

Apoio: CNPq, CAPES, PROPESQ

Entende-se que o último ancestral comum de todas as sinapses teve sua origem pouco antes dos cnidários e que o surgimento dos sistemas nervosos o seguiu pouco depois. Animais desprovidos de sistema nervoso possuem componentes sinápticos e mecanismos de transmissão de sinais. Estudos também indicam que genes associados a sinalização sináptica participam de respostas ambientais nos eucariotos unicelulares. A presença desses correlatos de neurotransmissão em diversos taxons eucarióticos nos levam a questionar como a rede de genes responsáveis pelas conexões nervosas se estabeleceu no processo evolutivo. Neste trabalho, buscamos a origem de genes que orquestraram a evolução das sinapses humanas e suas interações: através do KEGG Pathway Database, coletamos identificadores de genes relacionados às vias de neurotransmissão sináptica disponíveis (GABAérgica, glutamatérgica, serotoninérgica, dopaminérgica e colinérgica). De forma a inferir a origem evolutiva dos genes e suas respectivas proteínas, obteve-se a partir da base STRING uma árvore de espécies que foi posteriormente manualmente curada, bem como anotações de ortologia das proteínas. O posicionamento filogenético dos grupos de ortólogos e seus respectivos genes foi obtido por meio do pacote em R GenePlast. Dados de expressão (Expression Atlas) foram utilizados para corroborar a perspectiva sistêmica da análise. Importantes observações puderam ser feitas a respeito do surgimento dos genes e da topologia da rede PPI ao longo da árvore evolutiva. Apesar de muitos genes do conjunto estarem presentes no ancestral comum dos eucariotos, observou-se que a maior parte da rede foi progressivamente estabelecida até a divergência entre Cnidaria e Bilateria. Curiosamente não se observou aparecimento de genes relacionados a neurotransmissão sináptica desde a divergência entre humanos e peixes. Por fim, as análises demonstram a importância da perspectiva evolutiva na biologia sistêmica e no entendimento do sistema nervoso e sua evolução.



Biologia de sistemas e integração de dados

PACOTE EM R PARA ANÁLISE TRANSCRICIONAL COM TRANSCRIPTOGRAMAS

Autores: Diego Arthur de Azevedo Morais¹; Rodrigo Juliani Siqueira Dalmolin^{1,2};

E-mail para correspondência: arthur.vinx@gmail.com

Instituições: ¹Bioinformatics Multidisciplinary Environment, Instituto Metr pole Digital, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil; ²Departamento de Bioqu mica, Centro de Bioci ncias, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil;

Palavras-chave: An lise transcricional; Biologia de sistemas; Software

Apoio: Este trabalho foi apoiado pelo CNPq [444856/2014-5] e CAPES [3381/2013].

O transcriptograma, um m todo de an lise de transcriptomas, utiliza dados de intera o prote na-prote na (PPI) para construir uma lista de genes. Esta lista   constru da de forma a posicionar os genes de acordo com a probabilidade de intera o entre seus produtos, assim, a probabilidade de associa o funcional entre dois genes decai exponencialmente com o aumento da dist ncia. O pacote transcriptogramer, dispon vel no Bioconductor.org, utiliza transcriptogramas para realizar an lises topol gicas, identificar grupos g nicos cuja express o encontra-se alterada entre duas condi es biol gicas, e enriquecer ontologias do Gene Ontology. Os datasets do pacote incluem listas ordenadas de genes para 4 esp cies (Homo sapiens, Mus musculus, Saccharomyces cerevisiae e Rattus norvegicus), constru das com base em PPI obtidas do STRINGdb release 10.5. O pacote projeta dados de express o, provenientes de Microarray ou RNA-Seq, na lista ordenada de genes em 2 etapas. A primeira etapa consiste em atribuir os valores de express o a cada gene da lista, sendo utilizada a m dia da express o de todos os identificadores, sondas no caso de um dado de Microarray, referentes a um mesmo gene. A segunda etapa consiste em utilizar uma janela deslizante de raio definido pelo usu rio para obter a m dia da express o de um grupo de genes associados funcionalmente, atribuindo o valor obtido ao gene presente no centro da janela. Condi es peri dicas s o levadas em considera o em janelas envolvendo genes pr ximos dos limites da lista de genes. Esta estrat gia reduz o ru do presente nos dados e torna poss vel a express o diferencial de grupos g nicos definidos de acordo com o raio escolhido. O enriquecimento   realizado de acordo com uma das 3 ontologias do Gene Ontology, considerando, ou n o, a hierarquia dos termos. O pacote transcriptogramer   open source, multiplataforma (podendo ser executado em Linux, Windows e Mac OS X) e compat vel com processadores multicore, tornando poss vel a paraleliza o de algumas fun es apenas informando o n mero de threads a ser utilizada, reduzindo assim o tempo computacional necess rio. O pacote transcriptogramer produz como sa da data.frames e figuras que auxiliam a interpreta o biol gica de dados transcricionais.

Biologia de sistemas e integração de dados

CONSTRUÇÃO DA VIA DE INTERAÇÃO DO VÍRUS EBOLA COM O HOSPEDEIRO E ESTIMATIVA DA ORIGEM DOS GENES E PROCESSOS

Autores: ELISSON NOGUEIRA LOPES¹; CARLOS XAVIER GONCALVES¹; LISSUR AZEVEDO ORSINE¹; IARA D SOUZA²; TETSU SAKAMOTO¹; RODRIGO JULIANI SIQUEIRA DALMOLIN²; JOSE MIGUEL ORTEGA¹;

E-mail para correspondência: elisson.nlopes@gmail.com

Instituições: ¹UFMG; ²UFRN;

Palavras-chave: ebola; evolução; vias metabólicas

Apoio: Capes biologia computacional rede BSC

O vírus Ebola é o causador de uma doença devastadora, com taxa de mortalidade de cerca de 50-90%. Os sintomas desenvolvidos por pacientes infectados são febre, mal-estar e dor muscular, podendo ser seguidos por sangramento e falha de órgãos e tecidos. As primeiras células infectadas pelo vírus são macrófagos e células dendríticas, mas seu tropismo abrange um grande universo celular. O vírus faz sua adsorção através da membrana plasmática, utilizando sua glicoproteína. O processo de penetração é desencadeado em consequência, por mimetismo do mecanismo de endocitose, para tanto é necessário a ação das proteínas da membrana celular dos alvos virais. Durante a penetração, o vírus é transportado no endossomo, que resultam na liberação das partículas virais no citoplasma do hospedeiro, onde este pode realizar os processos restantes que culminam na liberação de novas partículas virais. Para elucidar a via de infecção e todos os processos desencadeados desde o primeiro contato viral-celular, foi feito o estudo do mecanismo de infecção, analisando todos os dados relativos conhecidos até o momento, através de ferramentas de mineração de texto. Esta abordagem de mineração recrutou 133 proteínas hospedeiras, de um total de mais de 1000 artigos científicos. Através dessa metodologia, foi construído uma via completa do interatoma para representar a infecção por Ebola. Para uma abordagem mais ampla, foi feita a análise também dos homólogos de cada proteína coletada, na etapa anterior, para inferir seu clado/época de origem. Os resultados da análise de origem evolutiva permitem inferir que o vírus poderia infectar desde *Euleostomi*, sugerindo que animais como peixes e pets poderiam ser hospedeiros e retransmitir o vírus para outros hospedeiros, como o homem. Além disso, analisamos duas séries de dados GEO para expressão gênica após a infecção pelo Ebola, criando um perfil de infecção viral experimental. Através da análise de enriquecimento dos genes encontrados com vias do Kegg, foram encontrados processos envolvendo desde controle celular a outros processos de imortalização e supressão do sistema imune. Em conclusão, a infecção pelo Ebola ocorre desde proteínas antigas, com uma datação que traz em pauta hospedeiros ainda não descritos. A construção da via possibilita a criação de um estudo sistemático a infecção do vírus e quais seriam os possíveis alvos terapêuticos.



Biologia de sistemas e integração de dados

DESENVOLVIMENTO DE UMA FERRAMENTA PARA AVALIAÇÃO DE VARIAÇÕES CONFORMACIONAIS E MODELOS TEÓRICOS DE PROTEÍNAS A PARTIR DE REDES DE INTERAÇÃO DE RESÍDUOS

Autores: Felipe Vieira da Fonseca¹; Eden Silva e Souza¹; João Paulo Matos Santos Lima¹;

E-mail para correspondência: ffvieira@outlook.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: Desenvolvimento de software; RINs; Modelagem por homologia

Apoio: Bioinformatics Multidisciplinary Environment - BIOME / UFRN

Alterações na sequência primária de aminoácidos podem resultar em alterações na estrutura tridimensional de proteínas e perda parcial ou total da sua função. Um dos métodos utilizados para avaliar alterações estruturais é a modelagem por homologia e posterior comparação das estruturas geradas. Entretanto, a modelagem não permite a comparação de grandes volumes de dados. Uma forma de representar as ligações e interações entre todos os aminoácidos de uma proteína é por meio das redes de interação de resíduos (RINs). As RINs são uma forma de apresentar a estrutura 3D de proteínas na forma de grafos, onde os nós representam os resíduos e as arestas representam as interações físico-químicas entre os aminoácidos. Acredita-se que as RINs apresentam um grande potencial para avaliação das diferenças conformacionais entre proteínas, o que pode representar uma importante ferramenta na avaliação de modelos teóricos. Portanto, o objetivo do presente trabalho foi construir uma ferramenta para comparação de diferentes RINs para uma proteína e utilizar tais dados para pontuar diferenças conformacionais entre proteínas e na validação de modelos gerados por homologia. As RINs foram criadas utilizando o RING2.0 (*Residue Interaction Network Generator*). A ferramenta desenvolvida compara os nós dos modelos teóricos gerados, levando em consideração a posição, a cadeia e o resíduo, bem como suas interações com os outros aminoácidos. Realizou-se a modelagem do peptídeo ALX72355.1 utilizando o programa Modeller e como molde a proteína com o código pdb 6ATW (modelo original). Em seguida, realizou-se manualmente substituições de quatro resíduos no modelo criado para outros resíduos com propriedades químicas semelhantes (modelo mutante), e todos os pdbs gerados foram comparados entre si. A ferramenta desenvolvida também foi utilizada na avaliação de pdbs com mesma sequência primária e diferentes estruturas conformacionais. A RIN do modelo original foi comparada com a RIN do modelo mutante utilizando a ferramenta apresentada. A ferramenta obteve êxito na identificação de todas as diferenças previstas em nossa base de dados, tanto para os resíduos como para suas interações, comprovando sua eficácia em descrever as alterações conformacionais e diferenças entre os modelos gerados. Futuramente, serão empregadas novas abordagens para a otimização da ferramenta e a implementação de uma interface web para posteriormente ser disponibilizada ao meio acadêmico.

Biologia de sistemas e integração de dados

HUMAN DIGESTIVE SYSTEM AS A MODEL ON THE STUDY OF THE EVOLUTION OF HETEROTROPHY

Autores: Fenícia Brito Santos¹; Tetsu Sakamoto¹; José Miguel Ortega¹;

E-mail para correspondência: fenisanto@gmail.com

Instituições: ¹Universidade Federal de Minas Gerais;

Palavras-chave: heterotrophy; digestion; evolution

Apoio: CAPES BIOLOGIA BOMPUTACIONAL - BSC NETWORK

The key processes for the maintenance of life require energy and for all heterotrophic organisms, this energy comes from organic sources of carbon available in the environment. Since the origin of the first forms of life, the processes and systems involved in obtaining and metabolizing food have been determinant in the evolutionary success of the species. The emergence of multicellularity and compartmentalized systems allowed metazoans to accumulate energy and reallocate it for the development of new functions. Although studies of comparative anatomy in metazoa are well documented and the systemic knowledge of several processes is reported in many databases, studies addressing the comparative genomics and the evolution of the components of this system are scarce. Our goal was to study the sequential origin of the genes involved in heterotrophy using the human digestive system as model. For this, we draw diagrams for the digestive system secretion pathways based on models available in KEGG Pathway database. The origin of each component was estimated using tools for homologous clustering and for lowest common ancestor inference. This allowed us to infer the origin of the system based on the origin of its genes. We showed that the most ancestral genes found in the pathways act on cell signaling processes and arose before the systems have been originated. The most recent components, such as receptors and some transporters, which are essential for secretion function, appeared from Metazoa. In addition, some components with auxiliary function, such as bicarbonate secretion in the pancreas and bile, have a more recent origin, indicating that this process appears as a refinement in these secretory pathways. Salivary secretion has the highest number of recent components and many of the proteins secreted are exclusive in mammals. In addition, we performed an analysis using the ELDOgraph program to identify which organisms or OTUs have more proteins close to human. The results show that the species that has more ELDO with human belong to the genus Pan. Analyzing at the order level the species with more ELDO with human belong to the orders Dermoptera and Rodentia. Our results show that the secretion pathways of the digestive system in mammals share many similarities, although some proteins are more distant. The data presented here allowed us to draw a scenario about the evolution of the digestive process, contributing to the evolutionary history of this system.

Biologia de sistemas e integração de dados

DESENVOLVIMENTO DE FRAMEWORK PARA CRIAÇÃO DE MODELOS COMPUTACIONAIS DE CÉLULA COMPLETA

Autores: Frederico Chaves Carvalho¹; Paulo Eduardo Ambrósio¹;

E-mail para correspondência: fcc073@gmail.com

Instituições: ¹Universidade Estadual de Santa Cruz;

Palavras-chave: Modelo de célula completa; Framework; Integração de dados

Apoio: Fundação de Amparo à Pesquisa do Estado a Bahia (FAPESB)

Propostas de criação modelos computacionais de células completas são recentes e buscam representar todos os processos bioquímicos intracelulares de maneira a fornecer uma maneira rápida e eficiente de obter resultados simulados confiáveis e equivalentes aos obtidos com métodos *in vitro* ou *in vivo*. O desenvolvimento desses modelos é importante tanto para consolidar o conhecimento atual da biologia, como para fornecer subsídios para avanços científicos mais rápidos em áreas como a medicina e a bioengenharia.

Os primeiros modelos construídos foram da bactéria *M. genitalium* devido à simplicidade de seu genoma, que possui apenas 580 kb, sendo o organismo de menor genoma conhecido. A complexidade de tais modelos traz consigo alguns desafios, como a necessidade de integrar dados heterogêneos, a alta demanda computacional das simulações e a não escalabilidade dos atuais modelos.

Este trabalho tem como proposta a criação de um framework que represente a formalização metodológica para a criação de modelos de célula completa, buscando integrar técnicas utilizadas para o desenvolvimento de modelos anteriores com ferramentas que auxiliarão na obtenção de modelos escaláveis de maneira simplificada. Tais ferramentas têm entre suas funções a coleta de informações diretamente de bancos de dados, paralelização do processamento do modelo e utilização de placas gráficas para acelerar o processamento. O framework será desenvolvido nas linguagens C e Python, utilizando também recursos da plataforma CUDA.

O framework BRCELL, em desenvolvimento, permitirá ao modelador registrar sistematicamente suas hipóteses e considerações durante a construção do modelo, optar entre utilizar um conjunto de regras e/ou comandos de programação, simular o modelo utilizando processamento paralelo e/ou GPU como coprocessador e visualizar os resultados em forma de gráficos e tabelas. A bactéria *M. genitalium* será utilizada como base para o desenvolvimento do framework, mas outras células também poderão ser modeladas.

No presente momento, o framework já possui uma interface gráfica que reflete as necessidades específicas da criação de modelos de célula completa, organizada de maneira a guiar o usuário através de todo o processo, desde a coleta de dados até a leitura dos resultados. As primeiras funcionalidades estão sendo implementadas e testadas, demonstrando que o BRCELL proporciona um significativo ganho, em tempo computacional e em confiabilidade do modelo, na modelagem da bactéria.



Biologia de sistemas e integração de dados

INFERÊNCIA DA RAIZ EVOLUTIVA DE GRUPOS DE ORTÓLOGOS RELACIONADOS A GENES ESSENCIAIS EM EUKARIOTOS

Autores: Iara Dantas de Souza¹; Clovis Ferreira dos Reis¹; Diego Arthur de Azevedo Moraes¹; Rodrigo Juliani Siqueira Dalmolin^{1,2};

E-mail para correspondência: iaradsouza1@gmail.com

Instituições: ¹Bioinformatics Multidisciplinary Environment (BioME) - Instituto MetrÓpole Digital, Universidade Federal do Rio Grande do Norte; ²Departamento de Bioquímica - Centro de Biociências, Universidade Federal do Rio Grande do Norte;

Palavras-chave: Genes essenciais; Genes letais; Evolução

Apoio: CAPES, CNPQ, PROPESQ-UFRN, PPG-BIOINFO/UFRN

Genes essenciais são assim chamados por terem papéis importantes para a manutenção celular. Deste modo, mutações deletérias nesses genes resultam em inviabilidade da célula e, em maior nível, do organismo. Em organismos multicelulares, alterações deletérias em genes essenciais produzem um espectro de fenótipos, desde o comprometimento do processo de fertilização, a interrupção do desenvolvimento fetal, até a perda da capacidade reprodutiva em indivíduos adultos. Nós supomos que, considerando a importância de processos regulados por genes essenciais, estes genes surgiram anteriormente durante a história evolutiva e permaneceram conservados em muitos organismos. Este trabalho tem o objetivo de identificar os genes essenciais em organismos eucarióticos e estimar seu surgimento durante o processo evolutivo. Foram buscadas informações fenotípicas provenientes de genótipos mutantes em bancos de dados especializados em organismos modelo. Com base na capacidade do genótipo anotado ser letal, classificamos os genes em essenciais e não-essenciais. Considerando o modelo de camundongo, categorizamos os genes essenciais do camundongo em três grupos: letalidade precoce, letalidade intermediária e letalidade tardia, de acordo com o estágio de desenvolvimento em que ocorreu a letalidade. Nós identificamos os grupos de ortólogos pertencentes a genes essenciais e não-essenciais a partir do banco de dados STRING (v10.5). Usando o pacote Geneplast (R/Bioconductor), inferimos a raiz evolutiva dos genes essenciais e não-essenciais. Os genes essenciais, em média, possuem um índice de ancestralidade maior que os não-essenciais em todos os organismos considerados. Da mesma forma, a distribuição de raízes de categorias de genes essenciais em camundongos mostra uma prevalência de genes de letalidade precoce em raízes primitivas. Portanto, concluímos que os genes essenciais emergiram mais cedo do que genes não-essenciais na história evolutiva.



Biologia de sistemas e integração de dados

ANÁLISE COMPUTACIONAL PARA PREDIÇÃO DE VARIANTES NÃO-SINÔNIMAS NA REGIÃO CITOPLASMÁTICA DO GENE TOLL- LIKE RECEPTOR 4

Autores: Isabella Luiza Ralph de Oliveira ¹; Maria Amélia Carlos Souto Maior Borba ¹; Carlos Henrique Madeiros Castelletti ¹; Danyelly Bruneska Gondim Martins ³;

E-mail para correspondência: bellaralph01@hotmail.com

Instituições: ¹Laboratório de Imunopatologia Keizo Asami-LIKA/UFPE; ²Departamento de Bioquímica-UFPE ; ³Instituto Agrônomico de Pernambuco-IPA;

Palavras-chave: algoritmos; domínio TIR ; variantes não-sinônimas

O receptor TLR4 modula a inflamação e também participa da resposta imunológica contra infecções por vírus. O domínio TIR é um domínio citoplasmático da TLR, essencial para interações com moléculas adaptadoras e responsável pela cascata de sinalização. Variantes não-sinônimas no gene TLR4 estão relacionados a maior susceptibilidade para desenvolvimento de cânceres, inclusive o cervical, causado pelo Papilomavírus humano (HPV). Assim, o trabalho tem como objetivo utilizar ferramentas computacionais de predição para determinar as variantes não-sinônimas no domínio citoplasmático da TLR4 que podem causar alterações na integridade gênica e predispor a infecção/lesão cervical pelo HPV. As variantes foram coletadas nas bases dbSNP e Ensembl. Quinze algoritmos de predição foram utilizados para determinar as variantes deletórias, dos quais treze pertencentes ao banco de dados do dbNSFP (local), e dois são preditores web-based (SNP&GO; SNAP2). O algoritmo Maestro, foi utilizado para predição estrutural. A sequência da TLR4 foi obtida do Uniprot para análise de modelagem das estruturas contendo as mutações pelo SwissModel. O Molprobit foi utilizado para avaliar a qualidade das estruturas geradas e corrigir os rotâmeros. Das 1.323 variantes catalogadas na TLR4, 308 são não-sinônimas, dos quais apenas 180 foram encontradas em ambos, dbSNP e Ensembl. Duas variantes não-sinônimas apresentaram predição deletória em 10 dos 15 algoritmos testados: NP_612564.1:p.Gly715Ser e NP_612564.1:p.Thr793Ile. O resíduo Gly715 é localizado no BB-loop na interface do domínio TIR. A região que compreende do resíduo 714 ao 716 é considerada importante na formação de uma superfície de interações entre proteínas. Desta forma, a variante Ser715 pode levar a desregulação da cascata de sinalização da TLR4. Não há estudos ainda que relacionem variantes não-sinônimas no resíduo Thr793 do TLR4. Gly715Ser e Thr793Ile, se acumuladas, apresentam menor estabilização na proteína. O estudo possibilitou identificar duas variantes não-sinônimas capazes de causar dano a proteína. Estudos adicionais irão determinar se estas variantes podem auxiliar o rastreamento para predisposição da lesão/infecção pelo HPV além de um possível alvo terapêutico.



Biologia de sistemas e integração de dados

PREDIÇÃO COMPUTACIONAL DA FARMACOCINÉTICA E FARMACODINÂMICA DO BENZOTIAZOL COM POTENCIAL ANTITUMORAL EM LINHAGEM DE ADENOCARCINOMA GÁSTRICO

Autores: Luina Benevides Lima¹; Felipe Pantoja Mesquita¹; Julio Paulino Daniel¹; Adrhyann Jullyanne de Sousa Portilho¹; Lais Lacerda Brasil¹;

E-mail para correspondência: luina@ufc.br

Instituições: ¹Universidade Federal do Ceará; ²Universidade Federal Fluminense;

Palavras-chave: FARMACOCINÉTICA; FARMACODINÂMICA; BENZOTIAZOL

Apoio: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

O adenocarcinoma gástrico é uma das malignidades originadas no estômago com considerável incidência e mortalidade no mundo. Com base na necessidade de novos fármacos como alternativas ao tratamento do adenocarcinoma gástrico, esse trabalho teve por objetivo testar a atividade antitumoral do benzotiazol (E)-2-((2-(benzo[d]tiazol-2-ila)hidrazono)metil)-4-nitrofenol na linhagem de adenocarcinoma gástrico ACP03, realizar a predição computacional da farmacocinética e farmacodinâmica dessa molécula, bem como o docking molecular com o alvo farmacológico relevante encontrado na predição. Para a avaliação do potencial citotóxico do composto foi realizado o ensaio do MTT, enquanto que a capacidade de indução de morte celular foi feita analisando as células morfológicamente por microscopia de fluorescência utilizando os fluoróforos diacetato de fluoresceína, hoechst 33342 e iodeto de propídeo. Para a predição das propriedades farmacocinéticas e farmacodinâmica foram utilizados os servidores online ADMETSar e SwissTargetPrediction, respectivamente. O docking molecular foi realizado utilizando o programa Autodock® Vina. O resultado da curva concentração-resposta obtido no ensaio do MTT revelou um efeito citotóxico relevante contra a linhagem ACP03 com valores de CI₅₀ [CI_{95%}] em torno de 17.24 µM [14.92 – 19.92] para 24h, 8.23 µM [7.02 – 9.65] para 48h e 1.39 µM [1.14 – 1.69] para 72h. A análise por microscopia de fluorescência mostrou que o composto AFN01 foi capaz de induzir de forma significativa apoptose nas concentrações de 1 e 2 µM (P<0.05), assim como o controle positivo doxorrubicina. As análises computacionais identificaram propriedades farmacocinéticas relevantes como capacidade de absorção intestinal positiva, inibição do CYP1A2, CYP2C9 e CYP3A4, bem como inibição fraca de hERG e toxicidade oral aguda classe III (LD₅₀ entre 500 e 5000 mg/kg). A predição farmacodinâmica permitiu a triagem de alvos moleculares de importância para a atividade antitumoral. Portanto, a enzima TDP1, uma enzima importante no reparo do DNA, foi selecionada para o docking molecular indicando que o benzotiazol possui uma energia de ligação de -7.5 kcal/mol e importantes interações com resíduos do sítio catalítico da enzima. Em conclusão, o composto benzotiazólico AFN01 é um excelente candidato para o tratamento do adenocarcinoma gástrico, com parâmetros farmacocinéticos preditos aceitáveis e possível alvo farmacológico relevante do ponto de vista molecular.



Biologia de sistemas e integração de dados

USANDO RINS PARA ENTENDER AS MUTAÇÕES EM CÂNCER: MUTAÇÕES DELETÉRIAS SÃO MAIS COMUMENTE ASSOCIADAS A AMINOÁCIDOS ALTAMENTE CONECTADOS

Autores: Laise Cavalcanti Florentino¹; Diego Arthur de A. Morais¹; Diego Gomes Teixeira¹; Jorge Estefano S. de Souza¹; Rodrigo Juliani S. Dalmolin¹;

E-mail para correspondência: l.unifolie@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: efeito das mutações; redes de interação de resíduos; análise de dados

Apoio: Biologia Sistêmica do Câncer (BSC); BioME UFRN; Npad - UFRN.

Muitos esforços para identificar mutações driver em câncer foram feitos, a maioria no nível de sequência. A compreensão do impacto estrutural de uma determinada mudança de aminoácido é de grande importância para a pesquisa médica do câncer. Aqui, utilizamos parâmetros de redes de interação de resíduos (RINs) para uma análise em larga escala de mutações missense em 16 tipos de câncer, permitindo inferir seus respectivos efeitos estruturais e verificar se mudanças em aminoácidos altamente conectados têm maior probabilidade de originar mutações driver. Usamos as RINs para analisar quais parâmetros de rede são mais comuns em resíduos (nó) com ocorrência relatada de mutações missense em câncer. A distribuição da quantidade de mutações por degree (conectividade do nó) varia significativamente comparada às simulações aleatórias e também à distribuição de um conjunto de dados de polimorfismos de nucleotídeo único humano (SNPs), tendendo a se manter em nós com conectividade mais baixa. Além disso, a proporção de mutações driver foi significativamente aumentada em nós com alto grau de conectividade, quando utilizados dois critérios diferentes para sua classificação: proporções de preditores de software (NDamage) e classificação de base de dados ClinVar (clinproc). Portanto, levando em conta esses resultados, podemos concluir que as alterações nos aminoácidos altamente conectados têm maior probabilidade de gerar mutações driver, uma vez que sua maior proporção de ocorrência nesses nós e a análise de RINs podem ser usadas como um parâmetro adicional para auxiliar a previsão de mutações driver em câncer.

Biologia de sistemas e integração de dados

ELABORAÇÃO DE VIA DE DESENVOLVIMENTO DA GLÂNDULA MAMÁRIA E ESTIMATIVA DA ORIGEM DOS SEUS GENES E DO SISTEMA

Autores: Lissur Azevedo Orsine¹; Elisa Rennó Donnard Moreira²; José Miguel Ortega¹;

E-mail para correspondência: lissurorsine@gmail.com

Instituições: ¹Universidade Federal de Minas Gerais; ²University of Massachusetts Medical School;

Palavras-chave: Via biológica; glândula mamária; origem evolutiva

Apoio: CAPES BIOLOGIA COMPUTACIONAL - REDE BSC

O conhecimento dos mecanismos moleculares que governam dado fenômeno biológico consiste em etapa importante para sua compreensão. Neste contexto, vias biológicas são de grande valia por apresentarem visão sistêmica dos fenômenos biológicos. A glândula mamária está ligada ao sucesso evolutivo dos mamíferos, bem como é alvo do tipo mais frequente de câncer em mulheres. Apesar disso, não existia uma via de desenvolvimento da glândula mamária. Outra demanda se referia à necessidade de elucidar sua origem evolutiva.

O objetivo deste trabalho era elaborar uma via de desenvolvimento da glândula mamária e estimar a origem dos seus genes e do sistema.

Para isso, o primeiro passo consistiu em realizar um levantamento dos genes (e das interações entre eles) envolvidos no desenvolvimento da mama através de ferramentas de mineração de texto (Medline Ranker e PESCADOR), e elaborar um desenho e uma descrição para a via. Em seguida, a origem evolutiva de cada gene foi estimada através da determinação do ancestral comum mais recente dos organismos contendo cópias do gene usando ferramentas para o agrupamento de genes homólogos (SeedServer e LCA). A origem do sistema foi inferida com base na origem dos seus genes.

Ao todo 310 genes e 795 interações foram encontrados. Foram produzidas quatro subvias de acordo com o estágio de desenvolvimento da mama: (1) Desenvolvimento Embrionário, (2) Puberdade, (3) Gravidez & Lactação e (4) Involução. A comparação entre as subvias revelou especificidades e generalidades entre as etapas de desenvolvimento: cada subvia possui um ou mais “reguladores-mestre”, enquanto alguns processos biológicos são compartilhados entre as subvias. Com relação à origem evolutiva foram encontrados genes com origem predita deste Organismos Celulares até Boreoeutheria, sendo que entre 80 e 97% dos genes já estavam presentes nos peixes (Gnathostomata-Teleostomi-Euteleostomi). Um processo frequente na história da vida é a co-opção, isto é, o recrutamento de vias pré-existentes para a geração de novas estruturas e funções. Este parece ser o caso da glândula mamária, já que o potencial genético para a geração de uma glândula mamária já existia muito antes da origem dos mamíferos. No entanto, a existência dos componentes não implica necessariamente na existência do sistema, mas na possibilidade da existência do sistema.

Este estudo traz contribuições para a compreensão tanto da biologia quanto da evolução da glândula mamária.



Biologia de sistemas e integração de dados

ENGENHARIA REVERSA DA REDE REGULATÓRIA DO SARCOMA DE EWING E DESCOBERTA DE MESTRES REGULADORES

Autores: Marcel da Câmara Ribeiro Dantas¹; Marialva Sinigaglia³; Caroline Brunetto de Farias^{3,4}; André Tesainer Brunetto³; Algemir Lunardi Brunetto³; Rodrigo Juliani Siqueira Dalmolin^{1,2};

E-mail para correspondência: vuln@ufrn.edu.br

Instituições: ¹Programa de Pós-Graduação em Bioinformática, Universidade Federal do Rio Grande do Norte; ²Departamento de Bioquímica, Universidade Federal do Rio Grande do Norte; ³Instituto de Câncer Infantil, Porto Alegre, RS; ⁴Laboratório de Câncer e Neurobiologia, HCPA, Universidade Federal do Rio Grande do Sul;

Palavras-chave: Sarcoma de Ewing; Transcriptômica; Redes Complexas

Apoio: CAPES, CNPQ, PROPESQ UFRN, Raffael Koff Acordi, Instituto do Cancer Infantil, PRONON

O Sarcoma de Ewing afeta ossos e tecidos moles e é o segundo câncer ósseo mais comum entre crianças. É uma doença caracterizada pela presença de uma translocação envolvendo o gene EWS e um outro da família ETS, geralmente FLI1. A oncoproteína EWS-FLI1 induz remodelagem da cromatina e altera a expressão gênica, ativando oncogenes e silenciando genes supressores de tumor, contribuindo para diferenciação e sobrevivência do tumor. A obtenção de redes regulatórias é crucial para a compreensão da fisiologia celular e de fenótipos de patologias poligênicas. Através de redes complexas, podemos visualizar este comportamento, onde os nós e arestas são calculados de acordo com a correlação dos dados de expressão e da informação mútua entre essas unidades regulatórias. O algoritmo utilizado aqui foi uma reimplementação do ARACNe e do Análise de Mestres Reguladores, em R no pacote RTN, encontrado no repositório Bioconductor. Os datasets utilizados foram de biópsias de sarcoma de ewing encontrados no Gene Expression Omnibus (GSE34620 e GSE63157), total de 202 pacientes, que levou a obtenção de uma rede com 783 unidades regulatórias. 1388 fatores de transcrição humanos foram utilizados a partir do levantamento de Fletcher 2013. De modo a executar a Análise de Mestres Reguladores, dependemos da assinatura da doença que pode ser obtida através de uma análise de expressão diferencial. O tipo celular de origem deste sarcoma é desconhecido mas largamente apontado pela literatura como ou célula tronco mesenquimal ou célula da crista neural, sugerindo inclusive a possibilidade de ambas. Nós as obtivemos e as utilizamos para executar o algoritmo, filtrando a posteriori pelos mestres reguladores encontrados em comum entre as duas análises. Estes são: CREB3L1, AEBP1, MEF2C, GLI3, PBX3, ARNT2, RUNX3 e PAX7. Após essa etapa, aplicamos um segundo filtro baseado na similaridade do conteúdo das unidades regulatórias entre coortes diferentes, o que nos levou a dois mestres reguladores: MEF2C e AEBP1. O MEF2C está envolvido no processo da apoptose e no controle do crescimento celular, enquanto que o AEBP1 é um repressor transcricional e participa da diferenciação celular. Eles tem sido apontados na literatura como envolvidos, respectivamente, no rabdomiossarcoma e glioblastoma. Esses resultados sugerem novos fatores de transcrição relacionados ao sarcoma de ewing ainda não apontados pela literatura que podem ser fundamentais para a compreensão da patofisiologia do sarcoma de ewing.



Biologia de sistemas e integração de dados

ANÁLISE DAS MUTAÇÕES MISSENSES RELACIONADAS AO CÂNCER DE ESTÔMAGO E SUAS CONSEQUÊNCIAS ESTRUTURAIS

Autores: Marília Viana Albuquerque de Almeida¹; Laíse Cavalcanti Florentino¹; Danilo Lopes Martins¹; Diego Arthur de Azevedo Moraes¹; Rodrigo Juliani Siqueira Dalmolin¹;

E-mail para correspondência: mariliav1985@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: câncer de estômago; mutações missense; efeito das mutações

Apoio: UFRN - Universidade Federal do Rio Grande do Norte BIOME - Bioinformatics Multidisciplinary Environment
IMD - Instituto Metr pole Digital CAPES - Coordena o de Aperfei amento de Pessoal de N vel Superior

O c ncer g strico est  entre as principais neoplasias malignas no mundo. Com o surgimento das tecnologias  micas se obteve entendimento mais aprofundado dos genes relacionados ao c ncer e suas respectivas muta es: drivers (MD) e passenger (MP). Abordagens de ferramentas computacionais e explora o de redes de intera o de res duos (RINs) em prote nas v m sendo desenvolvidas como forma de predi o desses tipos de muta es. O objetivo do estudo   analisar muta es *missenses* (MM) e suas consequ ncias estruturais relacionadas ao c ncer de est mago (STAD). Genes e suas respectivas muta es foram minerados a partir dos bancos *The Cancer Genome Atlas* (TCGA) e *Catalogue Of Somatic Mutations In Cancer* (COSMIC), e filtrados pelos par metros *S-Score* (identifica o e predi o de genes candidatos) e *Ndamage* (classifica o do impacto da muta o). As MM que ocorriam em regi es codificantes foram selecionadas. Estruturas tridimensionais de prote nas (PDBs) foram obtidas no banco pdb utilizando c digos UniProt. O mapeamento da posi o dos PDBs foi realizado pelo banco de informa es SIFTS e o efeito estrutural das muta es foi estimado a partir da constru o de RINs, utilizando o *software* RING 2.0. Arquivos contendo *Nodes* (cada amino cido da prote na) e *Edges* (conex es entre amino cidos na forma de intera es qu micas) foram gerados, tais informa es adicionadas ao banco de dados e relacionadas com as muta es. Os dados foram armazenados em MySQL. Observou-se o *Degree* (conectividade entre Nodes) e limitou-se este par metro at  15. Notou-se que existe alta quantidade de muta es em *Degrees* com valor at  cinco. Conforme aumenta o *Degree*, diminui a quantidade de MP e aumenta a propor o de MD. Observa-se nas MD maior propor o entre trocas de arom ticos para carga negativa, arom ticos para polar e carga negativa para polar. Nos *Edges*, as MD obtiveram maior propor o para intera es do tipo IAC e *pipistacks*. J  nas MP, tem-se maior propor o entre *pication* e pontes dissulfeto. As muta es est o em *degrees* baixos, mas MD apresentam tend ncia de ocorrer em n s que est o em *degrees* maiores, pois   esperado que maior conectividade possa provocar maior impacto nas prote nas. Altera es nos grupos qu micos e suas intera es podem estar associadas a poss veis mudan as estruturais nas prote nas, tendo potencial para afetar a atividade biol gica. A partir da avalia o de seus impactos na RIN, pode-se prever o efeito de uma muta o e associ -lo ao surgimento e a progress o do STAD.

Biologia de sistemas e integração de dados

CONSTRUÇÃO E ANÁLISE SISTÊMICA DA REDE REGULATÓRIA DA SEPSE

Autores: Raffael Azevedo de Carvalho Oliveira¹; Rodrigo Juliani Siqueira Dalmolin¹;

E-mail para correspondência: raffael.azevedo@gmail.com

Instituições: ¹Centro Multiusuário de Bioinformática - IMD, UFRN;

Palavras-chave: seps; biologia de sistemas; regulador mestre

Apoio: Capes

A seps (infecção generalizada) é um dos maiores problemas de saúde pública do mundo. Segundo dados da OMS, a maior causa de morte nos leitos de UTI é de complicações por pacientes sépticos. Por se tratar de uma síndrome inflamatória complexa, existe uma variedade de desfechos e respostas a tratamentos convencionais já estabelecidos. Hoje, a identificação da seps utiliza critérios exclusivamente clínicos, com poucos marcadores genéticos e/ou bioquímicos disponíveis. Apesar do grande volume de resultados promissores de pesquisas utilizando modelos animais de seps, uma porcentagem extremamente baixa dessas descobertas mostrou utilidade clínica efetiva. Uma abordagem recente e que tem apresentado resultados relevantes na compreensão de doenças complexas é a identificação de redes regulatórias. A partir de dados transcricionais, é possível identificar quais são os fatores de transcrição (FTs) majoritariamente envolvidos com os fenômenos biológicos avaliados. No presente trabalho, foram utilizados dados transcricionais de bancos de dados públicos com o objetivo de construir a rede regulatória da seps e identificar seus reguladores mestres (RMs). Para construir a rede, foi utilizado o pacote do R *RTN*, que utiliza informação mútua para construir as redes regulatórias, utilizando assinatura de choque séptico. Após construída a rede, foram identificados cerca de 100 RMs. Em seguida, foi aplicado um filtro utilizando dados transcricionais de outras doenças inflamatórias para eliminar RMs inespecíficos. Também utilizando o pacote *RTN* foi avaliada a atividade de cada regulon (RM + genes regulados) em todas as amostras. No total, 15 RMs apresentaram diferença na atividade entre as amostras e estes se mostraram candidatos: NR2E1, MEF2A, EPAS1, KLF5, PHTF1, KLF7, HES1, GAS7, CEBPB, JUNB, TRIM25, FOSL2, ZNF467, MTF1 e BCL6. Também foi avaliado se diferentes assinaturas de outras doenças inflamatórias enriqueciam o mesmo grupo de regulons destacado. Utilizando assinaturas de artrite reumatóide, dengue e esclerose múltipla, quando confrontada a rede de seps com estas assinaturas, foi verificado que pelo menos 95% dos 15 RMs continuavam sendo ativados. Conclui-se que diferentes assinaturas, de diferentes doenças inflamatórias, continuam ativando os mesmos regulons na rede de seps, o que pode indicar que eles realmente têm um papel-chave no desenrolar da seps, necessários para a progressão da doença, podendo também ser designados como alvos terapêuticos ou diagnósticos.



Desenvolvimento de softwares e banco de dados

DESENVOLVIMENTO DE UMA FERRAMENTA COMPUTACIONAL INTEGRADA PARA A DETECÇÃO DE POTENCIAIS NEOANTIGENOS

Autores: Ana Carolina Miranda Fernandes Coelho¹; Paulo Roberto Branco Lins¹; André Luiz Fonseca Faustino¹; Lucas Marques da Cunha¹; Sandro José de Souza¹;

E-mail para correspondência: amcoelho14@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: neoantígenos; cancer ; resposta imunológica

Neoantígenos são pequenos peptídeos originados por proteínas mutadas, resultantes de mutações somáticas não sinônimas, que são expressas pelas células tumorais. Essas moléculas são reconhecidas como não-próprias pelas células T e, conseqüentemente, podem elicitar potentes respostas imunológicas. Propomos aqui uma ferramenta computacional capaz de identificar potenciais candidatos a neoantígenos, através da automatização de diversas etapas, incluindo a predição de HLA, integração de dados de mutação e de expressão (RNA-Seq) e a predição da afinidade de ligação entre os neoepítotos e as moléculas de HLA.



Desenvolvimento de softwares e banco de dados

MOLECULAR DIAGNOSIS OF METABOLIC PATHWAYS: FROM IN SILICO PREDICTION BASED ON GENOME TO IN VITRO VERIFICATION

Autores: André Fonseca¹; Carolina Fonseca Minnicelli²; Marbella Maria da Fonsêca²; Inácio Medeiros^{1,2}; Rita Silva-Portela²; Jorge Estefano¹; Lucymara Agnez-Lima^{1,2};

E-mail para correspondência: carolminnicelli@gmail.com

Instituições: ¹Programa de Pós Graduação em Bioinformática, Instituto MetrÓpole Digital, Universidade Federal do Rio Grande do Norte; ²Laboratório de biologia molecular e genômica, Universidade Federal do Rio Grande do Norte;

Palavras-chave: primers; software; vias metabólicas

Metagenomic sequencing points to gene diversity, however, monitoring individual pathways and/or organisms requires a more feasible approach. The present work aimed to generate a package of publicly available programs and in-house scripts for designing clade specific primers. The program was written in Python and the input data was a FASTA file from gene sequences of interest. Selected sequences were aligned by MAFFT (<http://www.ebi.ac.uk/Tools/msa/mafft/>). Alignments were trimmed to remove ambiguously aligned regions and columns composed of more than 60% gaps using trimAl (<http://trimal.cgenomics.org>), avoiding excessively degenerate consensus. The program allowed generation of a phylogenetic tree by being coupled to FastTree2 (<http://www.microbesonline.org/fasttree/>). A greater phylogenetic distance between groups justified the design of different primers pairs. Primers were designed using the primer3 program (<http://primer3.sourceforge.net>), which allows user to apply biochemical parameters (e.g. amplicon size). Primers were obtained for a given target and chosen through in silico amplifications using TNTBLAST (<http://public.lanl.gov/jgans/tntblast>) by a best-alignment approach. For in vitro validation, PCRs conditions were standardized, followed by gene amplification and 6% polyacrylamide gel electrophoresis. Based on previously metagenome sequencing and analysis of unpublished work from our group, four genes were selected for amplification according to their presence in eight different genera. Additionally, two pairs of multiclade primers for the constitutive gene *gyrA* were generated as positive control of amplifications. All primers showed in silico amplification as part of the last step of pipeline. In vitro PCRs confirmed the presence of specific metabolic pathway genes in six metagenomes from our group with the expected molecular size. Because three samples had the total metagenomic DNA sequenced, it was possible, using the Integrated Microbial Genomes & Microbiomes (IMG), to confirm the presence of genes and organisms amplified by our diagnostic panel. We developed a bench validated computational pipeline for PCR primer design specific for a set of genes, whose architecture can be applied in different contexts. We are currently working on an algorithm to, through a dendrogram analysis, predict the maximum number of homologous sequences that can be amplified with the lowest number of primers.



Desenvolvimento de softwares e banco de dados

VCF PROCESSES: FERRAMENTA PARA ANÁLISE DE MUTAÇÕES GENÔMICAS.

Autores: Eliseu Jayro de Souza Medeiros¹; Jorge Estefano Santana de Souza¹;

E-mail para correspondência: eliseujayro@gmail.com

Instituições: ¹Instituto Metr pole Digital (IMD) - Universidade Federal do Rio Grande do Norte (UFRN);

Palavras-chave: Mutac o; SNP; Gene

Na atualidade a demanda por chamada de variantes de um genoma vem crescendo principalmente devido aos avanços na biologia molecular e na medicina de precisão, que visa o tratamento mais adequado e otimizado de acordo com o contexto genômico do paciente, é inegável que os estudos e análises pontuais do material genético de determinado organismo possam direcionar para um tratamento medicamentoso mais adequado e efetivo. O processo de anotação de variantes genômicas é um processo laborioso que exige muito esforço técnico/computacional, muitas vezes requerendo conhecimentos avançados de programação e sistemas operacionais como o linux. Dado e exposto o objetivo desse projeto foi o desenvolvimento de software para o sistema operacional Windows® com uma interface friendly que oferece uma forma facilitada de anotação e análise de variantes genômicas. Os procedimentos de anotação e identificação de relevância de mutações, são inicialmente padronizados em um arquivo bruto, denominado Variant Call Format (VCF), que servirá de input para ferramenta proposta. Para a anotação dos alelos variantes a plataforma utiliza bancos de polimorfismos (dbSNP), preditores de efeito deletério (Sift, Polyphen, Proven) e anotadores de variantes como o SnpEff e SnpSift. Sobre os alelos variantes essas ferramentas fornecem diversas características que serão convertidas em features internamente em nosso sistema. Desenvolvemos uma interface com objetivo de fornecer de forma facilitada a anotação de mutações, dando relevância quanto a sua patogenicidade. Além da anotação também foi construído um sistema de filtros de seleção de variantes sobre as colunas (features), atuando como filtro direto sobre as features para, por exemplo, fazer buscas por genes, posições genômicas, patogenicidades, etc. Remoção e seleção dinâmica de colunas de forma a facilitar o processo de análise em colunas de interesse de estudo, no arquivo. Por fim a ferramenta oferece a possibilidade de salvar os resultados obtidos pelas anotações e filtros utilizados, em arquivos no formato EXCEL ou CSV. Foi desenvolvido um software Windows® que facilita a análise de variantes genômicas, fornecendo a anotação gênica, informações de patogenicidade além de oferecer filtros para determinar a relevância mutacional no contexto gênico estudado. Link para o Software: <https://github.com/eliseumedeiros/VCF-ProcessesProject>



Desenvolvimento de softwares e banco de dados

GETVIEW - UMA FERRAMENTA PARA VISUALIZAÇÃO E MANIPULAÇÃO DE DADOS BIOLÓGICOS

Autores: Gabriel Augusto Fabricio Bezerra¹; João Pacifico Bezerra-Neto¹; Flávia Figueira Aburjaile¹; Ana Maria Benko Iseppon¹;

E-mail para correspondência: pggabrielfbezerra@gmail.com

Instituições: ¹Departamento de Genética - Universidade Federal de Pernambuco;

Palavras-chave: Desenvolvimento de software; Visualização de dados; Big Data

Apoio: CAPES , CNPQ e FACEPE

Com o aumento exponencial no volume de dados gerados em pesquisas na área da bioinformática, novas formas de visualizar, filtrar e analisar grandes quantidades de dados se fazem necessárias. Nos últimos anos, softwares foram desenvolvidos com foco na análise e visualização de dados, porém o volume destes dados tem sido um desafio. Neste contexto, existe uma demanda crescente por aprimorar as ferramentas computacionais que auxiliam no manuseio dos dados, principalmente quanto à sua mineração e visualização em plataformas web-dirigidas, que facilitem a identificação dos dados. Assim, este trabalho desenvolve a ferramenta GETView para a visualização de dados biológicos de grande volume por meio de interface gráfica, o que viabiliza a rápida interpretação dos mesmos. Esta ferramenta surge a partir do TrinotateWeb, um software baseado em comunidades abertas, onde contribuições e modificações do seu código são permitidas. Assim, a fim de melhorar seu funcionamento e implementação para nosso volume de dados, seu código original , foi modificado/reescrito favorecendo a interação com banco de dados, além de facilitar a ações por futuros usuários que venham a aprimorar o código atual. Em termos de função, o GETView propõe otimizar as formas de visualizar e analisar informações de expressão, bem como de anotação gênica, associada com os recursos da biblioteca em Javascript CanvasXpress. O programa permite a inserção de diversas informações em seu banco de dados, além de auxiliar na filtragem de tabelas com milhares de informações, dispondo ainda de recursos de visualização de dados de expressão gênica em heatmaps, além de ser possível o input de dados não previamente depositados na plataforma, permitindo a fácil manipulação de conjuntos de dados diversos. Assim, o tempo de análise é otimizado, visto que as análises poderão ser direcionadas para um conjunto maior de dados. Para explorar as diferentes formas de visualização o usuário pode inserir dados de expressão em formato “.tsv”, o que lhe permite gerar heatmaps interativos com parâmetros determinados. Todos os gráficos de heatmap gerados na plataforma desenvolvida são passíveis de download e manipulação pelo usuário para um conjunto maior de dados. Assim, o GETView permite diferentes formas de visualização de dados aos usuários visando tornar mais simples por meio de interface gráfica a análise de expressão diferencial dos genes e/ou proteínas, tornando-se um importante recurso para análises de ciências ômicas.



Desenvolvimento de softwares e banco de dados

SOFTWARE PARA SIMULAÇÃO DE CENÁRIOS EVOLUTIVOS EM POPULAÇÕES BIOLÓGICAS: A HIPÓTESE DO FENÓTIPO ESTENDIDO

Autores: Guilherme Fernandes de Araujo¹; Sandro Jose de Souza¹;

E-mail para correspondência: guilherme.araujo@imd.ufrn.br

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: Simulação biológica; Fenótipo estendido; Grafos

Simulações populacionais modeladas em grafos são utilizadas para estudar a fixação de genes que oferecem vantagens reprodutivas em indivíduos que competem em determinado ambiente. O coeficiente de seleção, bonificado nos indivíduos que possuem características vantajosas do ponto de vista reprodutivo pode levar a uma probabilidade maior de fixação desta característica, assim como pode contribuir para que ela aconteça de maneira mais rápida. Neste trabalho a simulação foi adaptada à hipótese do fenótipo estendido, onde a característica que oferece vantagem para um indivíduo não é um traço próprio, mas uma alteração realizada no meio ambiente, por exemplo, uma teia de aranha, que pode ser aproveitada por outros indivíduos mesmo após a morte daquele que a produziu. Foi desenvolvido um software na linguagem de programação Java para realizar simulações desta natureza em grafos de configurações variadas - grafo de Barabási-Albert, grafo completo e grafo regular - em diferentes tamanhos, 100 e 500 nós, sempre confirmando a hipótese de maior probabilidade de fixação quanto maior for o bônus do fenótipo estendido, mas nem sempre em relação a esta fixação ocorrer de maneira mais rápida. Os valores de bônus utilizados foram de 0, 0,05, 0,1, 0,15, 0,2, 0,25 e 0,3, sendo 1,0 o valor base para seleção aleatória para reprodução de um indivíduo. A simulação é executada até a fixação de uma das variantes, até o limite de 5000 ciclos, e a cada ciclo 4% da população é renovada - esta porcentagem de indivíduos é sorteada para morrer, e o mesmo número de indivíduos se reproduz, gerando uma cópia idêntica de si. Curiosamente, o tipo de grafo de relações entre os indivíduos tem efeito significativo sobre o tempo de fixação de uma variante na população, ocorrendo de maneira mais rápida em grafos completos e regulares, mas de maneira mais lenta em grafos de Barabási-Albert, sendo que este é um exemplo de rede livre de escala, utilizado em diversas simulações de interações em rede no mundo real.



Desenvolvimento de softwares e banco de dados

ATENDIMENTO À LEI AROUCA NO ENSINO DE FARMACOLOGIA ATRAVÉS DO PHARMASOFTWARE®

Autores: Laís Lacerda Brasil de Oliveira¹; Luina Benevides Lima¹; Raquel Carvalho Montenegro¹; Maria Elisabete Amaral de Moraes¹; Hellfada Vasconcelos Chaves¹;

E-mail para correspondência: laisbrasil@ufc.br

Instituições: ¹Universidade Federal do Ceará;

Palavras-chave: Alternativa ao uso de animais; Bioética; Ensino

A utilização de animais para fins didáticos é empregada em instituições de ensino para um melhor entendimento dos alunos e aprendizado mais eficaz. Indubitavelmente, aulas práticas de vivissecção são importantes no processo de ensino-aprendizagem do ciclo básico dos cursos de ciências da saúde, uma vez que permitem maior propriedade e facilidade do corpo discente em relação ao conteúdo trabalhado. De acordo com a Lei Arouca (Lei nº 11.794 – 2008), que estabelece critérios para “a criação e a utilização de animais em atividades de ensino e pesquisa científica, em todo o território nacional”, tem-se que, “sempre que possível, as práticas de ensino deverão ser fotografadas, filmadas ou gravadas, de forma a permitir sua reprodução para ilustração de práticas futuras, evitando-se a repetição desnecessária de procedimentos didáticos com animais”. Para atender essa lei, o objetivo deste estudo foi relatar o desenvolvimento de um software (Pharmasoftware®) como método alternativo de ensino em Farmacologia no curso de Medicina da Universidade Federal do Ceará (UFC), Campus Sobral. O protocolo experimental foi elaborado de acordo com as “Diretrizes Brasileiras para o Cuidado e a Utilização de Animais para Fins Científicos e Didáticos” e submetido à Comissão de Ética em Pesquisa Animal da UFC, Campus de Fortaleza. Para demonstrarmos as diferentes vias de administração (oral, subcutânea, intramuscular, intravenosa e inalatória), seis ratas (*Rattus norvegicus*) foram tratados por vias diferentes com hidrato de cloral (10%) e azul de Evans (2%), sendo que um dos animais foi submetido a ligadura do piloro no intestino antes da administração. Para registrar os vídeos foi utilizada uma câmera *full HD* e iluminação adequada. O Pharmasoftware® utilizou Java como linguagem de programação, que é uma linguagem multiplataforma onde o usuário tem a liberdade de utilizá-lo em qualquer sistema operacional. Para a certificação, foram utilizados dois grupos de alunos, um em aula prática convencional e outro com o Pharmasoftware®. A análise dos resultados revelou que ambas as atividades foram igualmente eficazes em auxiliar na consolidação do tema Vias de Administração dos Fármacos, evidenciando, assim, o Pharmasoftware® como uma ferramenta capaz de auxiliar no ensino da Farmacologia e, ainda, de fomentar a implantação de outros métodos alternativos de ensino de Farmacologia nas Instituições de Ensino Superior.



Desenvolvimento de softwares e banco de dados

UM ESQUEMA ESTRELA DE BANCO DE DADOS EM LARGA ESCALA PARA ANOTAÇÃO DE SNPS DE *ASPERGILLUS* SPP.

Autores: Rogério dos Santos Rosa¹; Antonyus Pyetro do Amaral Ferreira¹; Edna Natividade da Silva Barros^{1,2};

E-mail para correspondência: rogerio.bioinfo@gmail.com

Instituições: ¹Centro de Tecnologias Estratégicas do Nordeste; ²Universidade Federal de Pernambuco;

Palavras-chave: Banco de Dados; SNPs; *Aspergillus*

Apoio: CNPq

Há décadas o gênero de fungos *Aspergillus* tem despertado interesse da comunidade científica, tanto pelas características patogênicas de algumas espécies, como pelo potencial biotecnológico. Dentre as inúmeras aplicações desses microrganismos pode-se citar a produção de compostos de interesse comercial, a decomposição de biomassa, solubilização de compostos fosfatados. Dentre elas, a última é de interesse para a agricultura em regiões com baixa biodisponibilidade de fósforo, como no Nordeste brasileiro. Tais microrganismos têm capacidade de solubilizar tanto fosfatos naturais, quanto como os insolúveis residuais após a adição de fosfatos solúveis. Dessa forma, se aumenta a biodisponibilidade de fosfato, contribuindo para aumentar a produtividade. Identificar alterações genéticas pontuais, como *Single Nucleotide Polimorphysm* (SNPs) relacionadas a características específicas de cada espécie, pode vir a maximizar a aplicação do fungo em bioprocessos. Neste trabalho, aplicamos métodos de bioinformática para mapear SNPs em treze genomas de referência do gênero em questão. A estes genomas, *reads* coletados do SRA-NCBI foram alinhados. Sob os alinhamentos resultantes, aplicamos duas diferentes ferramentas de identificação de SNPs: *freebayes* e *mpileup*. Aos resultados obtidos foram empregados filtros estatísticos para selecionar apenas aquelas variações com no mínimo 99% de probabilidade de qualidade. Utilizando os mapas genéticos dos genomas de referência, cada SNP foi anotado de acordo com seu efeito. Por exemplo, ganho de códon de finalização, perda de códon de inicialização, etc. Cada efeito foi classificado como sendo de impacto “Alto”, “Baixo”, “Modificador” ou “Moderado”. Por exemplo, a ocorrência de SNP não sinônimo que gera o ganho de um códon de finalização prematuro é um SNP de “Alto” impacto. Para analisar as mais de 200 milhões de anotações resultantes do *pipeline* previamente descrito, foi modelado um esquema de banco de dados em formato estrela (*data warehouse*) no SGBD *PostgreSql*. O modelo desenvolvido possibilitou o cruzamento do SNP com a anotação do gene em que ele ocorre, para isto foram considerados os bancos de dados de anotação: KOG, KEGG e *Gene Ontology*. Como interface para visualização dos dados, foi empregada a plataforma OLAP *Pentaho*. Como resultado, montamos uma infraestrutura de *software* que permite identificar SNPs, anotá-los, cruzar informações de acordo com o gene que o contém, assim como consultar os resultados em uma plataforma *web*.



Desenvolvimento de softwares e banco de dados

APLICATIVO PARA MÉDICOS: CALCULADORA DE RISCO PARA CÂNCER HEREDITÁRIO

Autores: Sebastião Noberto Camêlo Pessoa Neto¹; Tirzah Braz Petta Lajus¹; Jorge Estefano Santana de Souza¹;

E-mail para correspondência: no.oli.pessoa@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: Cálculo de risco; Mutação germinativa; Hereditariedade

O câncer pode vir a se desenvolver com estímulo do ambiente e do estilo de vida, mas algumas pessoas possuem alterações genéticas hereditárias que as torna mais vulneráveis a desenvolver certos tumores. O câncer hereditário representa 5 a 10% de todos os casos da doença e predispõe os pacientes a tumores agressivos e prognóstico reservado. O sequenciamento genético pode revelar a presença de mutações que são conhecidas por aumentar o risco de desenvolver algum desse tipo de câncer, possibilitando a tomada de medidas preventivas e tratamento precoce, mas não é um serviço acessível a todos. É possível, todavia, restringir a população a um grupo de alto risco, onde o sequenciamento se faz realmente necessário para confirmar a hipótese diagnóstica. Para filtrar esse grupo de alto risco, o histórico familiar é analisado para averiguar o risco do paciente de ser portador de alguma mutação, já que esse tipo de câncer possui características de hereditariedade. Como as informações genéticas das síndromes que se procuram no histórico familiar são muito complexas, muitos médicos, desconhecendo os fatores etiológicos das síndromes, não se sentem seguros para indicar o teste genético. Além disso, muitos pacientes podem ter o teste solicitado quando não fazem parte do grupo de alto risco. Tomando como exemplo a Liga Contra o Câncer, no Rio Grande do Norte, que oferece um serviço de aconselhamento genético que atende pacientes encaminhados por oncologistas, ginecologistas, radioterapeutas e mastologistas, mais de um terço dos pacientes encaminhados não têm indicação para realizar o teste genético. Então, visando auxiliar os médicos a determinar se um paciente possui indicação para realizar o teste genético, foi desenvolvido o aplicativo web Riscan, para aplicação de um questionário nos pacientes. O aplicativo examina o histórico familiar e pessoal do paciente, verificando se atendem aos critérios das diretrizes da Agência Nacional de Saúde Suplementar (ANS), que regula os planos de saúde no Brasil. Os critérios da ANS se baseiam em diversos estudos oncológicos e protocolos clínicos, como os da National Comprehensive Cancer Network (NCCN). O grande diferencial do Riscan é que ele não somente faz indicação ao teste genético, mas também explicita quais genes testar, orientando o médico com métodos de análise escalonados cobertos pelo plano de saúde, para detecção de risco de doze síndromes diferentes como Glioblastoma, Síndrome de Cowden e Câncer de Mama e/ou Ovário.



Desenvolvimento de softwares e banco de dados

DESENVOLVIMENTO DE UMA FERRAMENTA PARA ANÁLISE DE VARIANTES NO BANCO DE DADOS 1,000 GENOMES

Autores: Hélber Palheta^{1,2}; Tânia Carlice-dos-Reis¹; Fabiano Moreira Cordeiro^{1,3}; Ândrea Ribeiro-dos-Santos^{1,3};

E-mail para correspondência: taniacarlice@gmail.com

Instituições: ¹Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, 66.075-110, Brasil; ²Centro de Tecnologia da Informação e Comunicação (CTIC), Universidade Federal do Pará, Belém, PA, 66.075-110, Brasil; ³Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Belém, PA, 66.073-005, Brasil;

Palavras-chave: Variantes; 1,000 Genomes; Patogenicidade

Apoio: i. Fundação de Amparo à Pesquisa do Estado do Pará (FAPESPA); ii. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); iii. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); iv. Rede de Pesquisa em Genômica Populacional Humana/RPGPH (Bio Comp/CAPES)

O uso de ferramentas *in silico* é uma excelente abordagem para a caracterização de variantes moleculares depositadas em banco de dados de genomas humanos como o *1,000 Genomes*. Neste banco estão depositados os sequenciamentos de 2.504 indivíduos fenotipicamente saudáveis. A análise do referido banco de dados pode auxiliar na melhor caracterização de variantes populacionais, assim como no entendimento de mecanismos de ação de várias doenças. Por outro lado, a manipulação de seus arquivos é difícil, principalmente para pesquisadores não programadores. Assim, o objetivo do presente trabalho foi desenvolver uma ferramenta para investigar a variabilidade de indivíduos cujo genoma esteja depositado em banco de dados *1,000 Genomes*, assim como caracterizar as variantes observadas e investigar padrão de patogenicidade de forma automatizada. No presente trabalho foram usados dados no formato .VCF do *1,000 Genomes* anotados pelo programa SNPEFF e armazenados num banco de dados usando o sistema Mysql. Foi desenvolvido um módulo VCF-extract, que seleciona os dados e os armazena. Para a visualização dos resultados desenvolveu-se uma aplicação web, o *Simple 4 Prediction* (S4P), que possui uma interface utilizada facilmente pelo usuário e permite aplicar filtros do seu interesse. Também possibilita realizar investigação de patogenicidade das variantes, por diferentes preditores, e no banco de dados CLINVAR. Para testar a funcionalidade da ferramenta, realizou-se simulação com o gene *CDH1*. Os resultados mostraram que a ferramenta permite acessar os dados anotados e a patogenicidade das variantes de forma organizada, simplificada e rápida. Conclui-se que a ferramenta desenvolvida *Simple 4 Prediction* (S4P), demonstrou ser um meio simplificado, rápido e eficiente para explorar dados de genomas.



Genes e genômica

RECONSTRUÇÃO DA HISTÓRIA EVOLUTIVA DE MOSQUITOS (DIPTERA:CULICIDAE) NEOTROPICAIS ATRAVÉS DE GENOMAS MITOCONDRIAIS

Autores: Alexandre Freitas da Silva¹; Laís Ceschini Machado¹; Carla Júlia da Silva Pessoa Vieira²; Roberta Vieira de Moraes Bronzoni²; Maria Alice Varjal de Melo Santos¹; Gabriel da Luz Wallau¹;

E-mail para correspondência: alexfreitasbiotec@gmail.com

Instituições: ¹Instituto Aggeu Magalhães-Fundação Oswaldo Cruz, Recife-PE.; ²Universidade Federal de Mato Grosso, Sinop-MT;

Palavras-chave: Filogenia; Vetores; Genômica

Apoio: CAPES, FACEPE e Núcleo de Bioinformática-IAM

Os mosquitos são insetos hematófagos de grande importância médica devido a capacidade de transmissão de patógenos causadores da malária, arboviroses e filarioses tanto para os animais quanto aos humanos. Devido ao avanço tecnológico nas plataformas de sequenciamento de DNA e bioinformática, o conhecimento da genética de mosquitos está expandindo. Apesar disso, atualmente pouco se conhece a respeito da evolução e a relação evolutiva entre várias espécies de mosquitos existentes. O genoma mitocondrial tem sido considerado um bom marcador molecular para estudo populacionais e filogenéticos sendo capaz de resolver a filogenia de vários grupos de organismos. Sendo assim, o objetivo deste trabalho foi caracterizar o genoma mitocondrial de mosquitos silvestres e estudar a sua relação evolutiva. Inicialmente foram realizadas diversas coletas de mosquitos em áreas de mata. Os mosquitos foram identificados morfológicamente, separados em pools, macerados em água ultra pura e utilizados para extração de DNA mitocondrial utilizando o kit QIAprep Spin Miniprep. O mtDNA foi quantificado em Qubit e a biblioteca genômica preparada através do kit Nextera XT, sendo sequenciada na plataforma Miseq Illumina. Além dos reads gerados, os reads brutos disponíveis em bancos de dados para diferentes espécies de mosquitos foram recuperados. Os reads foram trimmados através da ferramenta Trimmomatic 0.36 e utilizados para a montagem dos mitogenomas das diferentes espécies através do MITObim 1.8 e anotados no webservice MITOS. Os mitogenomas caracterizados neste trabalho e os recuperados em banco de dados foram alinhados através da ferramenta MAFFT 7. A partir do alinhamento foi reconstruída uma árvore filogenética de Máxima Verossimilhança através do PhyML 3.0. Foi possível caracterizar ao todo 23 novos genomas mitocondriais, 12 sequenciados neste trabalho e 11 a partir dos dados públicos. Além disso, foi possível realizar o posicionamento filogenético desses novos mitogenomas na árvore dos mosquitos, onde foram formados clados consistentes com o que já era previsto a partir dos estudos de morfologia. Dessa forma, através deste estudo foi possível conhecer a proximidade filogenética entre as espécies de mosquitos, incluindo genomas mitocondriais novos para a literatura, na filogenia da família Culicidae permitindo um maior entendimento da biologia desses insetos.



Genes e genômica

TRANSFERIBILIDADE DE MARCADORES MICROSSATÉLITES PLASTIDIAIS PARA ANÁLISES GENÉTICAS EM TRISCHIDIUM MOLLE (BENTH.) H.E. IRELAND, UMA LEGUMINOSA DA CAATINGA

Autores: Ana Lúcia Gonçalves da Silva¹; Paulo Aecyo Francisco da Silva¹; Tiago Esposito¹; Andrea Pedrosa-Harand¹; Inara Roberta Leal¹; Elâine Maria dos Santos Ribeiro¹;

E-mail para correspondência: lana48788@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Trischidium; Microsatélites; Caatinga

Apoio: PIBIC/CNPq/UFPE, CAPES, FACEPE

Os marcadores microsatélites ou SSR apresentam grande aplicabilidade em estudos sobre os efeitos da fragmentação de paisagens na diversidade genética e fluxo gênico entre populações. No entanto, o estabelecimento de novos marcadores encontra obstáculos devido ao alto custo e esforço empregados, sendo necessário o conhecimento prévio parcial do genoma da espécie a ser estudada. Desta forma, há interesse em transferir primers já desenvolvidos para espécies filogeneticamente relacionadas. *Trischidium molle* é uma leguminosa bem distribuída tanto em áreas preservadas quanto em regeneração da Caatinga, sendo por isso um bom modelo para investigar o impacto da perturbação antrópica nesse bioma. Sendo assim, o objetivo do presente estudo foi testar a transferibilidade de um conjunto de primers de SSR plastidial para *T. molle* a partir de primers universais e desenvolvidos para outra leguminosa. Para isso foram utilizados oito indivíduos adultos de *T. molle* provenientes de oito parcelas selecionadas no PARNA do Catimbau. As folhas foram dissecadas em sílica para extração de DNA baseada em CTAB. Os primers universais testados foram desenvolvidos para *Nicotiana tabacum* (ccmp1-ccmp10) e *Eucalyptus globulus* (EMCRC59, 60, 65, 67, 74, 84, 85, 86 e 90), além dos primers desenvolvidos para a leguminosa *Cenostigma microphyllum* (CmcpSSR4, 6-8, 12, 14-16, 18 e 19). Os locos amplificados foram marcados direta ou indiretamente com fluorocromos para genotipagem automática. Seis dos dez locos desenvolvidos para *N. tabacum* amplificaram (ccmp1-ccmp4, ccmp7 e ccmp10), assim como o loco EMCRC74 e nove dos locos desenvolvidos para *C. microphyllum*, exceto CmcpSSR16. Dos locos genotipados apenas os locos ccmp2 e CmcpSSR7 foram polimórficos (com 2 alelos cada) nessa subamostra do PARNA Catimbau, apesar dos locos genotipados apresentarem polimorfismo para as espécies para os quais foram desenvolvidos. O baixo polimorfismo apresentado indica que outros marcadores serão necessários para análises genéticas dessa espécie no Catimbau. Testes similares em indivíduos de outras localidades permitirão confirmar se o baixo polimorfismo observado é devido à alta estruturação genética resultado da baixa dispersão de suas sementes.



Genes e genômica

ANÁLISE DA EXPRESSÃO DO GENE AHSB ATRAVÉS DA BASE DE DADOS ALLEN BRAIN ATLAS PARA COMPREENSÃO DOS NÍVEIS ALTERADOS DE FETUÍNA-A EM PACIENTES COM CALCIFICAÇÃO NO GÂNGLIO BASAL

Autores: Anne Caroline Rosalen Meira¹;

E-mail para correspondência: annerosalen@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: AHSB gene; Fahr's Disease; Fetuin-A

Apoio: Advisor: Dr. João Ricardo Mendes de Oliveira

Introduction: Fetuin-A is a protein produced primarily by the liver, whose human gene is called α 2-Heremans-Schmid glycoprotein (AHSB) and resides on chromosome 3. It is an inhibitory protease that belongs to the cysteine superfamily. It has a high affinity for hydroxyapatite due to a dense order of acid residues extended on β -sheets within the D1 domain, one of its amino-terminals. Recently, serum fetuin-A levels were found to be lower in patients with bilateral basal ganglia calcification, suggesting that AHSB could also lead to dysfunction in stabilizing and preventing calcium and phosphorus precipitation. **Method:** The procedures were performed at the Laboratório de Imunopatologia Keizo Asami (LIKA) using Allen Brain Atlas database linked to bioinformatic tools. We investigated the expression of the AHSB gene in the basal ganglia through the analysis of microarray and brain explorer using postmortem brain from 6 male and female donors between 18 and 68 years of age with no known neuropsychiatric or neuropathological history. **Results:** By the analysis of gene expression using the heatmap, associated with observation of neuroanatomy and histology available on Allen Brain Atlas platform, the AHSB gene was found to be highly expressed in most parts of the brain, in both white and gray matter. Its expression, in general, increases in centrifugal direction. However, in areas of the basal ganglia, the focus of this study, its expression is much smaller. In globus pallidus, putamen and caudate nucleus the gene expression is low, with most of negative z-score values, between -2.46341 and -0.632968. In subthalamic nucleus the expression presents a slight increase, with mean values closer to zero, for example 0.124678. In substantia nigra, there is another subtle addition reaching expression values close to one, approximately 0.906397. **Conclusion:** The study concluded that low levels of AHSB gene expression in patients who do not have cerebral calcification indicate a possible vulnerability of the basal ganglia for imbalance in hydroxyapatite levels, which could explain the fact that low serum fetuin-A levels trigger calcium and phosphorus precipitation, characterizing a bilateral and symmetric pattern, often called as "Fahr's Disease". Due to the recent start of research on the subject, there are many unanswered questions, so this field needs to be better understood since it is such a complex disease.



Genes e genômica

IDENTIFICAÇÃO E CLASSIFICAÇÃO DA SUPERFAMÍLIA CITOCROMO P450S NO GENOMA DE FEIJÃO-CAUPI (VIGNA UNGUICULATA)

Autores: ARTEMISA NAZARE COSTA BORGES¹; José Ribamar Costa Ferreira Neto¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: borges.artemisanc@gmail.com

Instituições: ¹Universidade Federal de Pernambuco ;

Palavras-chave: CYP450; família gênica; estresse biótico

Apoio: CAPES, CNPq

O feijão-caupi sofre severas perdas devido à ocorrência de viroses, com ênfase para aquelas ocasionadas pelos vírus do Mosaico Severo do Caupi (CpSMV) e do Mosaico do Caupi Associado ao Afídeo (CABMV). Análises dos transcriptomas de acessos da referida espécie - 'BR14 Mulato', submetido a 60 min de infecção por CABMV, e 'IT85F', submetido a 60 min de infecção por CpSMV - mostraram que a superfamília proteica Citocromo P450 (CYP450) apresentou o maior número absoluto de transcritos diferencialmente expressos (induzidos ou reprimidos) dentre todas as famílias gênicas identificadas e categorizadas. Os acessos escrutinados são considerados resistentes às respectivas viroses, sendo sugerido que a família proteica CYP apresenta profundo impacto na reorganização fisiológicas dos genótipos nas condições impostas. Dessa forma, foi realizado um estudo genômico para identificação e classificação da referida superfamília no genoma de feijão-caupi. Para tanto, sequências de proteínas CYPs de várias espécies de leguminosas foram recuperadas do UNIPROT para serem utilizadas como sondas. Essas, por sua vez, foram alinhadas (BLASTp; cut off $\leq e-10$) contra o proteoma putativo (Vunguiculata_469_v1.1p.) de feijão-caupi, alocado no banco de dados Phytozome. A análise revelou a presença de 379 proteínas CYPs em feijão-caupi. Essas são codificadas por 285 loci (Vunguiculata_469_v1.1g.), número superior ao apresentado para outras leguminosas, a exemplo do feijão-comum (*Phaseolus vulgaris*). Os cromossomos 10 e 3 são os maiores ancoradores de loci codificadores de CYPs, com 48 e 44 loci, respectivamente. Em seguida, o alinhamento múltiplo das sequências foi realizado no software Clustal Omega e o método de *Neighbor-Joining* foi utilizado para realizar a análise fenética por meio da ferramenta MEGA7 (*bootstrap*=1000). Foi observado que essas proteínas foram pertencentes a oito clans, consistindo em 41 famílias. Dentre eles, o clan CYP71 (A-type) apresentou o maior número de proteínas CYPs (186), apresentando 17 famílias. As CYPs non-A-type corresponderam a sete diferentes clans (72, 727, 97, 86, 711, 74, 85), que contemplaram 24 famílias. Todos os clans analisados apresentaram os domínios característicos 'I-helix, K-helix', '(P)ER(F)' e 'heme-binding'. O presente trabalho agregou informações à classificação e caracterização estrutural das CYPs em feijão-caupi, bem como destacou seu potencial biotecnológico, devido à sua massiva participação na reposta às infecções virais analisadas.

Genes e genômica

ANÁLISE PARALELA DOS GENES LIGADOS À CALCIFICAÇÃO CEREBRAL FAMILIAR PRIMÁRIA

Autores: Arthur Cezar Pereira Barbosa^{1,2}; Cláudio César Silva Pereira^{1,2}; Paula Vitória Macêdo de Barros^{1,2}; João Ricardo Mendes de Oliveira^{1,2};

E-mail para correspondência: arthurbarbosa22@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco; ²Laboratório de Imunopatologia Keizo Asami;

Palavras-chave: Calcificação Cerebral; Allen Brain Atlas; Genética

Apoio: Agradecemos imensamente ao CNPq e à Propesq-UFPE

Calcificação Cerebral Familiar Primária (PFBC) é um distúrbio neuropsiquiátrico caracterizado por depósitos de cálcio nas regiões dos núcleos da base, tálamo e cerebelo, que se dispõem bilateral e simetricamente. Há 4 genes identificados envolvidos na fisiopatologia da PFBC: SLC20A2, PDGFB, PDGFRB e XPR1. O SLC20A2 codifica uma proteína transmembrana responsável pelo transporte de fosfato inorgânico denominada PiT-2. O PDGFB codifica a subunidade beta do fator de crescimento derivado de plaquetas, que forma dímeros e se liga ao receptor do PDGF. O PDGFRB codifica o receptor da subunidade beta. Esses 2 genes atuam na proliferação e recrutamento de pericitos durante a angiogênese no sistema nervoso central. O XPR1 codifica uma proteína responsável pelo efluxo de fosfato da célula. Nosso trabalho objetivou analisar os padrões de expressão desses genes no encéfalo humano. Para esse fim, utilizamos o banco de dados do Allen Institute (acessado via <http://human.brain-map.org/>), que dispõe de análises de expressão gênica de 6 doadores humanos. Esses dados são obtidos através da técnica de *microarrays*. Para a obtenção dos dados são utilizadas diferentes sondas e as análises são quantificadas e expressas em *Z-score*. Obtivemos como resultados a percepção de áreas no sistema nervoso central que apresentam padrões comuns de alta expressão gênica: o tálamo e os núcleos cerebelares; e áreas que apresentam padrão similar de baixa expressão gênica: o córtex cerebelar. Tomando 1 dos 6 doadores e utilizando a sonda número 1 para a análise do tálamo ventral, obtivemos *z-score* de 2,20108 (SLC20A2); 1,24887 (XPR1); 1,02028 (PDGFB); e 1,01626 (PDGFRB), demonstrando alta expressão desses genes. Tomando o mesmo doador e a sonda número 1, o *z-score* no córtex cerebelar foi -0,487653 (SLC20A2); -0,352154 (XPR1); -0,612076 (PDGFB); e -1,92748 (PDGFRB), demonstrando baixa expressão. A análise dos núcleos da base não demonstrou padrões homogêneos de expressão desses genes: No corpo estriado, PDGFB e PDGFRB apresentaram alta expressão; no globo pálido o SLC20A2 apresentou alta expressão, diferentemente do XPR1, apesar das proteínas codificadas por esses genes exercerem ações complementares na homeostase do fosfato. Em suma, encontramos regiões no sistema nervoso onde os padrões de expressão desses genes são similares, no tálamo e cerebelo, já nos núcleos da base esses padrões de expressão não se relacionam de modo claro. Assim, são necessários mais estudos para a elucidação da PFBC.



Genes e genômica

CARACTERIZAÇÃO E ANÁLISE FILOGENÉTICA DE ELEMENTOS TRANSPONÍVEIS DA SUPERFAMÍLIA TC1-MARINER DE EUCHROMA GIGANTEA (COLEOPTERA: BUPRESTIDAE)

Autores: Catarine Aragone de Albuquerque Mello¹; Igor Costa de Amorim¹; Crislaine Xavier¹; Gabriel da Luz Wallau²; Rita de Cássia de Moura¹;

E-mail para correspondência: catarinearagonemello@hotmail.com

Instituições: ¹Laboratório de Biodiversidade e Genética de Insetos ? Instituto de Ciências Biológicas, Universidade de Pernambuco, Recife, PE ? Brasil.; ²Departamento de Entomologia, Instituto de Pesquisas Aggeu Magalhães ? FIOCRUZ, Recife, PE, Brasil;

Palavras-chave: Transferência horizontal; TEs; Elementos estruturalmente conservados

Os elementos de transposição (TEs) são sequências repetitivas com capacidade de mobilidade dentro do genoma e podem ser classificados em diferentes superfamílias. Dentre essas, *Tc1-Mariner* se destaca pela diversidade e por apresentar a maior quantidade de eventos de transferência horizontal relatada. Essa transferência se caracteriza pela passagem de material genético independente da reprodução. Neste trabalho foi realizada a caracterização de elementos *Tc1-Mariner* no genoma de *Euchroma gigantea* (Coleoptera: Buprestidae), visando identificar TEs estruturalmente conservados e possíveis eventos de transferência horizontal. O sequenciamento genômico foi realizado na plataforma Illumina e a caracterização dos TEs no RepeatExplorer. Os clusters foram remontados no servidor do CAP3. As sequências obtidas foram examinadas no ORFfinder e submetidas ao Blastp (NCBI) e Censor (Repbase), visando obter TEs homólogos. As análises filogenéticas foram realizadas pelo método de máxima verossimilhança. No genoma de *E. gigantea* foram identificados 41 elementos estruturalmente conservados, com domínios de Tc1-Mariner íntegros (HTH, integrase, transposase e/ou DDE). Esses TEs pertencem às famílias *Mariner* (22 TEs), *Tc1* (10), *Lrsa* (sete) e *Pogo* (dois). As análises filogenéticas revelaram incongruência entre as filogenias dos elementos e as dos hospedeiros, visto que TEs de *E. gigantea* se agruparam preferencialmente com TEs de espécies de Hymenoptera e Diptera. Os elementos de *E. gigantea* apresentaram alta similaridade com TEs de espécies filogeneticamente distantes, incluindo as moscas *Bactrocera tryoni* (similaridade de 97.68%) e *Ceratitis rosa* (94.91%), as formigas *Camponotus floridanus* (96.77%) e *Linepithema humile* (94.01%) e a abelha *Apis mellifera* (95.36%). Quando comparados aos TEs de espécies próximas, esses valores foram semelhantes ou menores ao do buprestídeo *Agrilus planipennis* (97.4%) e maiores que os TEs pertencentes a outras famílias de Coleoptera, como *Onthophagus taurus* (Scarabaeidae) que apresentou similaridade de 69.69%. As espécies filogeneticamente distantes possuem interações ecológicas indiretas com *E. gigantea*, visto que em todos esses gêneros algumas espécies são parasitas ou polinizadores de plantas da família Malvaceae. A caracterização de TEs *Tc1-Mariner* em *E. gigantea* permitiu identificar diferentes elementos estruturalmente conservados e prováveis eventos de transferência horizontal entre diferentes espécies de insetos.



Genes e genômica

QUAIS SÃO AS ÁREAS DE EXPRESSÃO DOS GENES ISG15 E USP18 NO CÉREBRO HUMANO?

Autores: Cláudio César Silva Pereira¹; Arthur Cezar Pereira Barbosa¹; João Ricardo Mendes de Oliveira¹;

E-mail para correspondência: claudiocsp@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Áreas de expressão; Genes; Cérebro humano

Apoio: CNPq e Propesq-UFPE

O ISG15 e o USP18 são genes envolvidos na modulação da cascata do Interferon, a qual é envolvida no combate de infecções virais. Além de sua função imunológica, esses genes despertam a curiosidade da comunidade científica porque uma mutação neles pode desencadear o início de processos calcificativos no cérebro humano, com diversas manifestações clínicas associadas. Essas calcificações podem ter um padrão de dispersão semelhante às síndromes TORCH, como ocorre na deficiência do USP18 ou uma dispersão simétrica e bilateral, evidenciada na deficiência do ISG15. Sendo assim, devido à diferença de dispersão nos processos calcificativos vistos na mutação dos 2 genes, o presente estudo teve como objetivo a verificação do padrão de expressão genética de ambos, dando-se o enfoque para as regiões mais propensas às calcificações cerebrais. Assim, a base de dados selecionada para a produção do trabalho foi o Human Brain, presente no site do Allen Brain Atlas (acessado via <http://www.brain-map.org/>), devido a sua vasta quantidade de microarranjos de DNA em cérebros humanos. Esses microarranjos foram obtidos através da análise de amostras coletadas de 6 doadores, dispostas em 2 sondas teste, subdivididas em 169 estruturas cerebrais e quantificadas em *Z score*, na qual os maiores valores indicam uma maior expressão do gene e os menores valores, uma menor expressão. Desse modo, na elaboração do trabalho, os dados de microarranjo de DNA obtidos foram separados conforme as regiões cerebrais disponíveis no Atlas e tabelados em planilhas do Microsoft Excel®. Após, foram calculadas as médias dos valores de expressão genética para todas as regiões disponíveis, selecionando-se as maiores e menores médias para a produção de gráficos. Dessa maneira, os resultados obtidos para o gene ISG15 demonstraram um maior *Z score* em sub-regiões do tálamo dorsal ($Z=1,77917/1,59967/1,49947/1,40695$), bulbo ($Z=1,6472/1,59818/1,56712/1,32685$), ponte ($Z=1,70386$) e núcleo globoso do cerebelo ($Z=1,37393$). Um menor *Z score* foi visto para as áreas dos lobos do cerebelo posterior ($Z=-1,90270/-1,79698/-1,76120/-1,68597$) e anterior ($Z=-1,78361/-1,78098$) e posterior do vermis ($Z=-1,85855/-1,82421/-1,75776/-1,72878$). Os dados referentes à expressão genética do USP18 ainda estão em andamento. Por fim, conclui-se a partir do trabalho que o ISG15 é mais expresso em sub-regiões do tálamo dorsal, bulbo, ponte e núcleo globoso do cerebelo e menos expresso em sub-regiões do cerebelo posterior, anterior e vermis.



Genes e genômica

CARACTERIZAÇÃO PARCIAL DE HOX CLUSTERS GENES NA RELAÇÃO ESTRUTURA CORPORAL DE ARAPAIMA GIGAS.

Autores: Danilo Lopes Martins¹; Diego Gomes Teixeira¹; André Maurício Ribeiro dos Santos^{4,5}; Pitágoras de Azevedo Alves Sobrinho¹; João Paulo Matos Santos Lima^{1,2,4};

E-mail para correspondência: danilolmartins@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte; ²Bioinformatics Multidisciplinary Environment ? BioME; ³Instituto do Cérebro; ⁴Departamento de Bioquímica; ⁵Universidade Federal do Pará;

Palavras-chave: Hox clusters; Sintenia; Alinhamento

Apoio: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Os genes Hox são opções atrativas para a compreensão da base genética da diversidade morfológica em animais, por desempenharem papéis críticos na definição da identidade dos segmentos corporais. Vários estudos destacaram o papel crucial dos genes Hox no estabelecimento da estrutura corporal dos teleósteos. Por exemplo, a duplicação de clusters Hox na Enguia europeia (*Anguilla anguilla* L., 1758) estão envolvidos no padrão primário do eixo do corpo, o que implica um papel funcional para o excesso de Hox no alongamento axial. A proposta do presente trabalho é a identificação de genes Hox no Pirarucu (*Arapaima Gigas*), espécie endêmica da Bacia Amazônica, com forte apelo comercial podendo chegar até o peso total de 200 kg, que mostrem a relação com seu padrão de desenvolvimento corporal. A análise de sintenia foi processada utilizando como referência um draft genômico do *Arapaima Gigas*, realizado em nosso laboratório, e as proteínas das demais espécies utilizadas para comparação (*Danio Rerio*, *Anguilla Anguilla*, *Oryzias latipes* e *Takifugu rubripes*) obtidas do UniProt e o grau de identidade foi obtido pelo alinhamento das regiões com a ferramenta Blast. Com o resultado de alinhamento das seqüências de proteínas foi possível a identificação preliminar de clusters Hox (genes Hox e os seus genes flanqueadores), tornando possível a análise de homologia entre essas espécies. Como Identificamos sintenia a nível de proteínas entre o *A. Gigas* e as espécies analisadas, com parâmetros de filtragem para o alinhamento: e-value < 1E-5, porcentagem de identidade e porcentagem de cobertura > 80%. Com isso foram identificados preliminarmente 8 clusters Hox com 20 genes (Hox e flanqueadores) para diferentes scaffolds, justificando que essas regiões mais conservadas se encontram presentes no *A. Gigas*. Em um passo futuro iremos investigar a sintenia local nos hox clusters entre as espécies estudadas, tentando correlacionar o tamanho e formato do corpo com os mesmos e por fim realizar a quantificação dos níveis de expressão gênica para estes conjuntos de genes, em busca pela relação da variação de tamanho com o nível de expressão.

Genes e genômica

ANÁLISE DOS PADRÕES DE EXPRESSÃO NEUROANATÔMICOS DOS GENES ENVOLVIDOS COM CALCIFICAÇÕES CEREBRAIS

Autores: Denis Arthur Pinheiro de Moura¹; João Ricardo Mendes de Oliveira^{1,2};

E-mail para correspondência: drmouradap@gmail.com

Instituições: ¹Laboratório de Imunopatologia Keizo Asami, Universidade Federal de Pernambuco; ²Departamento de Neuropsiquiatria, Universidade Federal de Pernambuco;

Palavras-chave: Bioinformática; R; Calcificação Cerebral

Apoio: Fundação do Amparo a Ciência e Tecnologia - FACEPE

Calcificações cerebrais são achados relativamente freqüentes em exames de neuroimagem. Calcificações cerebrais, como a calcificação da glândula pineal e do plexo coróide, podem ser processos dependentes de idade, geralmente assintomáticos. Outras formas de calcificação são envolvidas com processos patológicos, associadas à gliose e morte neuronal. As áreas mais freqüentemente afetadas por calcificações patológicas são os gânglios da base, cerebelo, tálamo e o mesencéfalo. Os mecanismos causadores das calcificações ainda não foram elucidados para a maioria dos casos, embora alterações em algumas moléculas já tenham sido associadas com neuropatologias específicas. O objetivo deste trabalho foi analisar o perfil de expressão neuroanatômica dos genes associados a diversos tipos de calcificação cerebral em indivíduos sem calcificações. Os dados de *microarray* foram extraídos do banco de dados público Allen Human Brain Atlas, com amostras consistindo em seis cérebros humanos sem calcificações cerebrais e cerca de 60 mil sondas para cada região cerebral. Os dados do banco foram carregados na forma de *dataframe* na linguagem R (ver. 3.4.4) utilizando o R Studio (ver. 1.1.442) e foram analisados os padrões de expressão ao longo do cérebro para o conjunto de genes (*SLC20A2*, *SLC20A1*, *XPR1*, *PDGFB*, *PDGFRB*, *ZBTB20*, *PCDH12*, *BGLAP*, *BMP2*, *BMP4*, *CA2*, *CASR*, *GAPDH*, *GRIN2B*, *PPP1R1B*, *PTH2R*, *PTHLH*, *RUNX2*, *SP7*, *SPARC*, *SPP1* e *VDR*). Testes T foram utilizados para comparar os níveis de expressão entre áreas suscetíveis e áreas geralmente não afetadas pelas calcificações. Dos genes relacionados com as calcificações e hipoparatiroidismo, os genes *BGLAP* e *GRIN2B* tiveram níveis de expressão significativamente maior em áreas não afetadas, em relação às áreas suscetíveis, enquanto o gene *SPARC* teve níveis de expressão maiores nas áreas afetadas. Os genes relacionados à Calcificação Cerebral Familiar Primária não tiveram diferenças significativas dos níveis de expressão entre áreas suscetíveis e áreas não afetadas pelas calcificações. Os resultados indicam a existência de distintos processos que resultam em calcificações e envolvem moléculas que normalmente não têm expressão diferenciada entre áreas afetadas e não afetadas. É possível que as alterações associadas às calcificações cerebrais possam alterar os níveis de expressão regional de algumas dessas moléculas, levando a perda da homeostase do fosfato e do cálcio e ocasionando na formação das calcificações em indivíduos afetados.



Genes e genômica

ANALYSIS OF 16S GENE SEQUENCES IN METHANOGENIC ARCHEAS

Autores: Erick de Aquino Santos^{1,2}; Keyla Vitória Marques Xavier^{1,2}; Dailton Augusto Andrade Moraes¹; Camila Souza Santos¹; Michely Correia Diniz^{1,2};

E-mail para correspondência: erickdeaquinosantos@gmail.com

Instituições: ¹Universidade Federal do Vale do São Francisco; ²Geis - Grupo de Estudos Integrados do Semiárido;

Palavras-chave: biomethanization; conserved motifs; non-cultivated archaea

Anaerobic reactors have promoted a consolidation of the biotechnology in the treatment of waste water, mainly organic. The anaerobic microorganisms (in Bacteria and Archaea) transform organic matter into methane. Their joint action results in four stages of biomethanization: hydrolysis, acidogenesis, acetogenesis and methanogenesis. Methanogenic archaeas are essential in the anaerobic microbial degradation of organic waste resulting in methane, a gas with potential to cause environmental damage, but also has value as a non-fossil fuel. Despite its fundamental role as terminal oxidants in a complex microbial community, little is known about the methanogenic community. In bacterial groups, the gene encoding the 16S subunit of ribosomal RNA (*16S rRNA*) manifests a high conservation degree, used in microbial identification and diversity characterization researches. The objective of this study was to verify the conserved motifs in 16S ribosomal gene of non-cultivated archaea methanobacteria. Sequence search was performed in (NCBI) (<https://ncbi.nlm.nih.gov>), Taxonomy session by selecting environmental samples. The conserved motifs were traced through the MEME online tool (meme-suite.org/tools/meme) in default settings with three different motifs and an average size of 50 bases. Lower e-values results were considered significant. Bioedit v. v.7.2.6 software was used to obtain the consensus and nucleotide composition sequences. Of the 25 resulting sequences, 12% were from petroleum Brazil. Two groups were formed by similar size of sequence. The nucleotide composition in the first group was 15.24% A; 13.16% C; 21.25% G; 9.24% T, the second group has 19.97% A; 20.11% C; 28.58% G; 15.15% T. The group with 11 sequences has approximately 300 bp and the second group with 14 sequences, approximately 1000 bp. For the motif with the lowest e-value in the first group, 63.6% of motifs were in position 154-203, in the second group two motifs presented the lowest e-value with equal values, predominantly a motif with 35.7% in position 166-215, the other with 35.7% in position 915-964. The conservation verification corroborates to increase the knowledge related to the methanogenic bacteria, and to infer that the gene may present evolutionary importance to the group and maintenance of its functions.



Genes e genômica

ANÁLISE FUNCIONAL DE GENOMAS COMPLETOS DE XANTHOMONAS CAMPESTRIS E XANTHOMONAS CITRI

Autores: Juan Carlos Ariute Oliveira¹; Flávia Figueira Aburjaile¹; Ana Maria Benko Iseppon¹;

E-mail para correspondência: faburjaile@gmail.com

Instituições: ¹Laboratório de Genética e Biotecnologia Vegetal, Universidade Federal de Pernambuco, Recife, PE, Brasil;

Palavras-chave: genômica; fitopatógenos; bioinformática

Apoio: CAPES, CNPq, FACEPE.

A região nordeste tem se destacado pelo aumento da viticultura mesmo nas condições adversas do clima semiárido, com ênfase para o Vale do Submédio do Rio São Francisco. Entre os fatores cruciais que podem afetar a produção, destaca-se o cancro bacteriano, doença causada por bactérias do gênero *Xanthomonas*. As medidas preventivas bem como o controle através de produtos químicos ainda ocorrem de maneira ineficiente, causando prejuízos significativos. Deste modo, o objetivo do presente trabalho foi realizar a anotação e comparação dos genomas de diferentes linhagens de *X. campestris* e *X. citri*, com o intuito de caracterizar genes e/ou proteínas presentes em ilhas de virulência e patogenicidade. Para isso foram inicialmente selecionados e baixados 10 genomas de *X. campestris* e 53 de *X. citri* depositados no NCBI. Todos os genomas foram anotados automaticamente pelo RAST Server. Em seguida, um genoma foi curado manualmente, servindo de referência para a transferência de anotação de todos os outros. A curadoria foi realizada utilizando o programa Artemis e os bancos de dados biológicos do NCBI, Uniprot, InterProScan e Pfam. Para a categorização ontológica utilizou-se a ferramenta GOFeat. Os resultados de ontologia obtidos foram: 8656 hits no genoma de referência, distribuídos em 4266 com funções moleculares, muitos envolvidos com o processo de invasão da célula vegetal; 2497 envolvidos em processos biológicos, destacando proteínas envolvidas na regulação de expressão gênica e também no transporte de íons; e 1893 participantes de componentes celulares. Com o enriquecimento GO foi possível identificar proteínas específicas que atuam nos processos de regulação gênica e na secreção de proteínas. As análises permitiram a identificação e caracterização de novas regiões codificantes que não haviam sido descritas, consideradas essenciais para a compreensão da interação planta-patógeno. Observou-se que o repertório gênico é pouco variado, sendo compartilhado entre as duas espécies de *Xanthomonas*, o que facilitaria o estabelecimento de um método universal de combate ao cancro bacteriano em videira. Novos estudos complementares serão realizados a partir dos dados extraídos para a obtenção de um produto biotecnológico eficiente para a agricultura da região.



Genes e genômica

DESENVOLVIMENTO DE PIPELINE PARA IDENTIFICAÇÃO DE GENES ENVOLVIDOS EM VIAS DE SÍNTESE DE BIOSURFACTANTES EM DADOS DE METAGENÔMICA.

Autores: Inácio G. Medeiros^{1,3}; Marbella M. Fonseca³; Carolina Minnicelli³; Wydemberg J. Araújo³; Alaine de Brito Guerra³; Sinara Carla da Silva Araújo³; Jorge Estefano Santana de Souza^{1,2,4}; Lucymara Fassarella Agnez Lima^{1,3};

E-mail para correspondência: inaciogmedeiros@gmail.com

Instituições: ¹Programa de Pós Graduação em Bioinformática, Instituto MetrÓpole Digital, Universidade Federal do Rio Grande do Norte; ²Bioinformatics Multidisciplinary Environment, Instituto MetrÓpole Digital, Universidade Federal do Rio Grande do Norte; ³Laboratório de Biologia Molecular e Genômica, Centro de Biociências, Universidade Federal do Rio Grande do Norte; ⁴Instituto MetrÓpole Digital, Universidade Federal do Rio Grande do Norte;

Palavras-chave: biosurfactantes; metagenômica; pipeline

Identifying and producing biosurfactants at low cost is a high demand in scientific community. From many related works that emerged recently about bioproducts identification in metagenomic data, few are targeted on biosurfactants, which makes this field incipient. Facing that, this work proposes a pipeline for identification of genes involved in biosurfactant pathways in metagenomic data. Our pipeline adapts a deep-learning based approach called DeepARG (bench.cs.vt.edu/deeparg), using like training data set a reference subset of 478 genes from BioSurfDB (www.biosurfdb.org). It performs three initial steps over input data set: Quality control, removing low-quality sequences or regions through FASTX-Trimmer (hannonlab.cshl.edu/fastx_toolkit); assembly in contigs through MEGAHIT (github.com/voutcn/megahit); ORF prediction through FragGenScan (sourceforge.net/projects/fraggenescan). ORFs are submitted to adapted DeepARG and a list indicating which pathways each ORF is associated to is presented. Aiming to test accuracy of proposed method, two data sets with 232 sequences each were produced for testing sensibility and specificity of DeepARG, first with genes from BioSurfDB (biosurfactants), and second with non biosurfactants from RefSeq NR (www.ncbi.nlm.nih.gov/refseq). Data sets were applied to DeepARG, yielding a sensibility degree of 83% and a specificity degree of 100%. Once perceived DeepARG efficiency, proposed pipeline were applied to three petroleum well-biome metagenomes, from which it was obtained 45 ORFs related to five biosurfactant pathways. These ORFs were aligned to BioSurfDB and NR. Twelve ORFs aligned to BioSurfDB, being ten related to DnaK gene, and two related to DnaJ gene. Thirty-three ORFs aligned to NR, being related to eleven genes, from which DnaK and DnaJ were most representatives, both involved in the pathway of biosurfactant Putisolvin, which acts in soil bioremediation processes that contributes to petroleum-well microbiome maintenance. Bacteria genus Halomonas, Putisolvin producer, prevailed in ORF taxonomic analysis (about 70% of hits), enhancing assumption that obtained ORFs correspond to cited genes. Proposed method had good performance identifying biosurfactant pathway genes, which suggests its use in metagenome analysis aiming to find relevant genes for industrial application.



Genes e genômica

USANDO REDES DE INTERAÇÃO DE RESÍDUOS PARA ENTENDER AS CONSEQUÊNCIAS ESTRUTURAIS E A ORIGEM DE POLIMORFISMOS EM GENOMAS DO VÍRUS ZIKA.

Autores: Ivan Alisson Cavalcante Nunes De Lima¹; Diego Gomes Teixeira¹; Laise Cavalcanti Florentino¹; Selma Maria Bezerra Jerônimo¹; João Paulo Matos Santos Lima¹;

E-mail para correspondência: ivanalisson@hotmail.com

Instituições: ¹Universidade Federal Do Rio Grande Do Norte;

Palavras-chave: Flavivírus; RINs; Proteína não estrutural

Apoio: CNPQ

O vírus da Zika é do gênero Flavivirus, composto por uma fita única de RNA, com genoma de 10.7 kb e onze genes identificados, dentre estes, três compõem a região estrutural do vírus e os restante a não estrutural. Seguindo os novos estudos que associam as modificações nas regiões não estruturais como possíveis volantes para o surgimento de sintomas de microcefalia e a síndrome de Guillain-Barré, o presente estudo utiliza uma abordagem de construção de redes de interação de resíduos (RINs) para as proteínas do Zika, com o intuito de encontrar novas informações que possam auxiliar num melhor entendimento da dispersão desse vírus. Uma RIN permite a observação da estrutura das proteínas como um grafo, analisando cada aminoácido como um nó e as ligações químicas efetuadas com os demais aminoácidos na como arestas. Para isso, sequências das regiões NS1, NS3 e NS5 do vírus ZIKA foram obtidas através do banco Virus Variation, enquanto as estruturas das proteínas foram obtidas a partir do banco PDB e depois analisadas com o programa RING2.0. Todo o pipeline foi desenvolvido em python, estratégias de alinhamento local foram utilizadas para a categorização dos trechos a serem analisados e em seguida aplicamos scripts para a busca e mapeamento das regiões polimórficas. Uma vez identificadas as regiões polimórficas, determinamos o grau de conectividade de cada aminoácido e os tipos de interações químicas que estes efetuam com os demais aminoácidos na estrutura da proteína. A partir desses dados verificamos os impactos na estrutura da proteína com base nas mudanças das estruturas primárias. Além de verificaremos a origem dos polimorfismos em sítios que apresentam alta conectividade, por meio de uma análise filogenética. Após um extensivo processo de análise dos dados, determinamos que dentre as cinco proteínas estudadas, pertencentes a regiões mencionadas anteriormente, há padrões diferentes de distribuição dos polimorfismos ao longo da sequência. Notamos que as proteínas pertencentes a região NS1, possuem um número de trocas de aminoácidos menor quando comparadas as regiões NS3 e principalmente a NS5. Além disso, foi visto que aminoácidos com menor grau de conectividade possuem maior chance de sofrerem modificação e que essas mudanças tendem a ser entre aminoácidos com mesmas propriedades químicas.



Genes e genômica

ANÁLISE DE ENTROPIA DE QUATRO GENES HOUSEKEEPING DO GÊNERO XANTHOMONAS

Autores: José Bandeira do Nascimento Junior¹; Wilson José da Silva Junior¹; Antonio Roberto Gomes de Farias; Valdir de Queiroz Balbino¹;

E-mail para correspondência: bandeiraajr@gmail.com

Instituições: ¹Laboratório de Bioinformática e Biologia Evolutiva (LABBE), Departamento de Genética ? UFPE; ²Laboratório de Fitobacteriologia (LAFIBAC), Departamento de Agronomia ? UFRPE;

Palavras-chave: Marcador; Taxonomia; Bactéria

Apoio: CAPES, CNPq

Bactérias do gênero *Xanthomonas* destacam-se pela sua capacidade de infectar diversas

Variedades de plantas de importância econômica, causando perdas significativas pela limitação da produção. Apesar disso, a taxonomia do gênero ainda que bem definida, baseia-se em técnicas onerosas, como sequenciamento de genoma, análises comparativa de genomas e sequenciamento de várias regiões gênicas, elevando assim o custo e o tempo para a identificação das espécies. Para a identificação taxonômica de *Xanthomonas* sp. são utilizadas sequências parciais de genes encontrados em todas as espécies do gênero (housekeeping) que desempenham funções básicas para a sobrevivência da bactéria, como *gln*, *atpD*, *lrp*, *fusA*, *recA*, *groEL*, etc, *fusA*, entre outros, porém, os mais usados nas análises filogenéticas são o *dnaK*, *fyuA*, *rpoD*, *gyrB*. No entanto, nem um estudo verificou o grau de entropia dos genes em relação às regiões utilizadas atualmente. O cálculo de entropia de Shannon descreve as variações dos sítios de nucleotídeos nas regiões homólogas dos genes, permitindo assim selecionar as regiões que sejam mais adequadas para as análises filogenéticas. Este trabalho teve como objetivo avaliar a entropia dos genes housekeeping *dnaK*, *fyuA*, *rpoD* e *gyrB* e selecionar regiões mais adequadas para essas análises. 850 sequências de genomas disponíveis no NCBI foram utilizadas para montar um banco de dados local, de onde foram recuperadas 194 sequências completas de cada gene, constituindo quatro datasets. Cada dataset foi alinhado pelo MAFFT, esses alinhamentos foram tratados no GUIDANCE (MAFFT), por meio do algoritmo de alinhamento de sequências múltiplas (ASMs), e no GBLOCKS, que seleciona blocos conservados dentro do ASMs. Para visualização das sequências e avaliação dos parâmetros filogenéticos foi utilizado o MEGA7. A fórmula de Shannon foi utilizada para calcular a entropia para cada sítio nucleotídico e o DAMBE6 foi usado para realizar as análises. Para a seleção de regiões foi estipulado como critério o limite máximo de 0,20 de entropia, e tamanho mínimo de 400 de pares de base. Foi verificado que cada gene possui ao menos uma região de alta e baixa entropia, porém, para os genes *gyrB* e *rpoD* observou-se uma região que possui valores de entropia mais baixos do que as utilizadas atualmente. Para esses genes, as novas regiões selecionadas podem auxiliar os estudos de taxonomia molecular do gênero.



Genes e genômica

DETECÇÃO E ANÁLISE DAS FORMAS ÚNICAS RECOMBINANTES (URFs) BC DO HIV-1 CIRCULANTES NO ESTADO DO MARANHÃO &NDASH; BRASIL

Autores: Kleodoaldo Lima¹; Jessyca Kalynne Farias Rodrigues²; Élcio Leal; Andréa Vasconcelos Lobato de Lima; Marta de Oliveira Barreiros⁵;

E-mail para correspondência: kleodoaldo@gmail.com

Instituições: ¹Hospital das Clínicas - Universidade Federal de Pernambuco; ²Universidade Federal de Pernambuco; ³Instituto de Biotecnologia - Universidade Federal do Pará; ⁴Universidade Federal do Maranhão, Laboratório de Processamento da Informação Biológica, Programa de Pós-graduação em Rede e Nordeste de Biotecnologia; ⁵Universidade Federal do Maranhão, Laboratório de Processamento da Informação Biológica, Programa de Pós graduação em Engenharia de Eletricidade;

Palavras-chave: HIV; Formas Únicas Recombinantes (URFs); Epidemiologia molecular

Apoio: LABORATÓRIO CENTRAL DE SAÚDE PÚBLICA DO MARANHÃO

O HIV-1 apresenta uma grande diversidade genética no Brasil, com a circulação predominante dos subtipos B, F, C e seus recombinantes. No Nordeste brasileiro, além dos subtipos principais, foram detectados a circulação de recombinantes BF e BC, e a caracterização das Formas Recombinantes Circulantes (CRFs) 70, 71 e 90_BF. O objetivo deste estudo foi identificar e analisar o grau de recombinação genética de cepas BC do HIV-1 circulantes no Estado do Maranhão (Brasil). Os testes de genotipagem do HIV-1 foram realizados no Laboratório Central de Saúde Pública do Maranhão (LACEN-MA), totalizando 705 amostras testadas no período de janeiro de 2009 a dezembro de 2017. Os dados genéticos do HIV-1 foram obtidos através do kit de sequenciamento *Trugene HIV-1 genotyping assay* (Siemens Diagnostics), o qual amplificou e sequenciou a região da protease e parte da transcriptase reversa do vírus. A análise de recombinação intersubtipo foi realizada pelo SIMPLOT software. Detectou-se a presença de oito recombinantes BC do HIV-1 (1.13%), com padrão de recombinação genética similar à CRF31_BC primariamente identificada e com alta frequência na região Sul do Brasil. Todos os recombinantes apresentaram a região da protease caracterizada como derivada do subtipo C e a região da transcriptase reversa com um ponto de recombinação do subtipo B na posição aproximada de 2900 e 3200 pares de base em relação a cepa padrão do HIV-1 HXB2. Esta foi a primeira identificação da circulação de recombinantes BC do HIV-1 no estado do Maranhão. Além disso, sua análise de recombinação o caracterizou como um possível CRF 31_BC. Deste modo, denota-se a importância da epidemiologia molecular na caracterização das cepas circulantes do HIV-1 nas diferentes regiões, permitindo estudos mais aprofundados sobre a repercussão clínica, laboratorial, e de dispersão epidêmica e transmissibilidade dos diversos tipos virais.



Genes e genômica

CONSERVED MOTIFS IN MCRA GENE SEQUENCES OF METHANOBACTERIALES

Autores: Keyla Vitória Marques Xavier¹; Erick de Aquino Santos¹; Dailton Augusto Andrade ¹; Camila Souza Santos¹; Michely Correia Diniz²;

E-mail para correspondência: keyla_xavier@live.com

Instituições: ¹ Graduating Biological Sciences - UNIVASF; ² Doctor in Biotechnology. Adjunct teacher\ 's Biological Sciences College - UNIVASF. GEIS - Semi-Arid Integrated Studies Group;

Palavras-chave: MEMESUIT; Metil-coenzima M reductase ; methanogenesis

In anaerobic conditions, the material organic conversion into gases happens predominantly through methanogenesis, catalyzed by methanogenic Archaea. The enzyme Methyl-coenzyme M reductase (*mcrA* gene) acts directly in this process, exclusive to methanogen Archaea, and coding genes are evolutionarily conserved. The objective of this work was to search for motifs conserved in *mcrA* gene sequences of the order Methanobacteriales. The sequences were obtained at <https://www.ncbi.nlm.nih.gov/taxonomy>, using keyword Methanobacteria. The term nucleotide was selected and searched for the sequences within uncultured Methanobacteriales archaeon. An advanced search was made with the term *mcrA* Methanobacteriales. The nucleotide composition from the consensus sequence was verified using BioEdit v.7.2.6. The motifs were analyzed in MEME 4.12.0 (meme-suite.org), with the default settings, 3 different motifs and an average size of 50 bases. The lowest e-values were considered. Of the 59 sequences, 56 were selected for they presented complete data of their origin, 14 different origins: 14.28% were from marine sediments, 10.71% from reactors in general, and 75.01% from other environments (oil reservoir, rice field soil, mesophilic sludge). The means ranged from 325-761 nucleotides. The nucleotide composition was 16.58% A; 12.75% C; 17.22%; 11.37% T. On the conservation of the motifs, two had the best e-values, with 28 different dispositions, the smaller e-value the predominance of a site representing 28.57% of the motif in position 548-597. The motif with intermediate e-value in 2 sites, 28.57% in position 445-494 and 17.8% in position 171-220. The motif less significant e-value, presented 26 dispositions, with 3 sites, 26.9% in position 338-437 and 19.2% in positions 114-163 and 387-463. The motifs conservation corroborates the *mcrA* gene has a fundamental function in the evolution of these microorganisms.



Genes e genômica

DETECÇÃO DE CO-INFECÇÃO DE DIFERENTES GENÓTIPOS DO VÍRUS CHIKUNGUNYA, ATRAVÉS DE SEQUENCIAMENTO DE NOVA GERAÇÃO.

Autores: Laís Ceschini Machado¹; Mariana Carolina de Moraes Sobral¹; Tulio de Lima Campos¹; Mylena Ribeiro Pereira²; Maria de Fátima Pessoa Militão de Albuquerque¹; Clément Gilbert³; Rafael Freitas de Oliveira França¹; Gabriel Luz Wallau¹;

E-mail para correspondência: laisceschini@gmail.com

Instituições: ¹Instituto Aggeu Magalhães-FIOCRUZ/PE; ²Universidade Federal Rural de Pernambuco; ³CNRS Université Paris-Sud;

Palavras-chave: Co-circulação; genótipo Asiático/Caribenho; genótipo Africano

Apoio: CNPQ, CAPES

O vírus Chikungunya (CHIKV) é um arbovírus de RNA da família Togaviridae circulante em ciclos silvestres e urbanos. A febre Chikungunya (CHIKF) apresenta sintomas similares a outras arboviroses porém pode se manifestar de forma mais intensa e consequentemente mais debilitante. Esta arbovirose chegou no Brasil em 2014 através da introdução do genótipo Asiático e Caribenho pela região Norte e o genótipo ECSA Africano importado pela região Nordeste. Seguindo esta introdução ambos genótipos se estabeleceram no ciclo urbano causando surtos contínuos. Embora as sequências publicadas de genoma do CHIKV sejam abundantes em comparação com outros vírus da família Togaviridae, ainda se sabe pouco sobre as variantes genômicas virais e seu impacto na evolução do CHIKV e na progressão da doença após sua entrada nas Américas. A partir do sequenciamento de amostras clínicas com sintomas clássicos e mais severos da infecção do CHIKV, observou-se pela primeira vez a co-infecção dos genótipos Asiático/Caribenho e Africano na região nordeste do Brasil. A biblioteca de sequenciamento foi preparada utilizando o kit Nextera XT Library Prep (Illumina, San Diego, CA, EUA), onde foi sequenciado genomas do CHIKV diretamente de RNAs extraídos de amostras clínicas: 3 com sintomas clássicos e 5 sintomas mais severos, como mielite, miosite e neurite óptica. As sequências fastq foram trimadas utilizando o Trimmomatic 0.36 com parâmetros default. Posteriormente, as leituras foram mapeadas contra o genoma do KP164568 CHIKV (11812bp) usando leituras pareadas e não pareadas que passaram pelo filtro Trimmomatic utilizando o Bowtie 2 usando parâmetros default. A avaliação de SNP foi realizada com ferramentas samtools mpileup, vcf-annotate e snpEff software. Foi encontrado variantes de nucleotídeos (SNPs) em todas as amostras que apresentam manifestações neurológicas aguda, principalmente na região NS4, responsável pela codificação da RNA polimerase. A maior parte dos SNPs altamente suportados é sinônimo, porém SNPs não sinônimos também foram encontrados. Entretanto a grande maioria dos SNPs é exatamente correspondente ao genótipo Asiático e não a mutações novas. Nossos resultados mostraram que tanto os genótipos Asiático/Caribenho quanto o genótipo ECSA da África expandiram-se atingindo a co-circulação na região Nordeste, e posteriormente co-infecção. Além disso, co-infecção foi encontrada majoritariamente em amostras com manifestações mais severas da infecção por CHIKV.

Genes e genômica

AVALIAÇÃO CLÍNICA E MOLECULAR DOS POLIMORFISMOS NOS GENES MTHFR E VEGF NA SUSCEPTIBILIDADE À NEFROPATIA DIABÉTICA EM UMA POPULAÇÃO DO BRASIL CENTRAL

Autores: Letícia Ferreira Campos¹; Laura Raniere Borges dos Anjos¹; Elisangela Gomes da Silva³; Leandro do Prado Assunção²; Lorena Neves Ferreira¹; Gustavo Rodrigues Pedrino³; Rodrigo da Silva Santos⁴; Angela Adamski da Silva Reis¹;

E-mail para correspondência: leticiafecampos@gmail.com

Instituições: ¹Laboratório de Patologia Molecular, Departamento de Bioquímica e Biologia Molecular, Instituto de Ciências Biológicas (ICB), Universidade Federal de Goiás (UFG), Goiânia, GO, Brasil; ²Instituto de Patologia Tropical e Saúde Pública (IPTSP), Universidade Federal de Goiás (UFG), Goiânia, GO, Brasil; ³Instituto de Ciências Biológicas (ICB), Universidade Federal de Goiás (UFG), Goiânia, GO, Brasil; ⁴Departamento de Ciências da Natureza (LEdoC), Unidade Acadêmica Especial de Ciências Humanas (UAECH), Universidade Federal de Goiás (UFG), Goiás, GO, Brasil;

Palavras-chave: Metilenotetrafolato redutase; Fator de crescimento endotelial vascular; Nefropatia diabética

Apoio: Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

A nefropatia diabética (ND) é uma das principais complicações secundárias do *diabetes mellitus*, sendo caracterizada clinicamente por proteinúria, lesões glomerulares devido à pressão intraglomerular e declínio da taxa de filtração glomerular, levando o paciente à insuficiência renal, necessitando de diálise ou transplante. Vários fatores podem predispor o indivíduo diabético à nefropatia, tais como: hiperglicemia crônica, aumento da pressão arterial e determinantes genéticos. Dentre os determinantes genéticos, estudos têm destacado os polimorfismos nos genes metilenotetrahidrofolato redutase (*MTHFR*) e fator de crescimento endotelial vascular (*VEGF*). A enzima MTHFR desempenha papel fundamental no processamento de metionina, proteínas e outros compostos indispensáveis. O fator VEGF é uma proteína importante envolvida na sinalização que estimula a vasculogênese e angiogênese. Assim, o objetivo do estudo foi realizar a avaliação clínica e molecular dos polimorfismos de nucleotídeo único (SNP) nos genes *MTHFR C677T* e *VEGF 141 A → C* na susceptibilidade à nefropatia diabética em uma população do Brasil Central. Um total de 345 indivíduos foram genotipados para os polimorfismos descritos utilizando PCR-RFLP. O software RStudio, foi utilizado na análise estatística. A regressão logística múltipla não revelou interação entre os genes *MTHFR* e *VEGF*. No entanto, quando analisados individualmente, os resultados indicam que o polimorfismo de *MTHFR* aumenta o risco (OR: 2,37; p=0.004) do indivíduo diabético em desenvolver ND. Além disso, observamos que para o SNP no *VEGF*, os indivíduos diabéticos que apresentam genótipo CC apresentam 3,83 vezes mais chances de desenvolver ND (OR:3,83; p=0.000948). Ao analisar a associação de variáveis clínicas e polimorfismo no *MTHFR*, os resultados revelaram tendência estatística para alterações no IMC e na pressão arterial diastólica. Enquanto a associação de polimorfismo em *VEGF*, indicaram alterações na creatinina, taxa de filtração glomerular e pressão arterial diastólica. Estes resultados sugerem que os polimorfismos em *MTHFR* e *VEGF* podem conferir susceptibilidade à ND na população estudada. As alterações clínicas observadas indicam um mecanismo subjacente ao papel dos polimorfismos na disfunção renal. Outros estudos podem esclarecer a relação dos polimorfismos com a ND e auxiliar no tratamento dessas patologias.



Genes e genômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE ENZIMAS RELACIONADAS À BIOSÍNTESE DE LPG EM TRIPANOSSOMATÍDEOS POR FERRAMENTAS DE BIOINFORMÁTICA

Autores: Lucas Gentil Azevedo²; Luciane Amorim Santos^{1,2}; Pablo Ivan Pereira Ramos²;

E-mail para correspondência: lucasazevedo18116@gmail.com

Instituições: ¹Fundação Oswaldo Cruz (FIOCRUZ), Instituto Gonçalo Moniz (IGM), Laboratório de Enfermidades Infecciosas Transmitidas por Vetores (LEITV); ²Fundação Oswaldo Cruz (FIOCRUZ), Instituto Gonçalo Moniz (IGM), Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS);

Palavras-chave: Leishmania; lipofosfoglicano; Mineração de dados

Apoio: Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) e FIOCRUZ/IGM/CIDACS.

O lipofosfoglicano (LPG) é um constituinte da membrana de parasitas da família dos Tripanossomatídeos, sendo particularmente bem estudado no gênero *Leishmania*, agente etiológico da Leishmaniose. O LPG possui regiões de alta variação interespecíes e intraespecíes também sendo um importante fator modulador do sistema imunológico e virulência. As diferentes manifestações da leishmaniose são oriundas de características genéticas do parasito e de perfis da resposta imunológica do hospedeiro. Neste projeto, objetiva-se identificar e comparar sistematicamente a variabilidade dos genes relacionados à produção de LPG e suas relações filogenéticas. A metodologia foi dividida em duas etapas: 1) mineração de genes da biossíntese de LPG e 2) busca por genes ortólogos. Na primeira, foram utilizados o bancos de dados TriTrypDB e artigos que reportaram proteínas caracterizadas experimentalmente. Para identificar os genes ortólogos, foram realizadas buscas por similaridade das sequências encontradas contra os genomas de tripanossomatídeos completos. Algoritmos como BLASTp e PSI-BLAST permitiram encontrar sequências mais semelhantes buscando diretamente nas proteínas. O total de genomas/proteomas obtidos foi de 43 distribuídos entre a família de Tripanossomatídeos. Foram identificados 33 genes de *Leishmania major Friedlin* associados à biossíntese da âncora fosfolipídica, *core* e região de repetição. Observou-se alta conservação na sequência de genes importantes na montagem do LPG. Ao mesmo tempo, genes variavelmente presentes, a exemplo dos SCAs, foram identificados, os quais aparecem presentes em *L. infantum* e ausentes em *L. braziliensis*, representando, portanto, potenciais marcadores de infecção por *L. infantum*, além de levantar questões relacionadas à interação patógeno-vetor, uma vez que estes genes participam do desligamento do parasita do intestino médio do flebótomo. Outro achado foi a posição genômica dos genes pertencentes às famílias SCG/R/L e SCA nos genomas de *Crithidia fasciculata* e *Leptomonas pyrrhocoris*, todos presentes no cromossomo 16 e 27 respectivamente, sendo que a maioria dos representantes desses genes nas outras espécies estão em cromossomos diferentes, o que sugere uma evolução distinta em relação aos representantes do gênero *Leishmania*. Estes resultados ressaltam o poder que as análises genômicas comparativas possuem na melhor compreensão dos aspectos moleculares relacionados ao parasito.



Genes e genômica

CARACTERIZAÇÃO DE CÉLULAS VIÁVEIS MAS NÃO CULTIVÁVEIS EM *XANTHOMONAS CITRI* PV. *ANACARDII*

Autores: Lucas Pontes de Lucena¹; Flávia Figueira Aburjaile²; Ana Maria Benko Iseppon²; Marco Aurélio Siqueira da Gama¹;

E-mail para correspondência: lucasp194@yahoo.com.br

Instituições: ¹Universidade Federal Rural de Pernambuco; ²Universidade Federal de Pernambuco;

Palavras-chave: Fitopatógeno; Mancha angular; VBNC

Apoio: CAPES, CNPq e FACEPE

O cajueiro (*Anacardium occidentale* L.) é uma planta originária do Brasil, maior produtor mundial da cultura, e sua produção concentra-se na região nordeste do país. Dentre as doenças de natureza bacteriana na cultura do cajueiro, duas destacam-se no Nordeste brasileiro: a mancha angular bacteriana e a mancha de *Xanthomonas*, causadas por *Xanthomonas citri* pv. *anacardii*. A sobrevivência desta bactéria se dá principalmente em lesões foliares, e o manejo da doença é realizado principalmente através de medidas de exclusão, como uso de material propagativo sadio e de quebra-ventos, ou de erradicação, como a destruição de restos culturais e aplicação de fungicidas cúpricos. Porém, *X. citri* pv. *anacardii* bem como outras bactérias apresentam a capacidade de sobrevivência de células viáveis, mas não cultiváveis (VBNC - *Viable but nonculturable*), que se caracteriza como um período de baixa atividade metabólica que permite às bactérias sobreviverem na presença dos fatores de limitantes. Nesse sentido, o objetivo do presente estudo é caracterizar a ocorrência de células de *X. citri* pv. *anacardii* em estado VBNC nos isolados IBSBF 2579 e CCRMTAQ18 através da comparação simultânea das curvas de crescimento de células viáveis e de células totais durante 12 horas. A quantidade de células viáveis será determinada através do número de unidades formadoras de colônias/mL. O número de células totais será obtido a partir da medição de densidade ótica da suspensão bacteriana dos isolados em meio NYD, por um fotolorímetro com absorvância ajustada a 570 nm. Também será realizada a detecção de genes associados ao mecanismo com o auxílio do programa Artemis e validação destes genes por PCR quantitativa. Neste momento do estudo, os genes mais promissores de *X. citri* pv. *anacardii* comuns a outros fitopatógenos em VBNC foram os codificadores das proteínas ligadas à luxR, comumente relacionada ao *quorum-sensing*, e os genes de fatores transcricionais da LysR, que regulam a expressão de diversos genes ligados a funções como virulência e motilidade, ambos encontrados em *Pseudomonas syringae* pv. *syringae*.

Genes e genômica

CARACTERIZAÇÃO FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA H⁺ATPASE DE MEMBRANA PLASMÁTICA EM ESPÉCIES DA ORDEM POALES.

Autores: Lyndefânia Melo de Sousa¹; Clesivan Pereira dos Santos¹; Thais Andrade Germano¹; Moaciria de Souza Lemos¹; Rafael de Souza Miranda²;

E-mail para correspondência: lyndefania@hotmail.com

Instituições: ¹Universidade Federal do Ceará; ²Universidade Federal do Piauí;

Palavras-chave: Anotação gênica; distribuição filogenética; classificação

Apoio: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) Universidade Federal do Ceará (UFC)

A H⁺ATPase são proteínas integrais da membrana plasmática que apresentam uma massa de cerca de 100KDa e têm a capacidade de utilizar a energia química da hidrólise de ATP para expulsar os prótons para o ambiente extracelular através do transporte ativo. Dessa forma, as H⁺ATPase atuam na manutenção da homeostase iônica celular e no transporte de solutos. Esses transportadores também estão envolvidas na aclimação das plantas as condições de estresse, bem como na absorção de nutrientes, regulação do pH intracelular, abertura estomática e crescimento celular. O presente trabalho objetivou identificar, caracterizar e analisar a distribuição filogenética da família multigênica H⁺ATPase de membrana plasmática de espécies da ordem Poales. Para este fim, os membros dessa família multigênica foram identificados e anotados, manualmente, nas espécies *Aegilops tauschii*; *Brachypodium distachyon*; *Oryza sativa*; *Panicum hallii*; *Leersia perrieri*; *Setaria viridis*; *Zea mays*; *Zoysia japonica*; *Ananas comosus* através de buscas em bancos de genomas disponíveis no NCBI e *Phytozome*, utilizando a ferramenta BLAST. Em seguida, as sequências foram alinhadas através do algoritmo MUSCLE e agrupadas pelo método *Neighbor joining*, a partir de 1000 replicatas, através do programa MEGA 7.0. Os resultados revelaram um total de 9 a 12 genes codificando H⁺ATPase em Poales. O número de éxons variou de 3 a 21. As janelas de leituras aberta (ORFs) apresentaram entre 2.934 e 5.074 nucleotídeos, enquanto o número de aminoácidos variou de 857 a 971. A análise filogenética distribuiu as sequências em quatro clados distintos identificados como clado I-IV. Posteriormente, o clado I, por apresentar apenas uma sequência representativa de cada espécie foi nomeada de (H⁺ATPase1). Porém, aquelas dos clados II, III e IV foram nomeadas (H⁺ATPase2, 2a ou 2a1, 2b ou 2b1 e 2c), (H⁺ATPase3, 3a ou 3a1, 3b, ou 3b1, 3c) e (H⁺ATPase4, 4a ou 4a1, 4b), respectivamente, por apresentarem duas ou mais sequências da mesma espécie. Esses resultados fornecem suporte para a realização de estudos (em andamento) de expressão gênica em espécies como *O. sativa* e *Z. mays* no intuito de esclarecer sua (s) relevância (s) nos mecanismos de aclimação as condições de estresses.



Genes e genômica

ANÁLISE DO SATELITOMA EM PASSIFLORA L. (PASSIFLORACEAE)

Autores: Mariela Analía Sader¹; Magdalena Vaio Scvortzoff²; Luiz Augusto Cauz dos Santos³; Maria Lúcia Carneiro Vieira³; Marcelo Carnier Dornelas⁴; Nataniel Franklin de Melo⁵; Andrea Pedrosa-Harand¹;

E-mail para correspondência: mariela_sader@yahoo.com.ar

Instituições: ¹Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Universidade Federal de Pernambuco, Recife, Brasil.; ²Laboratorio de Evolución y Domesticación de Plantas, Facultad de Agronomía, Universidad de la República, Uruguay; ³Escola Superior de Agricultura Luiz de Queiroz, Departamento de Genética, Universidade de São Paulo, Piracicaba, Brasil; ⁴Instituto de Biologia, Departamento de Fisiologia Vegetal, Universidade Estadual de Campinas; ⁵Laboratório de Biotecnologia, Embrapa Semiárido, Empresa Brasileira de Pesquisa Agropecuária;

Palavras-chave: DNA satélite; Tamanho genômico; SatMiner

Apoio: FACEPE

O DNA nuclear é constituído por sequências codificadoras de cópia única assim como por várias classes de sequências de DNA repetitivo. O DNA repetitivo forma uma grande parte dos genomas de plantas, podendo constituir aproximadamente 80% do mesmo, e pode se diferenciar em sequências dispostas em *tandem*, as sequências de DNA satélite (DNAsat), ou as dispersas no genoma, os elementos transponíveis. A coleção de diferentes famílias de DNAsat em um genoma é chamada de satelitoma. As espécies do gênero *Passiflora* apresentam uma ampla variação no tamanho, número cromossômico e conteúdo de DNA. Nenhuma sequência de DNAsat foi descrita até o momento no gênero. Sendo assim, o objetivo deste trabalho foi determinar o satelitoma de três espécies do gênero para ser utilizadas como marcadores cromossômicos no estudo da evolução cromossômica do gênero: *P. quadrangularis* ($2n = 18$; $2C = 2.680$ pg) conhecida como maracujá gigante, com o maior genoma conhecido; *P. cincinnata*, o maracujá do mato ($2n = 18$; 2.202 pg); e *P. organensis* ($2n = 12$; 0.404 pg), com o menor genoma. Para isso utilizamos uma análise de cluster do pipeline RepeatExplorer combinado com a estratégia do Satminer. As sequências selecionadas foram analisadas no programa Geneious v.7.1.9 e classificadas de acordo com homologia de sequência e tamanho da unidade de repetição. Encontramos 45 DNAsats distintos nas três espécies, sendo seis para *P. quadrangularis*, dois para *P. cincinnata* e trinta e sete para *P. organensis*. Os satélites variaram quanto ao tamanho da unidade de repetição (entre 52 e 3100 pb) e em abundância estimada no genoma. Além de maior diversidade, *P. organensis* também apresentou a maior proporção de DNAsat, mostrando não ser o satelitoma das espécies a fração responsável pelos distintos tamanhos genômicos. Para todas as sequências classificadas foram desenhados *primers* para posterior amplificação de sondas para ser utilizadas nas futuras análises de evolução cromossômica.



Genes e genômica

ANÁLISE COMPARATIVA DE MONTAGEM GENÔMICA DE PROCARIOTOS USANDO DIFERENTES PARÂMETROS DO PATRIC®

Autores: Mailane dos Anjos Silva¹; Joel Augusto Moura Porto¹; Eric Roberto Guimarães Rocha Aguiar¹;

E-mail para correspondência: anjosmailane@gmail.com

Instituições: ¹Universidade Federal da Bahia;

Palavras-chave: montagem; parâmetros; PATRIC

Apoio: Universidade Federal da Bahia, Instituto de Ciências da Saúde, Departamento de Biotecnologia

Em Bioinformática, há uma variedade de programas que pode ser usada para montar as leituras que são produzidas a partir de máquinas de sequenciamento em contigs, mas elas podem exigir uma capacidade de programação avançada que os biólogos de pesquisa às vezes não têm. Para atender a essa necessidade, o PATRIC permite que os pesquisadores montem leituras curtas que são simples ou emparelhadas tipicamente em máquinas Illumina, e também leituras longas de máquinas PacBio ou Nanopore. Este trabalho tem como objetivo principal uma análise comparativa entre os diferentes tipos de montagem presentes no PATRIC. Importante que destaquemos que, dentre eles, excluímos o parâmetro *plasmid*, por entendermos que o seu objetivo foge aos interesses de montagem genômica de procariotos. Um outro ponto que motiva a realização deste trabalho é a desmistificação de que há uma qualidade quase que inquestionável relacionada ao parâmetro de montagem *Auto*. Notamos, ao comparar os diferentes parâmetros, que o *Auto* não apresenta dos melhores *assemblies* gerados pelo PATRIC. As consequências disso vão para muito além de somente uma montagem genômica, mas também de conclusões futuras equivocadas com base em uma montagem não tão bem estruturada. De acordo com a literatura fornecida pelo próprio PATRIC, a estratégia de montagem automática executa BayesHammer em leituras curtas, seguidas por três estratégias de montagem que incluem Velvet, IDBA e Spades. A estratégia de montagem Fast executa o MEGAHIT e o Velvet. Os usuários também podem escolher a estratégia Full Spades, que executa o BayesHammer seguido pelo Spades. Escolhendo o Kiki, roda o montador Kiki um script interno. As leituras do Illumina MiSeq devem ser montadas usando o miseq, que executa o Velvet. Um outro parâmetro usado neste trabalho é o Smart, que apresenta as montagens Gam_ng, Velvet e Spades. Utilizamos de um total de 20 procariotos, dentre eles, bactérias gram positivas e negativas (10 de cada) a fim de notarmos as diferenças de montagem entre elas. Foi observada também a influência da quantidade de repetições GC nos diferentes organismos, uma vez que a literatura garante uma significativa interferência nos processos e qualidades de montagem, o que também foi considerado em nosso trabalho. As nossas últimas análises comparativas apontam para o Fast (Spades) e o Smart (gam_ng) como os parâmetros de melhor qualidade de montagem genômica, conclusões parciais até o momento de feitura deste trabalho.



Genes e genômica

UTILIZAÇÃO DO BANCO DE DADOS GTEx PARA TRAÇAR UM PERFIL DE CO-EXPRESSION GÊNICA DAS VIAS BER E NER EM TECIDOS CEREBRAIS SADIOS

Autores: Maria Eduarda Andrade Lima Martins Lopes¹; Danilo Lopes Martins¹; Joaliton Luan Pereira Ferreira¹; Jorge Estefano Santana de Souza¹; Lucymara Fassarella Agnez Lima¹;

E-mail para correspondência: m.eduarda.al.martins@hotmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: Reparo de DNA; GTEx; Co-expressão gênica

Apoio: CAPES

A molécula de DNA está sujeita a uma variedade de lesões. Dentre estas, destacam-se as espécies reativas de oxigênio e/ou nitrogênio que têm sido associada ao desenvolvimento de algumas patologias neurodegenerativas, como Alzheimer, Parkinson, Esclerose Lateral Amiotrófica e Doença de Huntington. O sistema de reparo de DNA atua na detecção e correção de danos através de várias vias. A via de reparo por excisão de bases (BER) corrige bases incorretas ou danificadas, atua fortemente em neurônios e mitocôndrias. O reparo por excisão de nucleotídeos (NER) age contra danos que causam distorções na dupla hélice. Além disso, estudos indicam que BER e NER também atuam como reguladores transcricionais, interagindo entre si e com vários fatores de transcrição. Detectar os principais genes envolvidos nesse processo é de extrema importância para comparar com condições patológicas diversas. Este trabalho buscou estabelecer um perfil de co-expressão gênica em tecidos cerebrais saudáveis. Foram utilizados 54 genes para 07 tecidos cerebrais do banco GTEx. Mensuramos a correlação de expressão pelo Coeficiente de Correlação Linear de Pearson por meio de um script na linguagem python e em seguida analisamos os dados com o RStudio. Foram observadas correlações significativas entre os genes para todos os tecidos cerebrais analisados, o que indica uma co-expressão relevante entre os genes estudados. O valor de co-expressão se mostrou diferente para cada tecido, possibilitando sua comparação em relação a patologias, uma vez que cada uma atinge áreas cerebrais específicas. Apesar de BER ser descrita como atuante em neurônios, os genes da via NER se mostraram fortemente co-expressos nos tecidos cerebrais tanto em relação aos genes de BER quanto a reguladores transcricionais, demonstrando uma interação direta entre essas duas vias, indicando regulação mútua entre si e entre vários fatores de transcrição cujas funções estão associadas a diversos eventos celulares. Detectamos aproximadamente 30 genes que se mostraram comumente expressos nos diferentes tecidos cerebrais, o que nos possibilitou traçar um perfil de co-expressão em tecidos saudáveis. Este trabalho visa dar continuidade às suas análises comparando os seus resultados com diversos bancos de dados relacionados a patologias neurológicas, contribuindo para esclarecer os mecanismos moleculares de várias doenças e possíveis formas de diagnósticos e tratamentos.



Genes e genômica

FAMÍLIA DE GENES ACBP EM *TIBOLIUM CASTANEUM*: EXPRESSÃO, CARACTERIZAÇÃO E FUNÇÃO DE TCACBP-6

Autores: Mariana de Macedo Costa Batalha¹; Camilla Camerinno Santana Davino Freire¹; Kátia Capim Gondim²; David Majerowicz²; Camila Braga Dornelas¹;

E-mail para correspondência: marianamcbatalha@gmail.com

Instituições: ¹Universidade Federal de Alagoas ; ²Universidade Federal do Rio de Janeiro ;

Palavras-chave: transporte da acil-CoA; inseto; silenciamento

Apoio: CNPq; FAPEAL

Proteínas ligadoras de acil-coA (ACBP) constitui uma família que regula a disponibilidade desses lipídeos para os diversos processos metabólicos celulares. Pelo seu o alto grau de conservação por todas as espécies eucarióticas, e o fato desta proteína ser expressa na maioria das células e tecidos, sugerem que suas funções estão associadas a um ou mais processos básicos e comuns a todas as células. No entanto, seu papel preciso no metabolismo de lipídeo ainda não é completamente conhecido. O inseto *Tribolium castaneum*, modelo deste estudo, está em crescente aplicação em pesquisas de análise genética funcional de questões biológicas básicas, assim como a *Drosophila*. Portanto, o objetivo desse estudo é investigar sequências homólogas de ACBPs no *T. castaneum* e avaliar a expressão do gene TcACBP6 e compressão da função por silenciamento através de RNA interferente. A pesquisa de sequências homólogas e análise bioinformática foi realizada utilizando a sequência consenso do domínio ACBP do banco de dados do Pfam e investigação pelo Ensembl Metazoa com pesquisa de dados no bioMart. Foi prosseguida análise com o código da sequência primária da ACB1p de *S. cerevisiae* (AAA34384) com o algoritmo tBlastn no genoma do *T. castaneum*. A localização celular foi predita utilizando o programa PSORT II; pesquisa para domínios transmembrana foi confirmada nos 4 programas de análise (TMHMM 2.0, SOSUI, TMpred e HMMTop); presença de peptídeo sinal e predição de sítios de fosforilação pelo programa NetPhos 2.0 Server. Estimou - se os valores de massa molecular e seus pontos isoelétricos no programa ProtParam. A árvore filogenética foi construída pelo método da Máxima Verossimilhança com bootstrapping de 10.000 réplicas usando o software MEGA 7.0. A expressão relativa foi medida por RT-qPCR com os dados normalizados utilizando a proteína ribossomal 3 como gene de controle endógeno. O efeito do silenciamento em TcACBP6 foi confirmado por RT-qPCR, e o efeito, avaliado na β -oxidação e no conteúdo de triacilglicerol. Foram constatadas 6 sequências homólogas à ACBP no genoma do *T. castaneum* e evidenciadas diferenças significativas entre elas na predição por ferramentas bioinformática, evidenciando possíveis papéis distintos no endereçamento e sinalização celular por acil-coA graxo nesse modelo. O estudo do gene TcACBP6 permitiu evidenciar uma possível função no direcionamento de ácidos graxos para beta oxidação através do knockout de função por RNAi.



Genes e genômica

CYTB GENE ANALYSIS OF ANT GENUS ECTATOMMA

Autores: Raisia Maria Souza Rosas^{1,2}; Lorena Maria Souza Rosas¹; Benoit Jean Bernard Jahyny^{1,2}; Michely Correia Diniz^{1,2};

E-mail para correspondência: michely.diniz@univasf.edu.br

Instituições: ¹UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO; ²GEIS- GRUPO DE ESTUDOS INTEGRADOS DO SEMIÁRIDO;

Palavras-chave: cytochrome b; molecular phylogeny ; bioindicators

Apoio: UNIVASF

The ants form approximately half of the edaphic animal biomass. These small arthropods are little studied genetically and are currently considered important bioindicators of degraded environments. The ant genome provides information about its biology, allowing investigating its social organization at a molecular level, including its ecological domain. Bioinformatics tools can infer evolutionary relations, sequence characterization and homology analysis of organisms through sequenced genes. Mitochondrial DNA (mtDNA), in particular the cytochrome b (Cytb) gene, can provide important information to estimate the species divergence time and be useful in phylogenetic studies. This work aimed to analyze Cytb genes sequences in ants genus *Ectatomma* and infer their molecular phylogeny through bioinformatics tools. Sequences were selected from the NCBI (<https://www.ncbi.nlm.nih.gov/>). Alignment and consensus sequence was performed using BioEdit 7.2.5 software. The phylogenetic inference and mean sequence divergence were computed by MEGA 7.0. The phylogeny obtained used the Maximum Parsimony method, with the Hasegawa-Kishino-Yano (+G) model. The trees were replicated in a 1000 bootstrap. The 110 sequences available had an average of 712 bp, with 42.4% T; 15.2% C; 34.8% A and 7.6% G. The divergence index between the sequences was 0.066, considered low, there was probably a smaller number of nucleotide substitutions throughout the evolutionary process. The topology involved 33 nucleotide sequences and consensus sequences of the species *E. ruidum*, *E. neotropics*, *E. Oxaca 1*, *E. Oaxaca 2* isolate. Two large groups were formed, both strong supported (90%), which corroborate previous studies. However, other clusters remain sensitive to analytical methods probably due to the similarity between different species sequences, or due to data saturation and rapid diversification.

Genes e genômica

ASSOCIAÇÃO DO POLIMORFISMO MTHFR677C>T NA SUSCEPTIBILIDADE À LEUCEMIA LINFOCÍTICA CRÔNICA

Autores: Nayane Soares de Lima¹; Débora Acyole Rodrigues²; Elisangela Gomes da Silva²; Keila Correia de Alcântara²; Gustavo Rodrigues Pedrino²; Rodrigo da Silva Santos²; Angela Adamski da Silva Reis²;

E-mail para correspondência: nayanes00@gmail.com

Instituições: ¹Faculdade Araguaia; ²Universidade Federal de Goiás;

Palavras-chave: gene MTHFR; LLC; Polimorfismo genético

Apoio: Universidade Federal de Goiás (UFG)

A Leucemia Linfocítica Crônica (LLC) é uma doença crônica, hematológica linfoproliferativa, conhecida pelo acúmulo da população de linfócitos B de origem monoclonais na medula óssea, no sangue e tecidos linfáticos. É uma das formas de neoplasias maligna mais prevalente em indivíduos adultos relacionada com o avanço da idade. A LLC é uma doença predominantemente do idoso; a idade mediana ao diagnóstico é de 64-70 anos, sendo rara a sua ocorrência em indivíduos <30 anos. Um aumento exponencial na incidência da LLC é observado com o avançar da idade, para ambos os sexos, sendo a incidência em indivíduos >65 anos de 20,6 casos/100.000/ano enquanto em <65 anos é de apenas 1,3 casos /100.000. Há um predomínio no sexo masculino. A etiologia da LLC é ainda desconhecida; alguns fatores ambientais são descritos, porém as associações não são consistentes. Estudos relatam que eventos primários nesse tipo de neoplasia podem estar associados com as alterações genéticas e epigenéticas, como a metilação do DNA e os polimorfismos genéticos, os quais aumentam o risco do desenvolvimento de LLC. Indivíduos com alterações na atividade da enzima MTHFR, decorrente do polimorfismo no gene *MTHFR* 677C>T, podem estar mais susceptíveis às transformações malignas promovidas pela redução da atividade enzimática e da alteração na distribuição do folato intracelular. Assim, o presente estudo teve por objetivo avaliar o papel do polimorfismo C677T do gene *MTHFR* na susceptibilidade genética à LLC. Foi realizado um estudo caso-controle de base hospitalar com 68 pacientes com LLC e 145 indivíduos para o grupo controle. A genotipagem para o gene *MTHFR* 677C>T foram determinadas por PCR-RFLP. Observou-se diferença estatística para a distribuição por sexo, idade e tabagismo, pelo teste do X² (p>0,05), sendo maior percentual para o sexo masculino e tabagista. Verificamos que os indivíduos tabagistas portadores de genótipo TT apresentam 2,26X mais chances de desenvolver LLC (OD:2,26/IC:1,165-4.401/p=0,02). As frequências genotípicas para o sexo feminino e masculino foram CC (66,3,7%), CT (43,8%), TT (42,9%) e CC (33,7%), CT (56,2%), TT (57,1%), respectivamente. Tais resultados demonstram que o sexo masculino portador do genótipo TT possui mais risco de desenvolvimento de LLC (OD:2,52/IC:1,437-4.446/p=0,01). Nossos resultados demonstram que o risco de LLC está associado com o sexo e tabagismo.



Genes e genômica

AGRUPAMENTOS DE DADOS CLÍNICOS E GENÔMICOS DE CÂNCER ATRAVÉS DE MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN

Autores: Patrick Cesar Alves Terrematte¹; Daniel Sabino Amorim de Araújo¹; Adrião Duarte Dória Neto¹; Beatriz Stranky^{1,2}; Eric David Rebouças de Souza²; Affonso de Faria²; Fransueldo Florencio Ribeiro Do Ó²;

E-mail para correspondência: patrick.terrematte@ufersa.edu.br

Instituições: ¹Programa de Pós-graduação em Bioinformática (PPgBIO) - Instituto Metrópole Digital (IMD) - UFRN;

²Departamento de Engenharia Biomédica (DEB) - UFRN;

Palavras-chave: Clustering; Genômica; Cancer

A análise de dados heterogêneos de Câncer tem sido explorada através de algoritmos de agrupamentos afim de classificar subtipos de Câncer. No presente trabalho, analisamos dados de Câncer de próstata através de uma versão supervisionada das Redes Neurais de mapas auto organizáveis de Kohonen (SOM), com o objetivo de apresentar associações entre dados clínicos e genômicos, e analisar os resultados através de vias metabólicas (KEGG Cancer pathway). Os dados clínicos e genômicos utilizados são os dados públicos do TCGA, disponibilizados pelo Genomic Data Commons e cBioPortal. Para treinar a rede SOM utilizamos 70% das amostras para classificar 30% do total de 187 amostras do TCGA-PRAD com seus respectivos 8 subtipos de Câncer de próstata (ERG, ETV1/4, FLI1, SPOP, FOXA1, IDH1 e outros) associados aos dados de Idade, nível de PSA, soma de Gleason, Celularidade e Fração de Genoma Alterado. Como objetivo de reduzir a dimensionalidade, nós realizamos uma análise de componentes principais (PCA), e verificamos que através de três componentes principais é possível explicar apenas 74,7% da variação dos dados, portanto treinamos a rede SOM sem a PCA. Como resultado, obtivemos um coeficiente de concordância de Kappa de 92,6% para a predição 30% das amostras dos subtipos de Câncer de próstata. Notamos também que os resultados através da rede SOM permitem uma melhor visualização e identificação da contribuição de cada variável dos dados clínicos e moleculares para o agrupamento dos dados, quando comparados ao agrupamento integrativo obtido a partir do iCluster.



Genes e genômica

DESENVOLVIMENTO DE MICROSSATÉLITES PLASTIDIAIS POLIMÓRFICOS PARA ANÁLISES GENÉTICAS DE *CENOSTIGMA MICROPHYLLUM* (LEGUMINOSAE) EM ÁREAS DA CAATINGA

Autores: Paulo Aecyo Francisco da Silva¹; Tiago Esposito Oliveira Melo¹; André Seco Marques da Silva²; Andrea Pedrosa Harand¹;

E-mail para correspondência: pauloaecyo_1997@hotmail.com

Instituições: ¹Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco (UFPE); ²Laboratório de Recursos Genéticos, Instituto de Ciências Biológicas e Saúde, Universidade Federal de Alagoas (UFAL);

Palavras-chave: catingueira; cpSSR; marcador molecular

Apoio: FACEPE, CAPES e CNPq

Cenostigma microphyllum (Mart. ex G. Don) Gagnon & G. P. Lewis, conhecida como catingueira falsa, possui hábito arbóreo, sendo polinizada especialmente por insetos e dispersa balisticamente. É uma espécie endêmica da Caatinga, amplamente distribuída em toda a região, estando sujeita diretamente às perturbações antrópicas que ameaçam esse ambiente. Uma ferramenta importante para estudos de genética de populações num contexto conservacionista são os marcadores moleculares, especialmente os microssatélites (SSR) por serem codominantes, altamente polimórficos e facilmente encontrados em todos os genomas. Desta forma, este trabalho teve como objetivo desenvolver novos locos de microssatélites plastidiais (cpSSR) e testar a amplificação e o polimorfismo destes para que possam ser utilizados em análises genéticas futuras, como a influência das perturbações antrópicas na diversidade genética da Caatinga. Para isso, o DNA genômico de *C. microphyllum* foi sequenciado em plataforma Illumina com baixa cobertura e utilizado o programa NOVOPlasty v. 2.6 para a montagem do genoma plastidial (plastoma) completo. Assim, foi feita a procura por regiões de cpSSR no plastoma montado da espécie. Os *primers* foram desenhados com o programa PRIMER3PLUS e testados no OligoAnalyser, ambos online. O DNA genômico foi extraído utilizando o método CTAB, a partir de folhas de oito indivíduos, sendo cinco deles de áreas preservadas e três de áreas perturbadas da caatinga. As regiões foram amplificadas com temperatura de anelamento variando de 58°C a 60°C. O produto da PCR foi marcado indiretamente com o auxílio da cauda M13 e foram genotipados utilizando o ABI 3500 GeneticAnalyzer (Life). Para procura de polimorfismos, foi utilizado o software GeneMarker V.2.2.0. Foram desenvolvidos e testados 10 locos cpSSR, apresentando em sua maioria mononucleotídeos de repetição A ou T, com exceção do loco CmCPSSR6 que é um dinucleotídeo de repetição (AT)₅. Quatro locos estão localizados nos genes *rpoC2*, *trnG – UCC*, *trnK – UUU* e *ycf3*; e seis estão localizados em regiões intergênicas do cloroplasto. Todos os locos amplificaram regiões de tamanho esperado, variando de 125 pb a 429 pb. Destes, seis locos apresentaram polimorfismos de dois (CmCPSSR4, CmCPSSR7, CmCPSSR8, CmCPSSR15 e CmCPSSR16) a três alelos (CmCPSSR12). Sendo assim, esses locos podem ser utilizados para entender como os impactos antrópicos podem estar influenciando a diversidade e estruturação genética da *Cenostigma microphyllum*.



Genes e genômica

ANÁLISE DO PADRÃO DE EXPRESSÃO DOS GENES RELACIONADOS ÀS SÍNDROMES PSEUDO TORCH

Autores: Paula Vitória Macêdo de Barros¹; João Ricardo Mendes de Oliveira¹;

E-mail para correspondência: paulamacedo.vb@gmail.com

Instituições: ¹Laboratório de Imunopatologia Keizo Asami -LIKA e Universidade Federal de Pernambuco- UFPE ;

Palavras-chave: pseudo TORCH; Allen Brain Atlas; neurogenética

The pseudo-TORCH syndromes form a group of congenital genetic diseases that clinically mimic characteristics of the TORCH intrauterine infection without, though, show evidence of infection. Although they have heterogenous causes, they present clinical manifestations as microcephaly, intracranial calcifications, restriction of uterine growth, hepatosplenomegaly and thrombocytopenia. The intracranial calcifications, as well as microcephaly, are manifestations shared by many diseases that impair the central nervous system. In this context, the main objective of this work is analyze expression patterns of the genes related to these congenital conditions in the most injured encephalic structures so that be possible elucidate physiopathologic mechanisms involved in the alterations of the central nervous system which are observed in these syndromes as well as in another pathologic conditions. To this work, the BrainSpan Atlas of the Developing Human Brain, one of the atlases provided by the Allen Institute, was the main resource used to measure and investigate the patterns of the genetic expression in neuroanatomic structures involved in the syndromes. Therefore, the microarray data, available for more than 300 structures obtained by laser microdissection, were used to observe the expression of the 8 genes related in three different development stages. The quantitative genetic expression analysis was measured from the z-score data, also provided by the Allen Brain Atlas. The analysis of these genes (JAM3, TREX1, OCLN, RNASEH2A, RNASEH2B, RNASEH2C, PDGFR β e PCDH19) allowed establish genetic expression profiles that may indicate presumptive physiopathologic mechanisms shared by them. Of the 8 genes, 6 presented (at least one of its probes) a high expression in the ventricular and subventricular zones, as well as in the ganglionic eminences, embryonic regions that play a key role in the human brain development and which are intrinsically related to areas more affected by the calcifications, like the periventricular regions and the basal nuclei. Of these 6 genes, 4 (RNASEH2A, RNASEH2B, RNASEH2C e TREX1) are involved in the nucleic acids metabolism and 2 of them (JAM3 e PDGFR β) participate of the formation and maintenance of the blood brain barrier.



Genes e genômica

COMPARAÇÃO DE MONTADORES METAGENÔMICOS UTILIZANDO AMOSTRAS MICROBIANAS SIMULADAS

Autores: Pedro Henrique Dornele Machado Rocha¹; Renato Renison Moreira Oliveira¹; Regiane Silva Kawasaki Frances¹;

E-mail para correspondência: pedrohenriquedornele@gmail.com

Instituições: ¹Universidade Federal do Pará (UFPA);

Palavras-chave: Metagenoma; Comparação; Montadores

Apoio: Universidade Federal do Pará (UFPA), Laboratório de Bioinformática e Computação de Alto Desempenho (LaBioCAD), Instituto Tecnológico Vale (ITV), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

A metagenômica é um campo crescente de pesquisa que visa estudar populações de organismos não cultivados para entender a verdadeira diversidade dos mesmos, as suas funções, a cooperação e evolução, em ambientes como solo, água, restos antigos de animais, entre outros. Por conta deste crescimento, várias ferramentas de montagem voltadas para a metagenômica foram criadas, como IDBA-UD, (Meta-) SPAdes e MEGAHIT. Para verificar qual ferramenta possui um resultado melhor, foi feito um teste comparativo utilizando estatísticas como NGA50, tamanho do maior contig, erros de montagem, quantidade de contigs montados e maior alinhamento para averiguar qual dos montadores citados obteve a melhor performance ao se montar um metagenoma controlado (com sequenciamento simulado). Para realizar a comparação dos montadores, foram utilizadas duas amostras metagenômicas simuladas com coberturas diferentes, uma contendo 10 espécies de bactérias e outra contendo 100 espécies. O tratamento de qualidade das reads foi feito com a ferramenta PRINSEQ e a comparação dos resultados das montagens feitas pelo IDBA-UD, Meta-Spades e MEGAHIT foi feita com a ferramenta MetaQUAST. Para as montagens de maior cobertura (10 espécies), o Montador IDBA-UD obteve resultados melhores quando comparados com os outros montadores, apresentando um contig maior, melhores NGA50 e maior quantidade de contigs gerados, mesmo possuindo uma quantidade maior de erros de montagem. Na montagem da amostra de menor cobertura (100 espécies), o Montador MEGAHIT apresentou resultados melhores, com menor quantidade de erros de montagem, maior contig e maior alinhamento, enquanto o montador IDBA-UD obteve resultados inferiores. Já o montador (Meta-) SPAdes, não retornou resultados que se sobressaem aos outros, sendo inclusive o pior montador para a mostra de maior cobertura. Observando os resultados obtidos pelo MetaQUAST, o montador IDBA-UD se mostra ideal para montagem com metagenomas de alta cobertura, enquanto o MEGAHIT se mostra melhor em montagens de Metagenomas de baixa cobertura.



Genes e genômica

ANÁLISE CLADÍSTICA DO GEMINIVÍRUS: EVIDÊNCIA DE MULTIESPECIFICIDADE VIRAL EM RELAÇÃO AOS CULTIVARES.

Autores: Rafael Trindade Maia¹; Aparecida Yasmim Silva de Azevedo¹; Bartira Chaves¹;

E-mail para correspondência: rafael.rafatrin@gmail.com

Instituições: ¹CDSA-UFPG;

Palavras-chave: Máxima Verossimilhança; Vírus do mosaico; Tomateiro

Apoio: UAEB-CDSA-UFPG

Nos últimos anos os vírus da família *Geminiviridae* vem sendo intensamente estudados, tendo em vista a severidade das doenças causadas em diversas culturas de importância econômica nacional; como é o caso do feijão-comum, algodão (*Gossypium mustelinum* L.), milho (*Zea mays* L.), tomate e mandioca (*Manihot esculenta* Crantz.). A família *Geminiviridae* é caracterizada estruturalmente pela morfologia geminada da partícula viral, com 18-30 nanômetros; e geneticamente por possuir uma (monopartidos) ou duas (bipartidos) moléculas de DNA circular de fita simples (ssDNA). Cada uma das moléculas apresenta 2500-3000 nucleotídeos encapsulada por uma única proteína estrutural, que se arranja na forma de 22 capsômeros; formando dois icosaedros incompletos que confere o aspecto geminado das partículas virais, característico desta família de vírus de plantas. Apesar de família *Geminiviridae* apresentar grande diversidade de gêneros, um destes vêm se destacando nas pesquisas recentes, por desencadear várias doenças que acometem o tomateiro, pertencente ao gênero *Begomovirus*. Atualmente é o gênero de maior relevância pela alta incidência de viroses, apresentando relatos de perdas de até 100% em lavouras de tomate. O objetivo deste trabalho foi analisar cladisticamente populações de *Begomovirus* através de ferramentas computacionais. As sequências do genoma viral foram obtidas no NCBI (<https://www.ncbi.nlm.nih.gov/>), totalizando 297 sequências oriundas de diferentes países. As sequências foram alinhadas com o programa BIOEDIT, utilizando o algoritmo *ClustalW*. Após o alinhamento, o cladograma foi obtido através do método da Máxima Verossimilhança no programa MEGA, com análise de *bootstrap* de 1000 replicatas. Para escolha do modelo evolutivo mais adequado (Reversível G+I) utilizou-se o *ModelTest*. Os resultados obtidos mostraram que os valores de *bootstrap* dos clados variaram de 9 a 100. Sequências de vírus de pimenta, batata e melancia foram utilizados com *outgroup*, porém as mesmas se agruparam dentro de clados compartilhados pelos vírus obtidos no tomateiro, revelando assim uma evidência de que o vírus do mosaico nestes cultivares pertencem à mesma espécie ou mesmo gênero.

Genes e genômica

ANALYSIS OF THE GENETIC DIVERGENCE OF PASTEURELLA MULTOCIDA

Autores: Raquel Hurtado Castillo¹; Dennis Carhuaricra²; Siomar Soares³; Marcus Vinicius Canário Viana¹; Vasco Azevedo¹; Lenin Maturrano²; Flávia Aburjaile^{1,4,5};

E-mail para correspondência: raquelgen1@gmail.com

Instituições: ¹Laboratory of Cellular and Molecular Genetics, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; ²Laboratory of Molecular Biology and Genetics, Veterinary Medicine Faculty, San Marcos University, Lima, Peru; ³Department of Microbiology, Immunology and Parasitology, Institute of Biological and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Minas Gerais, Brazil; ⁴Laboratory of Plant Genetics and Biotechnology, Federal University of Pernambuco; ⁵Laboratory of Plant Genetics and Biotechnology, Federal University of Pernambuco, Recife, Pernambuco, Brazil;

Palavras-chave: virulence factors; genomic islands; pangenome analysis

Apoio: The research was supported by Coordination for the Improvement of Higher Education Personnel (CAPES), Brazilian National Research Council (CNPq) and Programa Nacional de Innovación para la Competitividad y Productividad (Innovate Perú).

Pasteurella multocida is an opportunistic pathogen causing a wide range of diseases such as snuffles, pneumonia, atrophic rhinitis, fowl cholera and hemorrhagic septicemia in different hosts. These diseases frequently lead to significant morbidity and mortality, causing important economic loss of livestock in the world. As means for controlling the infections are required, several research groups are currently investigating the molecular base of *P. multocida*. In this context, this work aimed to perform a pan-genomic analysis to infer phylogeny and evidence horizontal gene transfer among *P. multocida* isolated from distinct diseases and hosts. Genome sequences of 23 strains of *P. multocida* were retrieved from NCBI. The OrthoMCL prediction of orthologous proteins considered BLAST hits with threshold values of 85% nucleotide identity and 50% length coverage. Phylogenomic analysis was based on a presence/absence gene matrix, in the accessory genome, generated by PAST3 tool. Pan-genomic analysis allowed the prediction of accessory genes exclusively present in each group. Genomic islands (GEIs) in all strains were predicted using GIPSY and BRIG (BLAST Ring Image Generator). The results revealed a pan-genome containing 3,585 genes and an accessory genome presenting 1,200 genes. Accessory and unique genes presented a high content (46.35% and 49%) of unknown proteins, which reveals a high intraspecific diversity. Additionally, phylogenetic analysis using accessory genome content have shown great diversification among all strains demonstrating a clusterization conforming to the specific diseases. Genomic content analysis identified 6, 7 and 10 GEIs on each disease group. All these genomic islands and their genes were associated with a high presence of phage-related genes, hypothetical proteins and virulence-associated genes. The pan-genomic analysis will provide new insights for horizontal gene transfer into the intraspecies diversity of *P. multocida*. Additionally, the obtained data serve as a starting point to develop innovative strategies for screening *P. multocida* strains, which might be useful for an accurate diagnostic of specific diseases.



Genes e genômica

ALPORT AUTOSSÔMICA DOMINANTE: ESTUDO DE DUAS FAMÍLIAS POTIGUAR.

Autores: Raul Maia Falcão¹; Washington Candeia de Araújo^{2,3}; Selma Maria Bezerra Jeronimo^{2,3}; Jorge Estefano Santana de Souza^{1,4};

E-mail para correspondência: rmf4@cin.ufpe.br

Instituições: ¹BIOME - UFRN; ²DBQ - UFRN; ³IMT - UFRN; ⁴IMD - UFRN;

Palavras-chave: Síndrome de Alport; Genes de colágeno; Variação genética

Apoio: CAPES

Síndrome de Alport é uma nefropatia hereditária de colágeno tipo IV caracterizada por provocar a perda progressiva da função renal e auditiva durante a primeira infância. O diagnóstico precoce pode levar ao tratamento adequado antes do início da insuficiência renal terminal e, portanto, melhorar a expectativa de vida. Neste estudo foi investigado o exoma de duas famílias (F1 e F2) do Rio Grande do Norte (RN), ambas compostas por 4 indivíduos. Esta pesquisa teve como objetivo principal buscar relações diretas entre o impacto de alterações nos genes COL4A3, COL4A4, COL4A5 e o fenótipo da síndrome. Inicialmente o estudo consistiu no sequenciamento do exoma de 8 indivíduos seguido de mapeamento contra o genoma humano (build hg38). Posteriormente, um teste "cego" foi proposto afim de distinguir os progenitores, a prole, assim como apontar o indivíduo com a síndrome. Essa etapa constituiu-se de duas fases: 1- cálculo da cobertura nos cromossomos X e Y; 2- tabela de discordância de *match* dos alelos homozigóticas fixadas nos progenitores. Para identificar as amostras progenitoras inicialmente separou-se indivíduos machos de fêmeas pela busca do padrão: fêmeas apresentando ausência de cobertura no cromossomo Y e aumento de cobertura no cromossomo X. Na etapa seguinte foi feita a chamada de variantes usando as ferramentas GATK e VarScan. De forma a trabalhar apenas com variantes germinativas, foram eliminadas variantes com cobertura menor que 20x, frequência alélica menor que 30%, maior que 70% e mantendo os homozigotos. Em seguida, com base na distância euclidiana de diferença de número de alelos entre os indivíduos, foram confirmadas as amostras progenitoras e prole. Como resultado, as amostras progenitoras e prole foram classificadas com 100% de acerto e, também, com 100% de acerto os indivíduos com síndrome. A análise de variantes apontou 5 alelos não sinônimos possivelmente deletérios em homozigose nos indivíduos com a síndrome. Destes, 3 alelos estão presente nas duas famílias, 1 alelo único de F1 gerou um ganho de *stop* códon no gene COL4A3 e 1 alelo único de F2. A análise do padrão mutacional resultou na identificação de alelos homozigotos nos locus de COL4A4 e COL4A3 apenas nos indivíduos portadores de Alport. Ademais, como as duas famílias são de posições geográficas totalmente distintas e não possuem parentesco, em um passo futuro será possível estudar o efeito fundador que tenha originado mutações presentes nos genes COL4A4 e COL4A3 na população do RN.



Genes e genômica

FATORES DE TRANSCRIÇÃO EM CIANOBACTÉRIAS: PREDIÇÃO POR GENÔMICA COMPARATIVA

Autores: Roberto Brito Xavier Junior¹; Renato Renison Moreira Oliveira¹; Regiane Silva Kawasaki Frances¹; Danielle Costa C. Couto²;

E-mail para correspondência: regianekawasaki@gmail.com

Instituições: ¹Universidade Federal do Pará (UFPA) - Campus Belém; ²Universidade Federal do Pará (UFPA) - Campus Ananindeua;

Palavras-chave: Cianobactérias; Fatores de Transcrição; Banco de Dados Biológicos

Apoio: Universidade Federal do Pará (UFPA) LabioCad - (Laboratório de Bioinformática e Computação de Alto Desempenho) ITV - (Instituto Tecnológico Vale) LTB - Laboratório de Biologia Molecular CNPq

As cianobactérias são micro-organismos fotossintetizantes que possuem um grande poder de adaptação, são responsáveis pela abundância de oxigênio, nitrogênio e carbono na natureza, tornando-se alvos na indústria farmacêutica, terapêutica e de biotecnologia. Fatores de transcrição (TFs) são um grupo de genes responsáveis pela expressão gênica, e estão diretamente ligados aos mecanismos de desenvolvimento, como os que permitem aclimação e adaptação. Entender os fatores de transcrição é importante para a investigação da história evolutiva das cianobactérias e suas características. O cTFbase é um banco de dados de TFs em cianobactérias classificadas e analisadas, entretanto, sua última atualização ocorreu em 2007, tornando-se necessária a identificação de novos fatores de transcrição em novas cianobactérias. O objetivo deste trabalho foi o desenvolvimento de um *pipeline* para a identificação dos TFs de cianobactérias, por meio de um *pipeline* automático e abrangente usando uma combinação de softwares como AUGUSTUS, HMMER e Pfam. O AUGUSTUS foi utilizado para realizar a predição de genes a partir de 52 genomas de cianobactérias obtidas no NCBI. Na ferramenta HMMER, foram criados modelos HMM para cada família de TFs. Em seguida foi identificadas sequências putativas, ou seja, possíveis TFs – baseados nos modelos gerados e nos genes preditos nos passos anteriores. Por fim, foi realizada a validação dos TFs por domínio através do Pfam. Foram obtidos resultados importantes na validação de fatores de transcrição putativos identificados. O *pipeline* permitiu a identificação de novos fatores de transcrição em cianobactérias como a *Anabaena variabilis*, que é um organismo modelo para estudar os primórdios da vida multicelular, e a *Nostoc punctiforme*, uma das bactérias mais versáteis e adaptativas estudadas atualmente. Os resultados obtidos também contribuíram na validação e na identificação de novos fatores de transcrição em algumas cianobactérias que o cTFbase não conseguiu identificar, como a *Gloeobacter violaceus*, *Prochlorococcus marinus* e *Synechococcus elongatus*, que também são organismos alvos de estudos.



Genes e genômica

ESTUDO DA GENÔMICA FUNCIONAL DE *ACIDOVORAX CITRULLI*

Autores: Ricardo Severino Francelino¹; Flávia Figueira Aburjaile¹; Ana Maria Benko Iseppon¹;

E-mail para correspondência: Ricardo.france@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: anotação; fitopatôgeno; genoma

Apoio: CAPES, CNPQ e FACEPE.

Os fitopatógenos são microrganismos responsáveis por causar doenças em plantas através da secreção de enzimas, toxinas e fitoreguladores. Dentre estes, *Acidovorax citrulli* (também conhecido como *A. x avenae subsp. citrulli*, *Pseudomonas avenae subsp. citrulli* e *P. alcaligenes subsp. citrulli*) é um bastão Gram negativo, bactéria causadora da mancha aquosa, uma doença frequente no melão (*Cucumis melo*). Esta doença é responsável por muitos prejuízos da economia do Nordeste brasileiro. Neste contexto, o objetivo deste trabalho é identificar e caracterizar os genes e proteínas em genomas de *A. citrulli*. Para isso foram selecionados 7 genomas de *A. citrulli* disponíveis no GenBank do NCBI, sendo 2 com genomas completos e 5 incompletos. Posteriormente, a anotação automática dos genomas foi realizada pela ferramenta RAST (*Rapid Annotation using Subsystem Technology*). Houve uma curadoria manual para os principais genes de interesse, de acordo com a literatura. Na anotação manual, os bancos de dados utilizados foram NCBI, UniProt, InterProScan e Pfam. Em seguida, todos os genomas foram processados no GO FEAT (*Gene Ontology Functional Enrichment Annotation Tool*). Os resultados apresentaram a divisão de genes e proteínas de acordo com os subsistemas, sendo os mais representativos: (i) aminoácidos, (ii) carboidratos, (iii) cofatores, vitaminas, grupos protéticos e pigmentos, e (iv) virulência. Dentre os aminoácidos e derivados, em todos os genomas houve o predomínio dos aminoácidos: lisina; treonina; metionina e cisteína. Além disso, observa-se que grande parte dos subsistemas anotados está relacionada ao metabolismo primário, associados, por exemplo, à produção de energia e outros compostos essenciais, como os aminoácidos. Com relação à virulência foi observado que o número de genes varia entre 88 a 96 nos genomas analisados, estando a grande maioria relacionado com a produção de compostos de resistência a antibióticos e síntese de peptídeos antimicrobianos. Uma vez que a capacidade de infectar um hospedeiro pode estar diretamente relacionado, por exemplo, à produção de toxinas do fitopatôgeno. Novas análises serão realizadas a fim de compreender os mecanismos moleculares envolvidos na sobrevivência bacteriana e transmissão das fitopatologias ocasionadas por *A. citrulli*. Assim será possível o desenvolvimento de métodos que bloqueiam a propagação deste fitopatôgeno nas culturas de região do Nordeste.



Genes e genômica

O GANHO DE SÍTIOS DE LIGAÇÃO A FATORES DE TRANSCRIÇÃO ESTÁ ASSOCIADO A MUDANÇAS NA ASSINATURA DE EXPRESSÃO EM CÉREBRO E TESTÍCULO HUMANO E ESTÁ CORRELACIONADO A GENES COM MAIOR AMPLITUDE DE EXPRESSÃO.

Autores: Vandeclecio Lira da Silva²; André Mauricio Ribeiro dos Santos³; Wilfredo Blanco Figuerola⁴; Sandro José de Souza^{1,2};

E-mail para correspondência: vandeclecio@neuro.ufrn.br

Instituições: ¹Instituto do Cérebro, Universidade Federal do Rio Grande do Norte; ²Programa de Pós-graduação em Bioinformática - IMD, Universidade Federal do Rio Grande do Norte ; ³Programa de Pós-graduação em Genética e Biologia Molecular - Universidade Federal do Pará; ⁴Departamento de Ciência da Computação - Universidade Estadual do Rio Grande do Norte;

Palavras-chave: Fatores de Transcrição; expressão; amplitude de expressão

Apoio: CAPES, CNPQ

Acredita-se que o ganho de sítios de ligação a fatores de transcrição (TFBS) represente uma das principais causas da inovação biológica. Neste trabalho, utilizamos estratégias baseadas em genômica comparativa, analisando arquivos ".chain" resultantes de alinhamento humano/chimpanzé e humano/gorila obtidos do portal UCSC Genome Browser, resultando na identificação de 21.822 TFBS específicos da linhagem humana (TFBS-HS), que foram validados utilizando a ferramenta liftOver. Observando em uma janela de 6k bases (5kb upstream + 1kb downstream do sítio de inicialização da transcrição), região promotora dos genes humanos, foram encontrados mais de quarenta por cento (9.206) destes TFBS-HS associados a 1.283 genes. Foram utilizados dados públicos de expressão em diversos tecidos de humano, chimpanzé e gorila. Diferentes abordagens e metodologias, como expressão diferencial entre espécies e amplitude de expressão foram usados para analisar o padrão de expressão. Uma comparação do padrão de expressão dos 1.283 genes e dos correspondentes ortólogos no chimpanzé e no gorila identificou genes diferencialmente expressos em tecidos humanos. Esses genes mostram um padrão de expressão mais divergente no testículo e no cérebro humanos, sugerindo um papel para a seleção positiva na fixação dos ganhos de TFBS. Os genes associados ao TFBS-HS foram enriquecidos, com base no GO, em categorias de ontologias gênicas relacionadas à regulação transcricional, sinalização, diferenciação/desenvolvimento e sistema nervoso. Além disso, os genes associados ao TFBS-HS apresentam maior amplitude de expressão quando comparados aos genes em geral. Esta distribuição tendenciosa é devida a um ganho preferencial de TFBS em genes com maior amplitude de expressão, em vez de um desvio no padrão de expressão após o ganho de TFBS.



Genes e genômica

TIORREDOXINAS ATÍPICAS DE EUCALYPTUS GRANDIS PERTENCENTES À SUB-FAMÍLIA ATACHTS: DIVERSIDADE, CARACTERIZAÇÃO ESTRUTURAL, EXPRESSÃO GÊNICA E GENÔMICA COMPARATIVA.

Autores: Vitória Régia Alves Cavalcante¹; João Victor Villas-Bôas Spelta¹; Paulo Marinho¹;

E-mail para correspondência: vitoriaracavalcante@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: in silico; expressão gênica; oxidoredutases

Apoio: PPGCFL/EAJ/UFRN - Programa de Pós Graduação em Ciências Florestais - UFRN CAPES FAPERN

Tiorredoxinas (Trx) são pequenas proteínas oxidoredutases, presentes em todos os organismos, e surpreendentemente abundantes em plantas. O que as distingue e as caracteriza tanto bioquimicamente quanto nas plataformas de anotação e análise de sequências é a presença de um sítio ativo bastante conservado composto pelos aminoácidos WCGPC. Sua atividade enzimática se exerce a partir da formação de um ditiol entre as duas cisteínas deste sítio tornando-as doadores de hidrogênio para outras proteínas alvo pela quebra de pontes dissulfeto. A caracterização funcional dessas enzimas ao longo do últimos 20 anos e a confirmação da sua complexidade em vegetais é consequência direta da disponibilidade de informações em termos de sequências genômicas e de transcriptomas de plantas realizados. Neste sentido, foram descritos mais de 30 tipos de Trxs em *Arabidopsis thaliana*. Observou-se, no entanto, que esta diversidade se estendia às chamadas Trxs atípicas que apresentavam variações de aminoácidos no centro ativo, além de outros domínios ausentes nas primeiras Trxs estudadas. É o caso da pequena família de Trxs “like” do tipo ACHT (atypical Cys His-rich Trxs) identificadas inicialmente em *A. thaliana* e atuantes nos cloroplastos, precisamente nas membranas dos tilacóides. O objetivo deste trabalho foi de caracterizar esta pequena família de Trxs atípicas em um genoma de árvore, no caso em *Eucalyptus grandis*, a partir de um esforço em bioinformática. Neste sentido, foram utilizadas ferramentas clássicas de busca e análise de sequências em sites de domínio público (NCBI, Phytozome) e softwares para realizar alinhamentos múltiplos (Clustawl) bem como para construção de árvores filogenéticas (MEGA7). Igualmente, dados sobre expressão gênica foram analisados a partir das informações disponibilizadas pelo genoma de *E. grandis*. Os resultados obtidos mostram *E. grandis* codifica pelo menos 6 Trxs atípicas e que a expressão desses genes (dados em FPKM) ocorre majoritariamente em bibliotecas de tecidos de folhas adultas embora expressão específica em tecidos jovens seja identificada. As árvores filogenéticas geradas com Trxs de outras plantas confirmam a identificação e dão suporte à caracterização *in silico* aqui apresentada indicando uma forte conservação numérica e estrutural desta família multigênica em genomas de árvores.

Genes e genômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO IN SILICO DA FAMÍLIA GÊNICA E2F NA ESPÉCIE ZEA MAYS L.

Autores: Jéssica Figuerêdo Campos de Jesus¹; Jéssica Barboza da Silva¹; Wilson Dias de Oliveira¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: wilsondias.wo@gmail.com

Instituições: ¹Departamento de Genética, CCB, Universidade Federal de Pernambuco- UFPE; ²Universidade Federal Rural de Pernambuco;

Palavras-chave: fator de transcrição; bioinformática; milho

Apoio: CNPq, FACEPE.

Os fatores de transcrição E2F são componentes chave do ciclo celular, apresentando diferentes padrões de expressão no limite G1 / S do ciclo de divisão celular em eucariotos superiores. Essa família gênica pode ser dividida em dois grupos: E2F típicos, com a presença de domínios de ligação ao DNA e domínios de ativação transcricional, e E2F atípicos, em que compreende apenas a presença de dois domínios de ligação ao DNA. Embora essa família tenha sido estudada na planta modelo *Arabidopsis thaliana* (L.) Heynh., pouco se sabe sobre como esses fatores regulam os promotores nos vegetais. Diante disso, objetivou-se identificar e caracterizar sequências candidatas da família gênica E2F da espécie *Zea mays*. Inicialmente, foi selecionada uma sequência sonda da família gênica E2F da espécie *Arabidopsis thaliana* no banco de dados do NCBI. Esta sequência foi alinhada via tBLASTn contra sequências pertencentes à espécie *Zea mays* depositadas no NCBI e foram selecionadas sequências com cut-off de e-value $\geq e^{-10}$. As sequências nucleotídicas selecionadas foram anotadas, traduzidas e tiveram seus domínios identificados nas ferramentas ORFfinder e Batch CD-Search, respectivamente. Posteriormente, foi realizada a ancoragem dessas proteínas no genoma da espécie através do *Phytozome*, assim como, a predição do ponto isoelétrico e peso molecular utilizando o JVirGel 2.0 e a localização subcelular por meio do Cell-PLoc 2.0. Foram localizadas 21 sequências candidatas relacionadas com o gene de interesse. As proteínas traduzidas apresentaram variação de 301 a 467 aa. Foram identificadas 12 sequências E2F como típicas, devido à presença dos domínios de ligação (E2F_TDP) e dos domínios de ativação da transcrição (E2F_CC-MB e E2F_DD), e nove atípicas devido à presença apenas dos domínios responsáveis pela ligação ao DNA (E2F-TDP). As sequências apresentaram localização nos cromossomos 2, 4, 5, 6 e 10, pI entre 3,888 e 10,322 e peso molecular variando entre 3,233 e 50,689 kDa. Aproximadamente 62% das proteínas apresentaram localização subcelular no cloroplasto e 38% estavam direcionadas para o núcleo, corroborando com os dados descritos na literatura. Os resultados obtidos podem contribuir para uma melhor compreensão das características moleculares da família gênica E2F e sua participação no processo transcricional em *Zea mays*.



Proteínas e proteômica

DOCKING MOLECULAR DE ANÁLOGOS DE NUCLEOSÍDEOS DERIVADOS DA FRUTOSE COM AÇÃO INIBITÓRIA DA TRANSCRIPTASE REVERSA PARA PREDIÇÃO ANTI-HIV

Autores: Alex France Messias Monteiro¹; Marcus Tullius Scotti¹; Luciana Scotti^{1,2};

E-mail para correspondência: alexfrancem@ltf.ufpb.br

Instituições: ¹Universidade Federal da Paraíba, Laboratório de Tecnologias Farmacêuticas, 50670-910, João Pessoa, PB, Brasil; ²Gestão de Ensino e Pesquisa ? Hospital Universitário, Universidade Federal da Paraíba, João Pessoa, PB, Brasil;

Palavras-chave: HIV; TRANSCRIPTASE REVERSA; NUCLEOSÍDEOS

Apoio: CAPES - UFPB

A AIDS é uma infecção crônica que compromete o sistema imunológico do indivíduo infectado pelo HIV, deixando-o vulnerável a infecções secundárias. Segundo o Ministério da Saúde houveram 200 mil casos no Brasil em 2017, considerada uma pandemia mundial com registro de 36,7 milhões de casos. O HIV é um retrovírus, possuindo o RNA como material genético, o qual necessita da ação da *transcritase reversa* (TR) para se multiplicar. Um nucleosídeo é formado pela ligação *N*-glicosídica entre um carboidrato e uma base nitrogenada (púrica ou pirimidina), que no meio biológico ele é fosforilado e inserido no material genético ao final da multiplicação viral. Os bioativos análogos aos nucleosídeos naturais ao serem inseridos pela TR na fita de DNA de uma célula infectada, deixam de serem codificados e processo de multiplicação retroviral é imediatamente encerrado. A grande dificuldade hoje é a ação desses análogos de nucleosídeos em outros alvos biológicos não seletivos, como a enzima *protease* e *integrase*, conferindo toxicidade para células não infectadas. Este Trabalho consiste em uma análise computacional através do *Docking Molecular*, com o objetivo de prever a inibição da transcritase reversa, de uma série de 26 análogos de nucleosídeos derivados da frutopiranosose, comparados com moléculas bioativas já inseridas no tratamento anti-HIV. Para este estudo 36 moléculas foram desenhadas no ChemDraw Ultra 12.0 para obtenção de sua fórmula estrutural 2D, em seguida a molécula foi otimizada (RMS 0.1 kcal/Å.mol em no máximo 660 ciclos) pelos métodos de Mecânica Molecular (MM+) e Semi-empírico (AM1) com o auxílio do software HyperChemTM (Release 8.0.6 for Windows) para minimização de energia da estrutura 3D. Por último, foi selecionado um alvo da enzima *transcritase reversa* no PBD (PDB ID 1REV), e no Molegro Virtual Docker 6.0 foi realizada a ancoragem molecular. Analisando os resultados, é possível concluir que algumas moléculas apresentaram energias favoráveis à formação dos complexos ligante-enzima, bem como a presença de interações com resíduos de aminoácidos comuns a inibidores conhecidos. Com isso, este estudo contribui para a obtenção de novas biomoléculas anti-HIV através monossacarídeos facilmente encontrados na natureza.



Proteínas e proteômica

CÁLCULOS DAS ENTALPIAS DE LIGAÇÃO ENTRE AS SUBUNIDADES RTA E RTB DA RICINA ATRAVÉS MÉTODOS SEMIEMPÍRICOS

Autores: Acassio Rocha Santos¹; Gerd Bruno da Rocha¹; Elton José Ferreira Chaves²;

E-mail para correspondência: acassioroch@gmail.com

Instituições: ¹Departamento de Química - Universidade Federal da Paraíba (UFPB); ²Departamento de Biotecnologia - Universidade Federal da Paraíba (UFPB);

Palavras-chave: ricina; entalpia de ligação; métodos semiempíricos

Apoio: CAPES/FAPESQ, CAPES (biologia computacional, auxpe 1375/2014), CNPq, CENAPAD -SP e NPAD-UFRN

A ricina é proteína inativadora de ribossomos composta por duas subunidades (RTA e RTB) ligadas por uma ponte de dissulfeto, em que a RTA é a unidade catalítica¹. Por se tratar de uma proteína citotóxica, a ricina é utilizada como arma química, principalmente por grupos terroristas². Esse trabalho consiste em realizar o cálculo das entalpias de formação (ΔH_f) e o cálculo da geometria das subunidades RTA e RTB separadamente e unidas formando o complexo RTA-RTB através dos métodos semiempíricos de química quântica PM6, PM6-DH+, PM7 e RM1. A estrutura cristalográfica da ricina (2AAI), disponível no PDB foi utilizada nesse estudo. Como a subunidade RTB da ricina apresenta glicosilação em sua estrutura, os cálculos de ΔH_f foram conduzidos para os sistemas RTA-RTB sem glicosilações (8212 átomos) e RTA-RTB com glicosilações (8444 átomos). Outro objetivo foi verificar se as glicosilações da RTB influenciam na energética do complexo RTA-RTB e qual método semiempírico descreve melhor as entalpias de ligação (ΔH_{bind}) dos complexos RTA-RTB, pelo menos do ponto de vista qualitativo. Uma vez calculados os dados para o ΔH_f dos complexos RTA-RTB, obtém-se os valores de ΔH_{bind} para os sistemas RTA-RTB (sem glicosilações) e RTA-RTB (com glicosilações). Foram realizados tanto cálculos single-point quanto otimização de todos os átomos. Para todos os cálculos, usamos o algoritmo de escalonamento linear MOZYME³ e modelo implícito de solvente COSMO para proteínas solvatadas em meio aquoso, disponíveis no pacote MOPAC⁴. Cálculos de energia na geometria experimental com o método PM7 para a RTA-RTB sem glicosilações apresentou entalpia de formação desfavorável ($\Delta H_{bind} = 80,91 \text{ kcalmol}^{-1}$), porém para a RTA-RTB com glicosilações, a entalpia de formação foi altamente favorável ($\Delta H_{bind} = -6.568,31 \text{ kcalmol}^{-1}$). Essa mesma tendência foi observado para os métodos PM6, PM6-DH+ e RM1. Otimizações de todos os átomos com o método PM6 apresentaram ($\Delta H_{bind} = 223,44 \text{ kcalmol}^{-1}$) e ($\Delta H_{bind} = -15,51 \text{ kcalmol}^{-1}$) para a RTA-RTB sem glicosilações e com glicosilações respectivamente. Isso sugere que as glicosilações são muito importantes para a energética da ricina.

Proteínas e proteômica

ZIKA VIRUS PHYLOGENETIC ANALYSIS: A GLOBAL EPIDEMIOLOGICAL PERSPECTIVE

Autores: Almerinda Agrelli¹; Ronald Rodrigues de Moura²; Natalia Pereira da Silva¹; Heitor Horlando Sampaio Araújo da Silva¹; Ronaldo Celerino da Silva¹; Sergio Crovella^{1,2}; Lucas Andre Cavalcanti Brandao^{1,3};

E-mail para correspondência: almerindaapimentel@hotmail.com

Instituições: ¹Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE), Recife, Brazil.; ²Department of Genetics, Federal University of Pernambuco (UFPE), Recife, Brazil. ; ³Department of Pathology, Federal University of Pernambuco (UFPE), Recife, Brazil.;

Palavras-chave: Zika virus; ZIKV; Phylogenetics

Apoio: This work was supported by the National Council for Scientific and Technological Development ? CNPq [440371/2016-3]; the National Council for the Improvement of Higher Education ? CAPES [88887.130801/2016-00]; and the Department of Science and Technology of the Ministry of Health - Decit/SCTIE/MS.

Zika virus (ZIKV) is an enveloped and single-stranded RNA arbovirus of the *Flavivirus* genus, transmitted by *Aedes* mosquitoes, which was first isolated in Uganda in 1947 from a sentinel monkey. More recently, ZIKV has undergone a rapid geographic expansion. In 2007, a ZIKV outbreak was reported in Federated States of Micronesia. In 2013, French Polynesia experienced a ZIKV outbreak. In 2015, during the outbreak in Brazil, ZIKV was first linked to neurological disorders, and since then, the virus has spread rapidly throughout the world. Phylogenetic analysis revealed that two lineages (Asian and African) exist, however Asian lineage is responsible for the recent epidemics. To describe the relationship among Zika virus Asian and African lineages available in a public database. Two-hundred and three complete genomes were downloaded from GenBank (NCBI), translated and aligned by clustalW alignment algorithm, using CLC Main WorkBench. An alignment tree was generated for better comparison among the strains. The strains origin were stratified by countries and continents and observed for two lineages origins: Asian and African. Only two strains out of 203 analyzed are of the African lineage, distributed in Uganda (1) and Nigeria (1), both on the African continent. The strains of Uganda and Nigeria were isolated in 1947 and 1968, respectively, suggesting that these strains are not widespread and may have given rise to the Asian strain, which is widely spread around the world currently. Strains isolated from the Asian continent are distributed as follows: China (11), Thailand (6), Singapore (2), South Korea (2), Philippines (1), Cambodia (1), Malaysian (1) and Vietnam (1). In Europe: Italy (2). In the North America continent: Dominican Republic (21), USA (16), Honduras (15), Nicaragua (10), Haiti (9), Mexico (9), Panama (4), Puerto Rico (4), Jamaica (1), Guatemala (1) and Guadeloupe (1). In South America, where the outbreak had more severe clinical symptoms, strains were isolated in Colombia (41), Brazil (34), Venezuela (3) and Peru (1). In Oceania: French Polynesia (1), Fiji (1) and Australia (1). No strains were isolated in the Antarctic continent. These results corroborate the notion that Asian lineage is possibly originated from the African lineage and that the actual epidemic is due to Asian ZIKV.



Proteínas e proteômica

IDENTIFICAÇÃO DE PROTEÍNAS ESSENCIAIS EM CORYNEBACTERIUM PSEUDOTUBERCULOSIS BASEADO NA REDE DE INTERAÇÃO PROTEÍNA-PROTEÍNA

Autores: Annie Elisabeth Beltrão de Andrade¹; Marco da Costa Schulze¹; Edson Luiz Folador¹;

E-mail para correspondência: anniebeltrao@gmail.com

Instituições: ¹Universidade Federal da Paraíba ;

Palavras-chave: Proteínas hubs; Corynebacterium pseudotuberculosis; Alvo para drogas

Corynebacterium pseudotuberculosis, biovar *Ovis*, é o agente etiológico da linfadenite caseosa, bactéria causadora de perdas econômicas significativas no nordeste brasileiro que se distribui mundialmente em todas as áreas de criação de ovinos e caprinos. É uma bactéria de difícil controle e apresenta resistência aos tratamentos medicamentosos disponíveis. Este trabalho visa utilizar rede de interação proteína-proteína (RIP) para identificar proteínas hubs com potencial alvo para antibióticos que atuam no combate ao patógeno. As RIPs foram construídas como método de mapeamento de interações ortólogas. O grau de conectividade (k), um dos principais parâmetros topológicos avaliados, permite a identificação de proteínas hubs, as quais geralmente desempenham papéis essenciais nos sistemas biológicos, uma vez que a conservação de genes, essencialidade e repertório funcional são refletidos na conectividade das proteínas correspondentes nas RIPs. Proteínas essenciais são produtos de genes indispensáveis para a sobrevivência ou reprodução de um organismo, logo, a deleção de tais proteínas resultará em letalidade. As funções codificadas por genes essenciais são consideradas uma fundação da vida e, portanto tendem a ser muito conservadas nos organismos. A identificação dessas proteínas essenciais (hubs) diminui o custo nos experimentos para a identificação de antibióticos, tornando o processo mais rápido, direcionado e menos oneroso. No NCBI, estão publicamente disponíveis os genomas das oito linhagens utilizadas, o proteoma dessas bactérias foi utilizado na construção das RIPs. Na RIP foram identificadas 263 proteínas hubs conservadas. Das 263 proteínas 250 foram avaliadas por homologia como essenciais na base de dados de genes essenciais (DEG). Usando o pipeline do Mholline foram identificadas 138 proteínas com estrutura tridimensional, as quais foram validadas pelo RMSD e gráfico de Ramachandran. Avaliamos a homologia das 263 proteínas com os hospedeiros *Capra hircus* e *Ovis aries* e identificamos 70 proteínas hubs, comuns às oito linhagens do biovar *Ovis*, sem homologia com os hospedeiros, com estrutura tridimensional e, portanto, com grande potencial para serem usadas como alvo de antibióticos, considerando que a identificação experimental de proteínas essenciais é caro, essa metodologia pode diminuir o custo para a produção de fármacos, uma vez que podem ser testadas por método in-silico para a identificação de novas moléculas que atuem com fármacos.

Proteínas e proteômica

ABORDAGEM DE BIOINFORMÁTICA ESTRUTURAL PARA AVALIAR A REATIVIDADE IMUNOLÓGICA CRUZADA ENTRE ALÉRGENOS DE ÁCAROS E PROTEÍNAS DE HELMINTOS

Autores: Ayrton Breno Pimenta Lisboa¹; Luis Gustavo Carvalho Pacheco¹;

E-mail para correspondência: ayrton.breno@ufba.br

Instituições: ¹Universidade Federal da Bahia;

Palavras-chave: in silico; helmintos; ácaro

Apoio: CAPES

A hipótese de higiene foi estudada ao longo dos anos e ainda há algo a ser discutido. Os cientistas acreditavam que todas as infecções, incluindo as helmintíases, poderiam treinar o sistema imunológico e ajudar a prevenir reações e doenças alérgicas. No entanto, sabe-se que as helmintíases promovem a modulação do sistema imunológico em direção às respostas Th2, promovendo atopia e a produção de IgE. O objetivo do estudo é compreender o papel da conservação estrutural e funcional entre alérgenos de ácaros e ortólogos de helmintos na reatividade cruzada. Realizamos análises *in silico* em ambos os grupos de proteínas. As sequências principais foram obtidas do ácaro da poeira doméstica *Dermatophagoides pteronyssinus*. Sequências dos alérgenos dos grupos 1, 2, 5, 9, 10, 11, 18, 21 e 23 foram utilizadas no estudo. Foram analisadas as relações estruturais, funcionais e imunológicas entre alérgenos de ácaros e ortólogos de helmintos. A modelagem tridimensional das proteínas foi feita com a plataforma Robetta (*de novo*) e a predição de epítopos de células B foi realizada através das plataformas ElliPro, utilizando os arquivos PDB das proteínas modeladas, e CBTope utilizando as sequências de aminoácidos. A conservação estrutural foi avaliada através de alinhamento múltiplo das estruturas 3D (RaptorX Structure Alignment) das proteínas ortólogas e o nível de conservação foi avaliado pelos valores de RMSD. Nossos resultados sugerem que os alérgenos do grupo 1, 9, 10, 11, 18 e os ortólogos de helmintos possuem uma grande conservação estrutural, funcional e, portanto, imunológica. Esses grupos podem induzir sensibilização cruzada e desempenham um papel relevante na reatividade cruzada de IgE. No entanto, o grupo 2 mostra pouca conservação e os grupos 5, 21 e 23 não têm ortólogos encontrados em helmintos, sugerindo que esses grupos de alérgenos não desempenham um papel significativo na reatividade cruzada de IgE do helminto-ácaro. Concluímos que a reatividade cruzada entre as tropomiosinas dominou a discussão sobre a reatividade cruzada entre HDM e helmintos, mas nossos resultados sugerem que vários outros alérgenos desempenham um papel neste processo de reatividade cruzada. Outros grupos de alérgenos devem ser estudados, para que possamos entender o papel de cada alérgeno no processo de reatividade cruzada.



Proteínas e proteômica

DISTRIBUIÇÃO, FUNÇÃO E RELAÇÃO FILOGENÉTICA DA FAMÍLIA DE PROTEÍNAS POLICETÍDEO SINTASES EM GENOMAS DE FUNGOS.

Autores: CLARA DÁFNE ALVES DE FARIAS¹; OSMAN CAVALCANTE JÚNIOR¹; JOSÉ RODRIGO DA SILVA LEANDRO¹; LEONARDO BROETTO¹;

E-mail para correspondência: clara.farias@arapiraca.ufal.br

Instituições: ¹Universidade Federal de Alagoas ;

Palavras-chave: Genômica Comparativa; Fungos Patogênicos; Sequenciamento

Apoio: Universidade Federal de Alagoas - UFAL

Os fungos são importantes patógenos de plantas e animais, além de causarem doenças em seres humanos e nos demais animais, estão relacionados com perdas significativas da produção vegetal. Atualmente existem vários genomas completos para um grande número de fungos permitindo assim, descobertas a respeito de muitos aspectos da biologia desses microrganismos. O objetivo principal deste trabalho foi prospectar proteínas policetídeo sintases de interesse envolvidas na virulência e na patogenicidade a partir de genomas de fungos. Para esta finalidade, foi realizado o levantamento da distribuição das famílias de policetídeo sintases nos genomas de interesse. O *pipeline* de anotação das famílias nos genomas selecionados foi conduzido em dois passos distintos de identificação e anotação. Para prospecção, os genomas coletados foram inquiridos pela presença da família de proteínas utilizando modelo oculto de Markov, derivado de alinhamentos do banco de dados Pfam, a partir dos alinhamentos da família de proteínas. Os *hits* positivos foram submetidos ao procedimento de anotação usando BLASTP contra o banco de dados de proteínas não redundantes (nr) do NCBI. Os dados foram sumarizados e as sequências de interesse extraídas para as análises filogenéticas. Os resultados demonstram que as sequências estão bem distribuídas entre os filos, que se dividem nos três tipos de policetídeo sintases (PKSs). As proteínas PKSs foram agrupadas em oito clados divergentes, permitindo a visualização de um possível cenário evolutivo de suas famílias. A proximidade filogenética da família das PKSs, sugere que a duplicação gênica foi o principal fenômeno que contribuiu para o grande número e variedade das proteínas nos genomas. Muito provavelmente, a radiação de duas possíveis grandes duplicações em Ascomycota e Basidiomycota levou à expansão da família. A filogenia das PKSs é altamente conservada e consistente com a divergência dos táxons de fungos, podendo fornecer evidências sobre a ancestralidade do reino.



Proteínas e proteômica

AVALIAÇÃO DO POTENCIAL DE LIGAÇÃO DE PC-LTP1 EM RELAÇÃO À MEMBRANA LIPÍDICA UTILIZANDO A SIMULAÇÃO DE MONTE CARLO

Autores: Lívia Maria Batista Vilela¹; Carlos André dos Santos Silva¹; Mireli de Santana Rego¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: carlos.biomedicina@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: *Cajanus cajan*; Lipid Transfer Protein; servidor MCPep web

Apoio: CAPES, FACEPE, CNPq

As proteínas transportadoras de lipídeos (LTPs) também exercem função de peptídeos antimicrobianos em plantas e são classificadas em dois grupos LTPs1 e LTPs2, que diferem entre si por seu peso molecular e conformação estrutural, porém, compartilham características estruturais formando quatro α -hélices com oito resíduos de cisteína. Os dobramentos destes peptídeos adotam cavidades que permitem às famílias de LTPs se ligarem e transportar moléculas lipídicas. O trabalho teve por objetivo avaliar o potencial de um peptídeo modificado de LTP em relação à membrana lipídica, utilizando a simulação de Monte Carlo. Para tanto, foram identificadas e caracterizadas sequências de LTP1 em *Cajanus cajan* com o uso de ferramentas de bioinformática. Em seguida, foi selecionada uma dessas sequências, sendo modificada e reduzida de acordo com a predição da atividade antimicrobiana visando melhor atuação frente à membrana patogênica e facilidade de síntese/expressão em etapas posteriores, sendo o peptídeo denominado Pc-LTP1. Foi realizada modelagem *ab initio* pelo algoritmo ROSETTA e posteriormente o modelo gerado foi submetido à simulação de Monte Carlo (MC) utilizando o servidor MCPep. O modelo gerado apresentou duas α -hélices com formação de ponte dissulfeto entre os resíduos de cisteína 13 e 27. Analisando o gráfico de Ramachandran observou-se que 97% dos aminoácidos estavam em regiões mais favoráveis e 3% em regiões menos favoráveis. Na simulação de MC foi possível observar que as interações das α -hélices com as moléculas de água foram abaixo de 50%, indicando que a maioria dos resíduos de aminoácidos não teve interação com essas moléculas o que sugere a natureza hidrofóbica do peptídeo. A localização média dos aminoácidos de Pc-LTP1 em relação à região da bicamada lipídica adotou uma posição onde os resíduos polares (como a C, T, N, S, Y) aderiram à região da membrana hidrofílica, já os resíduos apolares (como a G, A, V, L, I, F) adentraram na membrana fosfolipídica se associando à região hidrofóbica dos lipídeos. Porém, o Pc-LTP1 não apresentou aderência em relação à superfície da membrana lipídica. Diante disso, conclui-se que o servidor MCPep é capaz de realizar interações de pequenos peptídeos helicoidais como o Pc-LTP1, possibilitando a observação de uma interação entre alguns resíduos de aminoácidos e a região transmembrana. Para a futura comprovação de tal interação é necessário combinar condições experimentais que se assemelhem aos dados simulados *in silico*.

Proteínas e proteômica

CARACTERIZAÇÃO IN SILICO DA PROTEÍNA QUINASE SIMILAR AO RECEPTOR LRR DE *STYLOSANTHES SCABRA*

Autores: Carlos Augusto Brandão¹; Sheyla Carla Barbosa da Silva Lima¹; Rayssa Guedes Gomes da Silva¹; Fabiana Aparecida Cavalcante Silva¹; Tercilio Calsa Junior¹;

E-mail para correspondência: carlos.augustob@ufpe.br

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Déficit hídrico; Fabaceae; Bioinformática

Apoio: CNPq e Facepe.

A *Stylosanthes scabra*, também conhecida como alfafa do Nordeste, é uma planta pertencente à família Fabaceae amplamente distribuída na América do Sul. Pode ser utilizada principalmente como pastagem em associação com gramíneas. A análise proteômica diferencial de *S. scabra* em condição de déficit hídrico resultou na identificação, dentre outras, da proteína quinase similar ao receptor (LRR-RLK), mais acumulada sob déficit hídrico. O objetivo deste trabalho foi buscar a validação *in silico* da proteína LRR-LRK de *S. scabra*. Foi realizada busca com a ferramenta tBLASTn no *GenBank* a partir da sonda G7K0W0. As seguintes ferramentas foram utilizadas: ORF *Finder* para observação dos quadros de leitura; CD-*Search* para visualização dos domínios; SignalP 4.1 para identificação do peptídeo sinal; DISULFIND para observação de pontes dissulfeto; PHILIUS para predição transmembrana; ProtParam para predição da massa e ponto isoelétrico, e o SWISS-MODEL para modelagem por homologia. O tBLASTn reportou como resultado os acessos XM_006595480.2, XM_024784281.1 e XM_014648453.2, referentes aos organismos *Glycine max*, *Medicago truncatula* e *Viagna radiata*, respectivamente, com e-value 0 em todos os casos. No CD-*Search*, o domínio PLN00113 foi observado em todas as sequências testadas, sendo para *G. max* intervalo de 22-884 e e-value de $1,55 \times 10^{-54}$, *V. radiata* apresentou intervalo 22-888 e e-value de $2,18 \times 10^{-45}$ e *M. truncatula* intervalo de 79-886 e e-value de $4,08 \times 10^{-44}$. No SignalP 4.1, *G. max* e *M. truncatula* apresentaram peptídeo sinal para região transmembrana, enquanto *V. radiata* não apresentou nenhum peptídeo sinal. No DISULFIND não foram observadas pontes dissulfeto nas sequências. No PHILIUS, foi observado que as proteínas apresentam região transmembrana. No ProtParam, foi observado para a proteína de *G. max* massa de 104.569,29 Da e pI teórico de 5,60, a massa para a proteína de *M. truncatula* apresentou 104.530,78 Da e pI teórico 5,89, e a massa para a de *V. radiata* foi 104.935,23 e pI teórico 5,57. A modelagem por homologia realizada no SWISS-MODEL mostrou identidade de 38,03%, 32,61% e 32,25%, para *G. max*, *M. truncatula* e *V. radiata*, respectivamente, com *template* utilizado 4mna.1.A para as duas primeiras e 4mn8.1.A para a terceira. Diante desses resultados, as proteínas foram validadas a partir de ferramentas de bioinformática, sendo necessário técnicas moleculares para validação experimental das proteínas.



Proteínas e proteômica

IDENTIFICAÇÃO DE EFETORES NO GÊNERO *MONILIOPHTHORA*

Autores: Ceslaine Santos Barbosa^{1,2}; Rute R da Fonseca^{3,4}; Thiago Mafra Batista⁵; Mariana Araújo Barreto^{1,2}; Caio Suzart Argolo¹; Carlos Priminho Pirovani¹; Fabienne Florence Lucienne Micheli^{1,6}; Karina Peres Gramacho^{1,2};

E-mail para correspondência: ceslaine1@gmail.com

Instituições: ¹Departamento de Ciências Biológicas, Centro de Biotecnologia e Genética/Universidade Estadual de Santa Cruz/Ilhéus, Bahia, Brasil.; ²Centro de Pesquisas do Cacau - CEPEC/ Comissão Executiva do Plano da Lavoura Cacaueira- CEPLAC/Ilhéus, Bahia, Brasil.; ³The Bioinformatics Centre, Department of Biology/ University of Copenhagen, Copenhagen, Denmark; ⁴CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal; ⁵Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais/Belo Horizonte, Minas Gerais, Brasil; ⁶CIRAD, UMR AGAP, F-34398 Montpellier, France;

Palavras-chave: Efetoroma; doença da Vassoura de Bruxa ; Monilíase do cacauero

Apoio: FAPESB - Fundação de Amparo à Pesquisa do Estado da Bahia e CNPq - Conselho Nacional de Desenvolvimento Científico

Fungos fitopatogênicos interagem com seus hospedeiros principalmente por meio da secreção de proteínas efetoras. Portanto, a detecção e caracterização funcional *in silico* de candidatos a efetores é um importante passo para entender como estes agentes causam doença. Este trabalho teve por objetivo predizer o repertório de proteínas secretadas candidatas a efetoras (PSCs) de isolados de *Moniliophthora perniciosa*, agente causal da doença da Vassoura de Bruxa do cacauero e *Moniliophthora roreri*, responsável pela Monilíase do cacauero. Foram utilizados 6 isolados representativos de subpopulações previamente descritas, a seguir: i) dois de *M. perniciosa* que diferem quanto a virulência aos genótipos de cacau na Bahia ii) um de *M. perniciosa* do Equador, iii) dois de *M. perniciosa* de hospedeiros solanáceos e iv) um de *M. roreri* do Peru (grupo Bolivar). Para a predição das PSCs, as proteínas com peptídeo sinal na região N-terminal direcionadas à secreção e não retidas na região transmembrana foram preditas como proteínas secretadas. Considerou-se que para serem consideradas efetores estas devem preencher pelo menos um dos seguintes critérios: (i) ter sinal de localização nuclear (NLS), (ii) Ser pequena e rica em cisteína (SCR) e (iii) ser proteína contendo repetições (RCP). As PSCs também foram preditas pelo EffectorP. Com o orthoVenn foi possível identificar clusters ortólogos entre os PSCs dos seis isolados, e os exclusivos de cada isolado e de cada subpopulação. O repertório de PSCs identificados em *M. perniciosa* foi: 157 e 134 proteínas de cacaueros da Bahia, 109 de cacaueros do Equador, e 92 e 80 de isolados de solanáceas selvagens (lobeira e caiçara, respectivamente). *Moniliophthora roreri* teve o maior número de candidatos a efetores, um total de 243 proteínas. Um conjunto de oito efetores foi compartilhado entre todos os isolados de *Moniliophthora*, enquanto outros foram exclusivos de cada subpopulação. Em sua maioria, os candidatos a efetores de *M. perniciosa* foram compartilhados entre as subpopulações, enquanto *M. roreri* apresentou quase 50% de exclusividade da espécie. O presente trabalho fornece uma base de dados do efetoroma putativo de subpopulações do gênero *Moniliophthora*, um primeiro passo para a compreensão do patossistema dessas duas doenças devastadoras do cacauero.



Proteínas e proteômica

ANÁLISE ESTRUTURAL, DOCKING E DINÂMICA MOLECULAR DE UMA LECTINA EDEMATOGÊNICA PURIFICADA A PARTIR DE SEMENTES DE *CENTROLOBIUM MICROCHAETE*

Autores: Cláudio Henrique Dahne de Souza-Filho¹; Vinícius José da Silva Osterne¹; Vanir Reis Pinto-Júnior¹; Antônio Hadson Bastos Neco¹; Kyria Santiago do Nascimento¹;

E-mail para correspondência: falecomclaudiodahne@gmail.com

Instituições: ¹Universidade Federal do Ceará (UFC);

Palavras-chave: Lectina; *Centrolobium microchaete*; Bioinformática

Apoio: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) e Universidade Federal do Ceará (UFC).

Lectinas constituem uma classe de proteínas ou glicoproteínas capazes de se ligar de maneira específica e reversível a carboidratos. Lectinas extraídas de sementes da tribo Dalbergieae (Leguminosae) possuem alta variabilidade nas suas propriedades estruturais, na especificidade a carboidratos e nos efeitos biológicos, tais quais inflamatório, vasorrelaxante e de ligação a antígenos de câncer. A fim de abordar de forma abrangente esses fatores, o presente trabalho objetivou estabelecer e caracterizar a estrutura tridimensional da lectina de *Centrolobium microchaete* (CML) por modelagem por homologia, investigar a interação proteína-carboidrato e avaliar seu efeito inflamatório em camundongos. O *docking* molecular foi realizado para analisar a interação da lectina com monossacarídeos, dissacarídeos e *N*-glicanos. Dois dimanosídeos, metil manose-1,3- α -D-manose (MDM) e manose-1,3- α -D-manose (M13), foram utilizados nas simulações de dinâmica molecular para estudar o comportamento do domínio de reconhecimento a carboidrato (CRD) ao longo do tempo. Os resultados mostraram um domínio expandido dentro do qual interações hidrofóbicas com um grupo metil na molécula MDM eram estabelecidas, revelando assim novas interações das lectinas manose-específicas da tribo Dalbergieae. Para examinar as atividade biológicas, CML foi purificada em cromatografia de afinidade em matriz de Sepharose-manose. A lectina elicitou resposta inflamatória no modelo de edema da pata e estimulou migração de leucócitos para as cavidades peritoneais, um efeito relacionado ao CRD. Pela primeira vez, esse trabalho mostrou a dinâmica molecular de uma lectina da tribo Dalbergieae.



Proteínas e proteômica

EVOLUÇÃO ESTRUTURAL DA PROTEÍNA FLORAL FT EM BRASSICALES

Autores: Deivid Almeida de Jesus¹; Darlisson Mesquita Batista¹; Thiago José de Carvalho André¹; Kauê Santana Costa¹;

E-mail para correspondência: dajesus.deivid.09@gmail.com

Instituições: ¹Universidade Federal do Oeste do Pará;

Palavras-chave: FT; Evolução; Biologia Estrutural

O gene *Flowering Locus T* (FT) é essencial no estímulo à floração em angiospermas. No presente estudo, realizamos uma análise estrutural da proteína com base em sua sequência ancestral estimada por métodos filogenéticos e sequência proteica atual em Brassicales. Sequências codificantes foram obtidas no GenBank através da ferramenta BLASTn utilizando-se como referência a sequência de *Arabidopsis thaliana*. Além de *A. thaliana*, obtivemos sequências para *A. halleri*, *Boechera stricta* e *Brassica oleracea*. Alinhamos as sequências obtidas na ferramenta MUSCLE implementada no programa Mega 7. O modelo evolutivo mais verossímil foi recuperado no programa jModelTest 2 (GTR +I +G), validada com mil repetições de *bootstrap*. Para inferências filogenéticas, usamos uma abordagem bayesiana no programa BEAST, utilizando três espécies da subclasse Pinidae como grupo externo. No programa Mega 7 usando máxima verossimilhança, obtivemos a sequência ancestral. As sequências de nucleotídeos foram traduzidas para aminoácidos. Obtivemos a estrutura tridimensional usando o método de modelagem comparativa no programa Modeller, utilizando como molde a estrutura cristalográfica da proteína FT de *A. thaliana* (PDB ID: 1WKP, cadeia A). O pacote AMBER 16 foi utilizado para minimização da energia das proteínas utilizando os algoritmos steepest descende e gradiente conjugado. A qualidade estereoquímica de cada um dos modelos foi otimizada pela alteração dos ângulos de torção no programa ModRefiner. O gráfico de Ramachandran para cada uma das estruturas foi obtido no programa PROCHECK e o perfil de energia atômica não-local pelo gráfico ANOLEA. Por fim, se alinhou a estrutura e calculamos RMSD-C α entre a estrutura molde e os modelos criados, no programa UCSF Chimera. Na análise da validação, todas as proteínas modeladas obtiveram (RMSD-C α < 0.9) e (Ramachandran > 91%). Houve pouca variação estrutural das proteínas FT ao longo da história evolutiva de Brassicales, conforme indicado pela comparação das proteínas atuais com sua ancestral (RMSD-C α < 2.0), assim como na região do quarto exon, importante para atividade da proteína (RMSD-C α < 2.1). Dada a importante função de ativação floral, sugere-se que a conservação da estrutura de FT encontrada aqui possa ter evoluído sob seleção natural estabilizante.

Proteínas e proteômica

EVIDÊNCIA IN SILICO DE LIGAÇÃO ENTRE INSETICIDAS E UMA GST CLASSE EPSILON (AGGSTE5) NO MOSQUITO TRANSMISSOR DA MALÁRIA

Autores: Débora Souza dos Santos¹; Henriqueta Monalisa Farias¹; Rafael Trindade Maia¹;

E-mail para correspondência: deborasouza.santos@outlook.com

Instituições: ¹Universidade Federal de Campina Grande;

Palavras-chave: GST; docking molecular; inseticidas

As enzimas da superfamília GSTs (glutathione s-transferase) constituem um dos mecanismos de barreira aos inseticidas químicos em insetos, herbicidas em plantas, recentemente estudado a viabilidade de sua influência à resposta contra o câncer de mama. De fundamental importância, esta enzima é responsável por proteger a célula de xenobióticos, catalisando-os e transformando-os em substâncias mais solúveis através da conjugação com a glutathione, sendo classificada, como pertencente à classe microssomais e a citosólicas, esta última composta por subunidades, e nos insetos dividem-se e, seis classes, a qual de interesse neste estudo é a epsilon. A GST em foco neste estudo é a AgGSTE5, a quinta proteína descoberta da GST de *Anopheles Gambiae* da classe epsilon, classe específica de artrópodes caracterizando 65% do total das GSTs. Neste estudo, inseticidas que foram submetidos a simulações de *docking molecular* contra um modelo teórico da AgGSTE5, através do programa Autodock. Os parâmetros destacados para estudo durante o docking molecular são: a Constante de Inibição Estimada (CIE), Energia Intermolecular Final (EIF), Proximidade com o Sítio Ativo (PSA) e Energia Livre de Ligação (ELL), tais de suma importância para o entendimento da qualidade da ancoragem. O complexo 3 do malathion obteve: CIE 6.09 mM, ELL (-3.02), EIF (-6.30), PSA 2.96Å. O complexo 7 do Alpha Cipermetrina: CIE 40.69nM, ELL (-10.08), EIF (-12.17), PSA 1.71Å. O complexo 4 do Acaristop: CIE 413.53, ELL (-0.44), EIF (-9.01), PSA 2.10Å. Os resultados encontrados dos três ligantes tiveram boa interação de PSA apontando mínimas distâncias entre o inseticida e o sítio ativo, taxas de ELL e EIF negativas, o que sugere uma afinidade entre as moléculas, levantando a possibilidade de inibição do efeito destes três inseticidas pela enzima GST.



Proteínas e proteômica

MODELAGEM E ATRACAMENTO MOLECULAR DE UMA LECTINA DE TAKIFUGO RUBRIPES

Autores: Eden Silva e Souza¹; João Paulo Matos Santos Lima¹; Michely Correia Diniz²;

E-mail para correspondência: souza.eden@outlook.com

Instituições: ¹Programa de Pós-graduação em Bioinformática, Universidade Federal do Rio Grande do Norte;
²Universidade Federal do Vale do São Francisco;

Palavras-chave: Molecular docking; Protein modeling; fish protein

Apoio: CAPES

A superfície epitelial de peixes é recoberta por uma camada de muco. O muco está relacionado às suas funções vitais, tais como resistência à agentes infecciosos, respiração, regulação osmótica, locomoção e comunicação intraespecífica. Lectinas, que podem estar presentes no muco, compõem uma família de proteínas ou glicoproteínas, caracterizada pela presença de um ou mais domínios com capacidade de reconhecer e ligar-se reversivelmente a carboidratos (CRDs). Uma lectina de *Takifugo rubripes*, foi isolada do muco epitelial e caracterizada com 116 resíduos, tem afinidade a manose, entretanto, a estrutura terciária desta proteína ainda não está disponível nos bancos de dados. O propósito desse trabalho foi prever *in silico* um modelo da estrutura terciária dessa lectina e o seu atracamento com a manose. Para isso, a sequência de aminoácidos foi obtida no National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) sob o código de acesso GenBank: NP_001027736_1. A modelagem por threading foi realizada no modo intensive do Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) e o I-TASSER, enquanto a modelagem por homologia foi realizada no Modeller. A validação dos modelos criados foi feita usando o Molprobit, ModEval, SAVES e Swiss-model. O atracamento molecular foi feito no SwissDock. O Phyre2 criou um modelo com 100% de confiança e cobertura de 95% baseado na estrutura homóloga de uma lectina de fungo (PDB: 4PDT). Para este modelo, o diagrama de Ramachandran mostrou que 93,4% dos resíduos estão em regiões favoráveis, 4,7% dos resíduos em regiões permitidas e 1,9% em regiões energeticamente não favoráveis. Os outros parâmetros de validação apontam que o modelo criado no Phyre2 é o mais apropriado para essa proteína. O melhor atracamento criado pelo SwissDock apresentou ΔG igual a -6,59 e *FullFitness* igual a -866.88 (kcal/mol), onde a manose interage com uma leucina na posição 99 da proteína. O modelo e atracamento criados indicam como a lectina *T. rubripes* interage com a manose. Esta proteína teve atividade biológica descrita na capacidade de se ligar a superfície de bactérias e do trematodo parasita *Heterobothrium okamotoi*, o atracamento molecular possibilita a visualização da interação proteína-ligante e da *insights* de como esta proteína desempenha sua atividade biológica. Para maior entendimento, mais estudos estruturais relacionados a dinâmica molecular serão realizados.

Proteínas e proteômica

MODELAGEM E CARACTERIZAÇÃO DA PROTEÍNA CDC48 EM SACCHARUM SPP.: UMA ABORDAGEM IN SILICO

Autores: Felipe de Lima Almeida¹; João Paulo Matos Santos Lima^{1,2}; Katia Castanho Scortecchi^{1,3};

E-mail para correspondência: felipe_limacb@yahoo.com.br

Instituições: ¹Instituto Metr pole Digital, Universidade Federal do Rio Grande do Norte; ²Departamento de Bioqu mica, Universidade Federal do Rio Grande do Norte; ³Departamento de Biologia Celular e Gen tica, Universidade Federal do Rio Grande do Norte;

Palavras-chave: CDC48; modelagem por homologia; Saccharum spp.

Apoio: CNPq, CAPES e UFRN

CDC48   uma prote na pertencente   fam lia AAA+ que est  envolvida em muitos eventos biol gicos incluindo processos associados a degrada o de prote nas, como por exemplo, o sistema UPS (Sistema-Ubiquitina-Proteassoma) e ERAD (Degrada o Associada ao Ret culo Endoplasm tico). Respons vel por manter a homeostase proteica, em eucariotos CDC48   altamente expressa durante condi oes normais de crescimento e regulada em situa oes de estresse. A an lise do proteoma de ra zes de cana-de-a o car submetidas previamente a um estresse oxidativo pela presen a de H₂O₂ revelou que CDC48 foi diferencialmente expresso e tolerante   condi o de estresse. Sob esta perspectiva, neste trabalho apresentamos a predi o tridimensional e an lise do modelo de CDC48 em *Saccharum spp.* Utilizando abordagens de modelagem por homologia, adotamos para a pesquisa o modelo com melhor  ndice *C-score* e *TM-score* de confian a. Tamb m empregamos o pacote GROMACS vers o 5.1.4 para simular as equa oes newtonianas de movimento na perspectiva da modelagem ser realizada em ambiente pr ximo de seu estado nativo. Para valida o do modelo predito analisamos os indicadores que fornecem confiabilidade   estrutura e, posteriormente, o submetemos a din mica molecular. A partir do modelo funcional de CDC48 em cana-de-a o car foi obtido um homohex mero est vel em forma de anel no qual cada mon mero   formado por um dom nio N-terminal, dois dom nios AAA ATPase (D1 e D2) e uma cauda C-terminal. An lises preliminares do modelo proteico sugerem dois estados de nucleot deos (ADP-AIFx e ADP) que ajudam a caracterizar as mudan as conformacionais de CDC48   medida que progride atrav s do ciclo de hidr lise de ATP. Al m disso, identificamos na cauda de CDC48 um res duo Hb (res duo hidrof bico), a tirosina (Y) e o res duo C-terminal asparagina (N), sugerindo o *motif* HbYN um mediador de intera o com PNGase que atua em mecanismos de degrada o proteica. Estes resultados sinalizam que a conforma o estrutural de CDC48 via hidr lise   necess ria e que a sua intera o com PNGase   importante para um subgrupo de processos biol gicos dependentes, fornecendo subs dios para compreens o do papel de CDC48 na resposta ao estresse oxidativo em *Saccharum spp.*



Proteínas e proteômica

ANÁLISE ESTRUTURAL DA PROTEÍNA GLUTAMIL-TRNA REDUTASE, CHOROPLASTIDIAL (GLU-TR) DE SACCHARUM SPP SOB ESTRESSE HÍDRICO.

Autores: Georon Ferreira de Sousa¹; Fabiana Aparecida Cavalcante Silva¹; Tercilio Calsa Júnior¹;

E-mail para correspondência: georon.sousa@gmail.com

Instituições: ¹Universidade Federal de Pernambuco, Centro de Biociências, Departamento de Genética, Laboratório de Genômica e Proteômica de Plantas;

Palavras-chave: Proteômica; Bioinformática; Biologia molecular

Apoio: CAPES

Os mecanismos pelos quais as plantas conseguem sobreviver e se adaptar às mudanças climáticas como seca e altos índices de radiação ultravioleta são alvo de diversos estudos. Sabe-se que a capacidade de adaptação da planta resulta de alterações no metabolismo da célula vegetal, entre elas a síntese de proteínas. Estudo prévio de análise do proteoma foliar de cana-de-açúcar submetida a estresses abióticos combinados permitiu a identificação e anotação presumível da proteína Glutamil-tRNA redutase (Glu-TR – Acesso Uniprot P49294) como exclusiva do tratamento sob estresse hídrico e alta radiação UVB) Essa proteína está diretamente envolvida na resposta ao estresse fotooxidativo atuando na biossíntese de clorofila. Diante da importância dessa proteínas o objetivo deste trabalho foi caracterizar *in silico* a sequência de aminoácidos da Glu-TR (Acesso do Uniprot P49294) com o auxílio de ferramentas de bioinformática. Para confirmação do domínio da proteína foi utilizado o programa online CD-Search. Em seguida, a sequência foi examinada quanto ao ponto isoelétrico (pI) e peso molecular (PM), presença de peptídeo sinal e pontes dissulfetos por meio dos programas online ProtParam, SignalP e DISULFIND, respectivamente. Posteriormente, foi utilizado o PHILIUS para avaliar predição transmembrana, SWISS-MODEL para modelagem por homologia. PROTCOMP para identificação da localização sub-celular da proteína. De acordo com as análises a proteína é globular não citoplasmática, de peso molecular 58291,64 kDa e ponto isoelétrico 8,82. A proteína apresenta o domínio conservado ‘Glutamil-tRNA redutase’ entre os aminoácidos 1 a 519. A localização subcelular identificada para a sequência é do tipo cloroplastidial de membrana, corroborando com os dados na literatura. A modelagem *in silico* da sequência de Glu-TR revelou identidade de 85,59% com o modelo 5che.1.A. de Arabidopsis. Os dados obtidos permitiram uma melhor compreensão da estrutura da proteína Glu-TR, contudo são necessários estudos complementares para determinação da associação entre os níveis transcricionais e os estresses aplicados.

Proteínas e proteômica

ANÁLISE IN SILICO DAS PROTEÍNAS RUBISCO E YCF4 EM ALGODÃO NATURALMENTE COLORIDO SOB ESTRESSE POR RADIAÇÃO UVB

Autores: Geisenilma Maria Gonçalves da Rocha¹; Tainara Fernandes Dantas¹; Fabiana Aparecida Cavalcante Silva¹; Elton Pedro Nunes Pena¹; Liziane Maria de Lima²; Tercilio Calsa Junior¹;

E-mail para correspondência: geisenilma@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco, Centro de Biociências, Departamento de Genética, Laboratório de Genômica e Proteômica de Plantas, Recife, PE, Brasil. ; ²Embrapa Algodão, Laboratório de Biotecnologia, Campina Grande, PB, Brasil.;

Palavras-chave: Malvaceae; *Gossypium hirsutum*; Bioinformática

Apoio: UFPE, CAPES, Embrapa Algodão e RENORBIO

O algodão é uma planta pertencente à família *Malvaceae* e gênero *Gossypium*, suas fibras são as mais usadas na indústria têxtil mundial. Contudo, o estresse abiótico, incluindo radiação UVB, provoca diversos efeitos adversos em plantas, como alterações no acúmulo de proteínas. A análise proteômica preliminar da cultivar BRS Rubi permitiu a anotação presumível das proteínas RBCL - Rubisco (Q8SL24) e Ycf4 (P25412) em condições de estresse por radiação UVB. A Rubisco atua na fixação do carbono no ciclo de Calvin-Benson, enquanto Ycf4 está envolvida na montagem do complexo fotossistema I e, se ausente, a fotossíntese é ineficiente, afetando o crescimento das plantas. Objetivou-se neste trabalho, caracterizar *in silico a sequência* das proteínas Rubisco e Ycf4 acumuladas em algodão naturalmente colorido sob estresse por radiação UVB. Foi realizada a predição da interação de cinco proteínas identificadas em BRS Rubi com proteínas existentes, através do banco de dados STRING versão 9.05. A sequência dos acessos Q8SL24 e P25412 foram utilizadas como sondas para a realização de tBLASTn contra o *Genbank* NCBI e considerou-se o alinhamento apenas para a família *Malvaceae*. Em seguida foram utilizadas as ferramentas ORF *Finder* para verificação do quadro de leitura; CD-*Search* para observação de domínios conservados; SignalP 4.1 para identificar a presença de peptídeo sinal; DISULFIND para identificar as pontes dissulfeto; PHILIUS para predição transmembrana; ProtParam para verificar o peso molecular e o ponto isoelétrico teórico da proteína, por fim, o Swiss-Model para modelagem por homologia. O modelo de rede de interações proposto inclui, a relação entre as proteínas Rubisco e Ycf4 relacionadas à fotossíntese. Foi observado no tBLASTn, para proteína Rubisco alinhada com *G. australe* (KP221928.1) e *G. populifolium* (KP221924.1), identidade de 95% e *e-value* 0 em ambos; para proteína Ycf4 alinhada com *G. hirsutum* (HQ901196.1) e *G. aridum* (KP170502.1) observou-se identidade de 72% e *e-value* 2×10^{-83} em ambos. Nas ferramentas DISULFIND e SignalP 4.1 observou-se que as proteínas não apresentaram pontes dissulfeto e peptídeo sinal, respectivamente. O resultado do PHILIUS demonstrou que as proteínas apresentaram característica do tipo globular. O presente trabalho refere-se a um estudo preliminar *in silico* das proteínas Rubisco e Ycf4, sugerindo sua relevância para futuros estudos funcionais no processo fotossintético e, portanto, no crescimento vegetativo e na produtividade.



Proteínas e proteômica

SELEÇÃO *IN SILICO* DE INIBIDORES EM POTENCIAL DE UMA PROTEÍNA DE REPLICAÇÃO VIRAL (NS5 RDRP) DO VÍRUS DA ZIKA

Autores: Henriqueta Monalisa Farias¹; Franklin Ferreira de Farias Nóbrega¹; Rafael Trindade Maia¹;

E-mail para correspondência: monalisa_miller@hotmail.com

Instituições: ¹Universidade Federal de Campina Grande;

Palavras-chave: Saúde Pública; Arbovírus; Docking Molecular

Apoio: CNPq/UFPG.

Apesar dos sintomas parecerem triviais, o vírus Zika (ZIKV) tornou-se uma das maiores preocupações na saúde pública mundial e do Brasil, principalmente depois da descoberta de sua associação com casos de microcefalia. Até o momento não existe vacina disponível e o tratamento é paliativo. No Brasil a confirmação de transmissão da febre causada pelo ZIKV surgiu em 2015, apresentou-se inicialmente nos estados do Rio Grande do Norte e da Bahia. Em uma recente pesquisa foi elucidada a estrutura tridimensional da ZIKV NS5 RdRp, uma proteína associada à RNA polimerase e importante ativadora do processo de replicação do vírus, com esta informação, veio a possibilidade de tornar a mesma um alvo de novos fármacos. Os estudos teórico-computacionais são de grande valia para sinalizar possíveis inibidores da NS5 RdRp que funcionem como antivirais, logo, esta pesquisa teve como objetivo principal selecionar *in silico* possíveis inibidores da proteína de replicação NS5. A metodologia seguiu os seguintes passos: Análise da estrutura tridimensional da NS5, bem como seu sítio ativo; busca na literatura e nos bancos de dados de moléculas por possíveis inibidores; Simulações de *docking* molecular e investigação de interações nos complexos proteína-ligantes e/ou proteína-proteína gerados. Buscou-se destacar quatro parâmetros considerados mais importantes na avaliação para o melhor complexo de cada *docking*: A proximidade com o sítio ativo da proteína (PSAP); a Constante de Inibição Estimada (CIE), a Energia intermolecular Final (EIF) e a Energia livre de ligação (ELL). O terpeno (-) apresentou em seu complexo 1: 2.98 Å de PSAP; 333.9 µM de CIE; -5.34 Kcal/mol de EIF; -4.74 Kcal/mol de ELL. O terpeno (+) apresentou em seu complexo 10: 3.46 Å de PSAP; 365.31 µM de CIE; -5.29 Kcal/mol de EIF; -4.69 Kcal/mol de ELL. O último composto em destaque é a 1-3 Indandione, seu complexo 10 demonstrou: 2.74 Å de PSAP; 310 µM de CIE; -4.79 Kcal/mol de EIF; -4.79 Kcal/mol de ELL. Estes três resultados em ênfase expressaram uma interação direta com o sítio ativo da proteína por meio de ligações covalentes, além de valores extremamente altos nas constantes de inibição e suas energias (Energia intermolecular final e Energia livre de ligação) apontaram valores negativos, o que indica maior afinidade no atracamento dessas moléculas. Os três compostos em questão demonstraram potencial na inibição da NS5 RdRp do ZIKV.



Proteínas e proteômica

A DOENÇA DE HUNTINGTON E UM ESTUDO COMPARATIVO DA HUNTINGTINA EM DIFERENTES ORGANISMOS

Autores: Joel Augusto Moura Porto¹; Mailane dos Anjos Silva¹; Eric Roberto Guimarães Rocha Aguiar¹;

E-mail para correspondência: joelaugustomp@gmail.com

Instituições: ¹Universidade Federal da Bahia;

Palavras-chave: doença de huntington; huntingtina; motifs

Apoio: Universidade Federal da Bahia, Instituto de Ciências da Saúde, Departamento de Biotecnologia

A Doença (ou Coreia) de Huntington é causada pela expansão instável do triplete CAG, uma desordem hereditária autossômica e dominante que resulta na morte de células cerebrais. A DH foi descrita inicialmente no século 19 pelo médico George Huntington. Ele identificou as características clínicas e o padrão de transmissão familiar. Apenas em 1993, um grupo de nome *Hereditary Disease Foundation* identificou uma expansão instável do triplete Citosina-Adenina-Guanina), na região codificante (éxon 1) do gene *HD*, que codifica a proteína huntingtina. O gene *HTT* está localizado no braço curto do cromossomo 4 em 4p16.3.

Este trabalho objetiva analisar comparativamente quais os *motifs* presentes genes referentes à produção da proteína huntingtina em organismos tais quais: *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Sus scrofa*, *Equus caballus*, *Danio rerio*, *Ratus norvegicus*, *Drosophila melanogaster*, *Canis lupus familiaris*. Esses organismos foram escolhidos com base, sobretudo, nas bibliotecas disponíveis de melhor qualidade do NCBI.

Como ferramenta para execução do presente trabalho, foram utilizados o InterPro para a identificação de família e superfamília de proteínas homólogas e seus domínios funcionais; o MAFFT e o MUSCLE, ambos para o alinhamento de sequências, e o MEMESuite (TOMTOM) para a identificação dos *motifs* conservados nos diferentes organismos.

Analisamos os *motifs* associados ao gene de produção da huntingtina e notamos a prevalência de três deles em quase todos os nove organismos estudados. Deles, o *motif* relacionado ao direcionamento de proteínas transmembranas tipo I da superfície celular para o compartimento lisossomal-endossomal esteve presente em oito dos nove organismos (apenas o peixe *Danio rerio* não o apresentou). Outros dois *motifs* expressos foram encontrados em sete dos nove organismos analisados. Um deles está vinculado à ligação aos receptores nucleares das células cerebrais. O outro relaciona-se com a ligação da paxilina como um peptídeo helicoidal anfipático ao domínio FAT, mecanismo ainda não muito bem elucidado e compreendido pelo meio acadêmico. Tais *motifs* não foram encontrados, mais uma vez, no peixe *Danio rerio* e em cavalos (*Equus caballus*).

Com relação às proteínas homólogas, mantiveram-se, em grande parte dos organismos analisados, os *loci* gênicos responsáveis por intitular a superfamília huntingtina. Isso mostra que, apesar de possíveis mecanismos de *splicing*, os organismos podem apresentar conformações análogas entre si.



Proteínas e proteômica

ESTUDOS DE DOCKING MOLECULAR DE DERIVADOS BENZOTIAZINONAS COMO POTENCIAIS INIBIDORES MULTI-TARGET EM PROTEÍNAS DA MYCOBACTERIUM TUBERCULOSIS

Autores: Jéssika de Oliveira Viana¹; Marcus Tullius Scotti¹; Luciana Scotti^{1,2};

E-mail para correspondência: viana_jess@hotmail.com

Instituições: ¹Programa de Pós-graduação em Produtos Naturais e Sintéticos Bioativos (PgPNSB), Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa-PB, Brasil; ²Gestão de Ensino e Pesquisa - Hospital Universitário da Universidade Federal da Paraíba, João Pessoa - PB, Brasil.;

Palavras-chave: Antituberculose; In silico; Multi-alvo

Apoio: Agradecemos o apoio financeiro disponibilizado pela Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) para o desenvolvimento da pesquisa.

A Tuberculose é uma doença infecciosa causada pelo bacilo *Mycobacterium tuberculosis* (Mtb) e que atualmente acomete cerca de 6,3 milhões de novos casos por ano. Diante deste prejuízo social do Mtb há a necessidade de planejamento novos fármacos antituberculose. Dentre estes compostos estão as benzotiazinonas (BTZ), que apresentam atividade nanomolar e bloqueiam a via sintética da formação da membrana micobacteriana. Nosso trabalho visou investigar, através de Docking Molecular, as interações e potencial inibição de 3 proteínas da Mtb: a DNA Gyrase que age na síntese proteica, o complexo CFP10-ESAT6 que interfere na sinalização da lise celular no hospedeiro e a PKS13 que atua na formação da membrana na micobactéria. Analisamos 24 derivados benzotiazinonas com 0.0001-0.05 µg/ml através do programa HyperChem v.8.0, onde os compostos foram desenhados e sua geometria molecular foi otimizada por mecânica molecular MM+ e pelo método semi-empírico AM1. No *Protein Data Bank* (PDB) foram selecionadas as estruturas proteicas DNA Gyrase (ID PDB 5BS8), CFP10-ESAT6 (ID PDB 3FAV) e PKS13 (ID PDB 5V3Y), que foram usadas como dados de entrada no programa Molegro v. 6.0 para realização do Docking Molecular. Foi observado que todos os compostos apresentaram interações com as proteínas investigadas, mostrando que a formação do complexo é favorável. Com a DNA Gyrase de Mtb, observamos que 10 compostos tiveram a energia de interação menor que o Moxifloxacina (-37.6 kcal/mol), o principal inibidor da enzima, e interações estéricas em comum. Para a CFP10-ESAT6 vimos que 5 moléculas apresentaram energia de interação superior ao imidazole (-25.4 kcal/mol) e ao inibidor de literatura da enzima (Benzimidazol), com interações estéricas em comum. Com relação a PKS13 observamos valores de energia próximos (-120 kcal/mol) ao inibidor TAM6 e interações de hidrogênio e estéricas em comum. Os resultados do estudo puderam confirmar por métodos *in silico* a potencial atividade *multi-target* das benzotiazinonas nas 3 enzimas, além de ampliar nossas confirmações *in silico* anteriores de que derivados benzotiazinonas atuam na enzima DprE1, sendo útil em bloquear várias vias de sinalização simultaneamente.



Proteínas e proteômica

DESENVOLVIMENTO DE ABORDAGENS COMPUTACIONAIS PARA ANÁLISE E IDENTIFICAÇÃO DE PEPTÍDEOS POLIMÓRFICOS

Autores: Lucas Marques da Cunha^{1,2}; Vandeclécio Lira da Silva^{1,2}; Gilderlânio Santana de Araújo^{1,2}; Gustavo Antônio de Souza^{1,2}; Sandro José de Souza^{1,2};

E-mail para correspondência: lucasmarques@ufrn.edu.br

Instituições: ¹Universidade Federal do Rio Grande do Norte; ²Centro Multiusuário de Bioinformática;

Palavras-chave: proteômica; banco de dados; peptídeos polimórficos

Apoio: A Rede Biologia Sistêmica do Câncer (BSC) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

A abordagem proteômica permite estudos em larga escala da expressão proteica em diferentes tecidos e fluidos corporais, tendo como objetivo identificar e quantificar o conteúdo proteico total, analisar a dinamicidade do genoma, determinar funções e interações proteicas e analisar modificações pós-traducionais. No processo de análise proteômica, a identificação de proteínas ainda apresenta lacunas, apesar dos grandes avanços na área. A espectrometria de massas é método mais utilizado nesse processo que permite gerar valores de massa/carga das amostras. A complementação desse processo é feita utilizando um software identificador que compara os valores da amostras com um banco de dados teórico. Porém, grande parte dos bancos de dados existentes não contemplam as variações que ocorrem no DNA que podem impactar na sequência de aminoácidos e impossibilitar o processo de identificação das proteínas. Assim, esta pesquisa propõe o desenvolvimento de uma abordagem computacional para análise e identificação de peptídeos polimórficos. Para isso, desenvolvemos um banco de dados com sequências hipotéticas contendo somente os peptídeos que apresentam mutação na proteína. O processo de identificação é realizado em duas etapas. Na primeira etapa utilizamos uma base de dados de referência e na segunda etapa a base de dados contendo os peptídeos polimórficos. A partir das saídas resultantes, verificamos a unicidade dos peptídeos das bases e, em casos de redundância, selecionamos o peptídeo com melhor score. Os peptídeos polimórficos identificados também são classificados de acordo com o tipo de mutação, frequência alélica e patogenicidade. Como resultado preliminar, foram identificados 3,013 novos peptídeos utilizando a base polimórfica. Desse total, 82% são SNPs, 13% INDEL, 5% frameshift e menos de 1% correspondendo a Lost Stop e UTR variation. As mutações identificadas estão relacionadas à Surdez não sindrômica (MAF<1%, AFR), Hipomielinização com envolvimento do tronco encefálico e da medula espinhal e espasticidade da perna (MAF<1%, AFR), Doença de Gaucher (MAF<1%, AFR, EAS, EUR, SAS) e câncer da mama (MAF=1,2%, AFR, MAF<1% AMR, EUR). Portanto, nossa abordagem mostra-se bastante promissora quanto ao objetivo estabelecido e aplicável a análises com novas amostras. Como trabalhos futuros, pretendemos desenvolver uma plataforma web que torne esses dados públicos para comunidade científica.



Proteínas e proteômica

PREDIÇÃO DE PROTEÍNAS ESSENCIAIS EM INTERATOMA POR MEIO DE SVM EM ÍNDICES DE CENTRALIDADES

Autores: Mirele Carolina Souza Ferreira Costa, mestranda em Informática, CIC-UnB¹; Thais Couto Laureano, mestranda em Bioinformática, ICB-UFMG²; Daniella Castanheira Bartholomeu, professora, ICB-UFMG²; Robson da Silva Lopes, professor Ciência da Computação, ICET-CUA-UFMT³;

E-mail para correspondência: carolinamirele@gmail.com

Instituições: ¹Universidade de Brasília - UnB; ²Universidade Federal de Minas Gerais - UFMG; ³Universidade Federal do Mato Grosso - UFMT;

Palavras-chave: Proteínas essenciais; rede de interação proteína-proteína; índices de centralidades

Proteínas são substâncias que atuam em inúmeras funções no organismo dos seres vivos, sendo que, muitas proteínas interagem com outras desempenhando papéis extremamente importantes em diversos processos biológicos, tornando-se indispensáveis para sobrevivência e reprodução. A deleção de tais proteínas, chamadas de proteínas essenciais, podem resultar na letalidade ou infertilidade, fazendo com que a identificação dessas proteínas possa contribuir para o desenvolvimento de medicamentos, que tratam doenças causadas por diversos patógenos. Uma forma de tentar compreender as proteínas essenciais é por meio do estudo das redes de interações protéicas, os interatomas. Muitos trabalhos têm estudado índices de centralidades de topologia de redes, tais como Centralidade de Grau, Centralidade de Proximidade, Centralidade de Intermediação, BottleNeck, Componente de Vizinhança Máximo (MNC), Densidade do Componente de Vizinhança Máximo (DMNC) e Componente Percolado de Borda, para identificar proteínas essenciais em redes de interação proteína-proteína (IPP). No entanto, estes trabalhos têm buscado identificar quais índices de topologia, de forma isolada, apresentam melhores resultados na identificação de proteínas essenciais, sem levar em consideração a combinação deles com técnicas de aprendizado de máquina (AM). Assim, o objetivo deste trabalho foi prever proteínas essenciais em redes IPP, por meio de índices de centralidades de topologia de redes aplicados em técnicas de AM. Foi calculado os índices de centralidades para cada proteína essencial e não essencial, experimentalmente identificada, de três redes de interações proteicas, duas em bactérias (*Escherichia coli* e *Helicobacter pylori*) e uma em um eucariota (*Saccharomyces cerevisiae*). Para prever proteínas essenciais foi utilizado a técnica de AM Support Vector Machine. As previsões apresentaram AUC (Área sob a curva) de $0.70(\pm 0.008)$, para *S. cerevisiae*, sendo que, nas previsões de forma isolada para o mesmo organismo, alguns índices alcançaram uma AUC de 0.69. Além disso, a AUC das espécies *E. coli* e *H. pylori* ficaram em torno de 0.50 na predição isolada, já nas previsões com os índices combinados com AM das duas espécies alcançaram AUC superior à 0.66. Os resultados mostram que a aplicação de índices de centralidades combinados com AM podem contribuir significativamente para a predição de proteínas essenciais, quando comparado com a identificação de proteínas essenciais utilizando índices de forma isolada.



Proteínas e proteômica

STRUCTURAL FEATURES, STRENGTH AFFINITY AND EVOLUTIONARY PRESSURE CORRELATES WITH P53 HIJACKING BY E6 AND E6AP ON RISK OF HPV-MEDIATED ONCOGENESIS

Autores: Madson Allan de Luna Aragao¹; Maira de Arruda Lima¹; Ricardo Duraes de Carvalho¹; Carlos Henrique Bezerra Cruz¹; Roberto Dias Lins Neto¹;

E-mail para correspondência: madsondeluna@gmail.com

Instituições: ¹Aggeu Magalhaes Institute - IAM, Oswaldo Cruz Foundation - FIOCRUZ;

Palavras-chave: Thermodynamics; Evolutionary Pressure; HPV

Apoio: Brazilian Coordination for the Improvement of Higher Education Personnel - CAPES, Brazilian National Council for Scientific and Technological Development - CNPq, Foundation for the Support of Science and Technology of Pernambuco - FACEPE, Oswaldo Cruz Foundation - FIOCRUZ, National Laboratory for Scientific Computing - LNCC

[Trabalho não publicado a pedido dos autores]

Proteínas e proteômica

ANÁLISE MODULAR DE REDE DE INTERAÇÃO PROTEICA NO DESENVOLVIMENTO DE GLIOMAS

Autores: Manuela Rocha de Menezes¹; Raul Emídio de Lima³; Luydson Richardson Silva Vasconcelos³; Herika Karla Negri Brito²; Fábio Lopes de Melo³;

E-mail para correspondência: manurmenezes@gmail.com

Instituições: ¹Instituto de Ciências Biológicas da Universidade de Pernambuco; ²Hospital da Restauração; ³Centro de Pesquisas Aggeu Magalhães;

Palavras-chave: Gliomas; Interação proteica; IDH1

Apoio: Centro de Pesquisas Aggeu Magalhães

Introdução: Glioma é o tipo de tumor cerebral mais comum e corresponde a quarta principal causa de morte oncológica no mundo. A descoberta de marcadores moleculares permitiu um melhor gerenciamento de pacientes com glioma. Mutações nos genes IDH1/2 estão sendo relacionadas ao surgimento desta neoplasia. **Objetivo:** Identificar redes de associação de proteínas funcionais que possam influenciar no desenvolvimento de gliomas. **Metodologia:** Utilizamos o STRING v.10.5, uma ferramenta on-line, projetada para avaliar e integrar interações proteína-proteína. Estas incluem associações diretas (físicas) e indiretas (funcionais), possibilitando realizar a análise de predição computacional, para verificar interações agregadas de outros bancos de dados primários. Avaliamos a correlação interacional desses Genes Diferencialmente Expressos (DEGs). Primeiramente foram desenhados DEGs em STRING e depois utilizou o software Cytoscape 3.6.1 para construir uma rede interação proteína-proteína (PPI), seguindo os parâmetros de número máximo de iteradores=0 e um escore de confiança ≥ 0.4 como critério de corte. Utilizou-se também o Molecular Complex Detection (MCODE) para selecionar módulos da rede PPI no Cytoscape de acordo com o cut-off=0.2, máxima profundidade=100 e k-core=2. **Ademais, foram realizadas análises de** correlação entre níveis de expressão, vizinhança gênica e de padrões de ocorrência nos genomas de várias espécies no STRING. **Resultados:** Gerou-se uma rede de interação proteica a partir da proteína IDH1 com: TERT, ATRX, TP53, p16INK4A, PTEN, EGFR e MGMT. Em análise de predição, incluiu-se outras proteínas na rede por mecanismo de algoritmo computacional que podem ter atuação na mesma via oncogênica (BCL2, MYC, EGF e PIK3CG). Foram encontrados os índices médios de 0.48 e 0.048 para co-expressão e vizinhança gênica, respectivamente para *H. sapiens*. Foi visto que existe uma similaridade entre as sequências de outras espécies, principalmente de demais eucariotos. As análises de interação a presença de alguns genes, destaca-se o MYC que desempenha um papel importante no desenvolvimento cortical e é uma molécula chave para o estabelecimento da gliogênese. PTEN e p53 foram sugeridos como reguladores de MYC no comportamento das células-tronco no glioma. **Conclusão:** A presente pesquisa oferece suporte à ideia do envolvimento da proteína IDH1 no processo de aparecimento do câncer em células do sistema nervoso e. Assim como, amplia os conhecimentos para novos alvos terapêuticos.



Proteínas e proteômica

CARACTERIZAÇÃO *IN SILICO* E ESTUDOS DE INTERAÇÃO MOLECULAR COM O SUBSTRATO DE UMA QUITOSANASE DE *CHROMOBACTERIUM VIOLACEUM*

Autores: Mayara Itala Geronimo de Azevedo¹; José Ednézio da Cruz Freire¹; Thalles Barbosa Grangeiro¹;

E-mail para correspondência: itala_azevedo@hotmail.com

Instituições: ¹Universidade Federal do Ceará;

Palavras-chave: Docking Molecular; Quitosana; Bactéria

Apoio: CNPq, CAPES

Com o seqüenciamento de genoma de *Chromobacterium violaceum*, uma bactéria Gram-negativa, saprófita e de vida livre, várias ORFs codificando proteínas com potencial biotecnológico foram reveladas. Dentre elas, estão várias hidrolases de glicosídeos, incluindo quitinases e quitosanas. Quitosanas (EC 3.2.1.132) são enzimas capazes de degradar a quitosana, um polímero de β -D-glucosamina (GlcN). O produto dessa degradação, os oligômeros de quitosana ou oligômeros de GlcN, são de grande interesse industrial, apresentando atividade antitumoral, anticolesterol, entre outras, além de agir contra microrganismos e serem utilizados na alimentação. Sabendo da importância do conhecimento prévio das características de uma proteína antes da sua aplicação, o presente trabalho teve por objetivo caracterizar *in silico* a quitosanase de *C. violaceum* ATCC 12472 codificada pela ORF Cv3931, assim como realizar análises de interação molecular. Para isso, diversas ferramentas de bioinformática disponíveis gratuitamente foram empregadas. Os resultados das análises da estrutura primária mostraram que Cv3931 possui um peptídeo sinal de 29 resíduos de aminoácidos, apresenta localização extracelular e dois domínios protéicos, um relacionado à família GH46, característico de quitosanas, e outro domínio relacionado a ligação à quitina. As análises físico-químicas indicaram que a proteína é estável, hidrofílica e possui massa molecular (*Mr*) em torno de 35,7 kDa e ponto isoelétrico de 8,54. A massa molecular encontrada é maior do que as massas encontradas para outras enzimas do mesmo grupo e isso está relacionado à presença do domínio de ligação a quitina, que não é comum nos membros da família GH46. Outras análises também mostraram que a quitosanase Cv3931 possui duas pontes dissulfeto, que são importantes para a estabilidade de proteínas secretadas. A partir do modelo da estrutura tridimensional da enzima e de estudos de interação molecular do domínio catalítico com um oligossacarídeo (GlcN)₂-GlcNAc, foi possível observar que a interação do substrato no sítio ativo de Cv3931 é semelhante ao que foi observado para uma quitosanase de *Bacillus circulans*. Essas interações enzima-substrato são mediadas principalmente por pontes de hidrogênio, várias delas intermediadas por moléculas de água. A partir dos modelos de interação, os aminoácidos Glu¹⁰⁹ e Asp¹²⁷ foram identificados como os prováveis resíduos catalíticos, sugerindo que Cv3931 é, muito provavelmente, uma quitosanase funcional.



Proteínas e proteômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE POTENCIAIS PEPTÍDEOS ANTIMICROBIANOS DO TIPO PROTEÍNA TRANSPORTADORA DE LIPÍDEO EM CAJANUS CAJAN COM O USO DA BIOINFORMÁTICA

Autores: Mireli de Santana Rego¹; Carlos André dos Santos Silva²; Lívia Maria Batista Vilela³; Ana Maria Benko Iseppon⁴;

E-mail para correspondência: mirelisantana@gmail.com

Instituições: ¹Universidade Federal de Pernambuco; ²Universidade Federal de Pernambuco; ³Universidade Federal de Pernambuco; ⁴Universidade Federal de Pernambuco;

Palavras-chave: LTPs; *Cajanus cajan*; AMPs

Apoio: Laboratório genética e biotecnologia vegetal (LGBV)

As proteínas transportadoras de lipídeos (LTPs) são definidas principalmente pela capacidade de facilitar a transferência de fosfolipídios entre membranas. Elas possuem um peptídeo sinal, oito resíduos de cisteína altamente conservadas, e podem ser divididas com base em sua massa molecular em LTP1(9 kDa) e LPT2(7 kDa). Em plantas exercem a função de proteção contra infecções microbianas. Sua atividade foi verificada também no *Pigeonpea* (*Cajanus cajan*), uma leguminosa da família Fabaceae de grande importância econômica no Brasil utilizada na alimentação animal, forragem e componente de mistura em silagens. O presente trabalho identificou e caracterizou potenciais LTPs em *Cajanus cajan* com o uso de ferramentas de bioinformática. Para tanto, foi realizada uma busca no banco de dados PhytAmp, utilizando o termo *lipid transfer protein*. Em seguida, as sequências obtidas no PhytAmp foram alinhadas contra o banco de dados de *Cajanus cajan*, disponível no GenBank através da ferramenta BLASTp. Os candidatos encontrados foram caracterizados observando a presença de peptídeo sinal (SignalP), identificando os domínios conservados (CDSearch), analisando a conservação dos resíduos de cisteínas na formação de pontes dissulfeto (Disulfind e DIANNA) e avaliando a prospecção de atividade antimicrobiana através da ferramenta CAMP_{R3}, com quatro algoritmos diferentes. Em seguida, as sequências foram modeladas (SWISS-MODEL), seguidas de refinamento. A qualidade da modelagem foi observada antes e depois do refinamento com o gráfico de Ramachandran. A partir dos resultados obtidos foram localizadas 30 sequências de LTPs no banco de dados PhytAmp. Após o alinhamento com o genoma disponível de *Cajanus cajan* (Pigeonpea-Version-2.0), foram obtidas 31 sequências candidatas. Ao caracteriza-las observamos que os 29 candidatos possuíam peptídeo sinal, 23 possuíam o domínio LTP1 e 9 o domínio LTP2. No que se refere ao número de resíduos de cisteínas 8 sequências apresentaram 8 cisteínas nas suas posições conservadas (C1-C4; C2-C6; C3-C7; C5-C8), além disso todas as sequências candidatas apresentaram predição para atividade antimicrobiana. Para a modelagem, pôde-se observar que na maioria dos casos a resolução (em Ångström) do homólogo resolvido era mais importante do que a porcentagem de similaridade entre as sequências. Desta forma, as análises *in silico* permitiram selecionar potenciais candidatos antimicrobianos de LTPs no genoma de *Cajanus Cajan*.



Proteínas e proteômica

ANÁLISE IN SILICO DA FENILALANINA AMÔNIA-LIASE DE CANA-DE-AÇÚCAR SOB SIMBIOSE COM *GLUCONACETOBACTER DIAZOTROPHICUS*

Autores: Rayssa Guedes Gomes da Silva¹; Fabiana Aparecida Cavalcante Silva¹; Tercilio Calsa Junior¹;

E-mail para correspondência: rayssa.guedesgsilva@gmail.com

Instituições: ¹Laboratório de Genômica e Proteômica de Plantas, Departamento de Genética - Centro de Biociências - Universidade Federal de Pernambuco;

Palavras-chave: *Saccharum* spp.; Fixação biológica do nitrogênio; Melhoramento vegetal

Apoio: CNPq

A cana-de-açúcar (*Saccharum* spp.), membro perene de Poaceae, possui metabolismo C_4 e é a principal fonte agroindustrial de açúcar e etanol no Brasil. No cultivo, a fixação simbiótica de nitrogênio (FBN) tem sido alternativa ao uso de fertilizantes, reduzindo gastos com insumos. A enzima fenilalanina amônia-liase (PAL) atua na biossíntese de fenilpropanoides, no crescimento vegetal e na resposta a fatores abióticos e bióticos tais como microrganismos simbioses para a FBN. Assim, objetivou-se analisar *in silico* a proteína fenilalanina amônia-liase identificada por proteômica diferencial e espectrometria de massas (MS) em colmo da cana-de-açúcar sob simbiose com *Gluconacetobacter diazotrophicus*. A sequência mostrou alta similaridade com proteína ortóloga de *Saccharum* sp. (M1MQ13), utilizada como sonda no tBLASTn contra *GenBank*. As sequências similares encontradas foram então analisadas nos programas ORFfinder (confirmação do quadro de leitura) e CD-Search (definição de domínios conservados). Também foram examinadas quanto ao pI e massa molecular (MM) com a ferramenta ProtParam, presença de pontes dissulfeto com o DISULFIND e peptídeo sinal pelo SignalP 4.1, assim como a predição da localização sub-celular no ProtComp 9.0 e alinhamento múltiplo das sequências através do software online Clustal Omega. Foi realizada modelagem comparativa para obter estruturas terciárias presumíveis pelo SWISS-MODEL. Via tBLASTn foram selecionadas (*e-value* = 0) quatro sequências dos seguintes acessos: XM_002454153.2, NM_001111864.2, XM_004953096.4 e XM_015769634.1, de *Sorghum bicolor*, *Zea mays*, *Setaria italica*, e *Oryza sativa*. No CD-Search foi observado domínio PLN do aminoácido 1 ao 701, *e-value* 0. A MM das sequências variou de 75,16 a 75,60 kDa, enquanto pI foi de 5,86 a 6,07. Nenhuma sequência apresentou pontes de dissulfeto ou peptídeo sinal. A localização subcelular da PAL é presumível no citosol. O alinhamento mostrou regiões conservadas e a modelagem por homologia revelou identidade entre 90,84% e 99,86% com o modelo 6at7.1.A. Assim, a análise *in silico* confirmou a sequência como ortóloga a PAL de cana-de-açúcar e será validada transcricionalmente.



Proteínas e proteômica

VALIDAÇÃO *IN SILICO* DA PROTEÍNA AFP3 (ABI FIVE BINDING PROTEIN 3) DE CANA-DE-AÇÚCAR (SACCHARUM SPP.) EM RESPOSTA AO DÉFICIT HÍDRICO

Autores: Rodrigo Rocha da Silva¹; Elton Pedro Nunes Pena¹; Rayssa Guedes Gomes da Silva¹; Sheyla Carla Barbosa da Silva Lima¹; Fabiana Aparecida Cavalcante Silva¹;

E-mail para correspondência: rodrigodogc2010@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: caracterização; Poaceae; bioinformática

Apoio: CNPq e FACEPE

A cana-de-açúcar é uma gramínea de grande importância econômica, devido ao alto teor de açúcares redutores em seu colmo, sendo esta, a principal fonte de sacarose e etanol no Brasil. O ácido abscísico (ABA) na maioria dos vegetais, é um fitormônio associado com a regulação de vários aspectos fisiológicos, como regulação da abertura estomática. O acúmulo de proteínas relacionadas à regulação do ABA pode ser influenciado por estresses abióticos, estando ligadas aos mecanismos de tolerância ao déficit hídrico. Uma análise proteômica diferencial da fração foliar rica em cloroplastos de cana-de-açúcar demonstrou que o acúmulo de AFP3 (ABI five binding protein 3 - A0A059PZR7), uma proteína reguladora do ABA, foi 1,89 vezes maior na condição de déficit hídrico do que no controle. Portanto, o objetivo deste trabalho foi validar por meio de ferramentas de bioinformática a proteína AFP3 de cana-de-açúcar. O acesso A0A059PZR7 foi utilizado como sonda e submetido à ferramenta tBlastn. Inicialmente, foram identificados os quadros de leitura e domínios conservados por meio das ferramentas ORFinder e CD-search, respectivamente. Em seguida, foram utilizadas as seguintes ferramentas: Signal-P 4.1 para predição e localização de peptídeo sinal, Disulfind para a detecção de pontes dissulfeto, ProtParam para analisar as sequências quanto ao ponto isoelétrico (pI) teórico e massa molecular (MM), Philius para a predição de transmembrana e ProtComp para determinação da localização subcelular. Posteriormente, utilizou-se o SWISS-MODEL para a realização do modelo por homologia. Foram obtidas 4 sequências pelo tBlastn (XM_002465599.2 [0,0], XM_008659854.3 [0,0], XM_004985139.3 [1,00E-110], XM_015777506.1 [4,00E-92]), nas quais foram localizados 3 domínios conservados em comum com a sonda: NINJA_B, EAR e JAS. Estes estão relacionados a proteínas que atuam como repressor transcricional, sendo o domínio EAR responsivo a etileno e o domínio JAS responsivo a jasmonato, que está envolvido em mecanismos de defesa a estresses. A caracterização da proteína AFP3 revelou ausência de peptídeo sinal e pontes dissulfeto, com massa molecular de 41,334 e ponto isoelétrico de 8.90, com característica globular localizada no núcleo e modelagem predita de 24 % de identidade com o modelo (PDB:1ufn.1). Os resultados obtidos permitiram validar *in silico* a sequência da proteína AFP3, contudo faz-se necessário uma análise a nível transcricional da expressão gênica correlacionando-a ao estresse aplicado.



Proteínas e proteômica

IDENTIFICAÇÃO DE LECTINAS DE CANDIDA ALBICANS E MAPEAMENTO DE SEUS DOMÍNIOS CONSERVADOS EM UMA ANÁLISE IN SILICO

Autores: Rosilda Cintra de Souza¹; Jéssica Barboza da Silva²; Carlos André dos Santos Silva²; Kyria Santiago Nascimento³; Cíntia Renata Costa Rocha¹;

E-mail para correspondência: rosildacindra@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco, LIKA; ²Universidade Federal de Pernambuco, LGBV; ³Universidade Federal do Ceará, UFC;

Palavras-chave: Lectinas; Candida albicans; Bioinformática

Apoio: Capes; UFPE-LIKA; UFPE-LGBV; UFC

Introdução: Lectinas são proteínas ubíquas de origem não-imune e compõem um grupo específico dentro de uma classe maior que são as adesinas, envolvidas em muitos processos essenciais de reconhecimento celular e molecular. O que as difere de adesinas não-lectínicas é sua capacidade de aglutinar células e glicoconjugados de maneira reversível e específica a açúcares livres ou glicanos presentes em glicoproteínas ou glicolipídeos, sem alterar suas estruturas. As lectinas tem sido alvo de pesquisas científicas devido às suas inúmeras aplicações biotecnológicas. Contudo, lectinas de fungos ainda são pouco estudadas e pouco se sabe sobre a história evolutiva e suas funções nos processos de instalação e desenvolvimento de infecções fúngicas. A disponibilidade de sequências completas do genoma de fungos patogênicos torna possível a exploração *in silico* destes dados visando o entendimento dos sistemas biológicos. **Objetivos:** Identificar através da análise de genoma *in silico* sequências com potencial codificador de lectinas em *Candida albicans*, uma vez que até o presente momento são desconhecidas. **Metodologia:** Foram selecionadas 105 sequências sonda de três famílias gênicas (L-type, CBM e FLO) no banco de dados do UniProt e alinhadas via tBLASTn contra o genoma de *C. albicans* depositado no NCBI, adotando-se e-value $\leq 0,05$ e similaridade $\geq 30\%$. As sequências foram coletadas, traduzidas e seus domínios identificados com o auxílio do Transdecoder e CD-search, respectivamente, além de realizadas a predição do ponto isoelétrico (pI) e peso molecular utilizando o JVirGel 2.0, e a localização subcelular através do Cell-PLoc 2.0. **Resultados e Discussão:** Após tradução e análise de domínio, foram identificadas 12 sequências com domínios completos (oito com o domínio Flocculin-t3, duas com o domínio *Lectin_leg-like*, uma com o domínio ChtBD1_GH16 e uma com o domínio Flo11). As candidatas apresentaram p.I entre 3.38 e 9.31, PM variando entre 22 e 111 KDa e localização subcelular na membrana celular e extracelular. Candidatas com o domínio FLO também podem ocorrer na parede celular, citoplasma e núcleo e no caso das L-type, também serem encontradas no retículo endoplasmático e complexo de golgi. Esses resultados corroboram com dados descritos na literatura para essas proteínas. **Conclusão:** Nossos resultados podem contribuir para uma melhor compreensão das características moleculares das famílias gênicas FLO, CBM e L-type e sua participação no processo de patogenicidade de fungos.



Proteínas e proteômica

DETERMINAÇÃO DE ESTRUTURA TRIDIMENSIONAL DA LECTINA DE SEMENTES DE PLATYPODIUM ELEGANS POR CRISTALOGRAFIA DE RAIOS X E FERRAMENTAS DE BIOINFORMÁTICA

Autores: Sarah Elizabeth Gomes Correia¹; Vinícius José da Silva Osterne²; Vanir Reis Pinto-Júnior³; Corneville Correia-Neto⁴; Cláudia Figueiredo Lóssio⁵; Jorge Luis Almeida Correia⁶; Ivanice Bezerra da Silva⁷; David de Alencar Araripe⁸; Benildo Sousa Cavada⁹; Kyria Santiago Nascimento¹⁰;

E-mail para correspondência: segc22@live.com

Instituições: ¹Universidade Federal do Ceará; ²Universidade Federal do Ceará; ³Universidade Federal do Ceará; ⁴Universidade Federal do Ceará; ⁵Universidade Federal do Ceará; ⁶Universidade Federal do Ceará; ⁷Universidade Federal do Rio Grande do Norte; ⁸Universidade Federal do Ceará; ⁹Universidade Federal do Ceará; ¹⁰Universidade Federal do Ceará;

Palavras-chave: Lectina; Cristalografia; Docking Molecular

Apoio: UFC, CNPq, CAPES, FUNCAP, INCT-NanoBioSimes

Uma lectina nativa (nPELa), purificada de sementes da espécie *Platypodium elegans*, da tribo Dalbergieae, foi cristalizada e estruturalmente caracterizada por ferramentas de cristalografia de difração de raios X e bioinformática. Os cristais obtidos difrataram para uma resolução de 1,6 Å e a estrutura nPELa foi resolvida através de substituição. Além disso, o nPELa possui um sítio de ligação de metal e um domínio de reconhecimento de carboidratos conservado (CRD) semelhante a outras lectinas da tribo Dalbergieae, tais como PAL (*Pterocarpus angolensis*) e CTL (*Centrolobium tomentosum*). A análise de ancoragem molecular indicou alta afinidade desta lectina para diferentes manosídeos, principalmente trimanosídeos, formados por ligações glicosídicas de -1,3 ou -1,6, evidenciados pelos escores obtidos. Ademais, foram realizadas simulações de dinâmica molecular para demonstrar o comportamento estrutural de nPELa em solução aquosa. Em solução, nPELa foi altamente estável, e modificações estruturais em seu sítio de reconhecimento de carboidratos permitiu a interação entre a lectina e os diferentes ligantes. Diferentes modificações foram observadas durante as simulações de cada um dos glicanos, que incluíram ligações de hidrogênio e interações hidrofóbicas através de alterações nos resíduos relevantes.



Proteínas e proteômica

IDENTIFICAÇÃO DE NODULINAS PRECOSES E CARACTERIZAÇÃO IN SILICO DA FAMÍLIA RWP-RK EM *STYLOSANTHES SCABRA*

Autores: Sheyla Carla Barbosa da Silva Lima¹; João Pacífico Bezerra-Neto²; Valesca Pandolfi²; Ana Maria Benko-Iseppon²; Tercilio Calsa Junior¹;

E-mail para correspondência: sheylac.b@gmail.com

Instituições: ¹Universidade Federal de Pernambuco, Laboratório de Genômica e Proteômica de Plantas, Recife-PE, Brasil; ²Universidade Federal de Pernambuco, Laboratório de Genética e Biotecnologia Vegetal, Recife-PE, Brasil. ;

Palavras-chave: Bioinformática; Simbiose; Fixação de nitrogênio

Apoio: CAPES, CNPq.

Stylosanthes scabra Vogel (Fabaceae) é uma forrageira nativa do semiárido brasileiro, a qual apresenta excelente potencial fixador de nitrogênio atmosférico através de associação simbiótica. Este trabalho teve como objetivo minerar e identificar possíveis nodulinas presentes no primeiro estágio de nodulação. Foram selecionadas 28 sondas proteicas de genes codificantes de nodulinas descritos como relacionados ao processo de infecção. As sondas foram confrontadas contra o proteoma conceitual de *S. scabra* (obtido a partir da tradução de seu transcriptoma), utilizando a ferramenta tBLASTn, com *cut-off* de e^{-10} . Em seguida foi feito o resgate das sequências de aminoácidos com a ferramenta Galaxy, e a caracterização dos domínios conservados com a ferramenta Batch/NCBI. Após remoção de redundâncias, as possíveis nodulinas foram classificadas por grupos. As sequências foram analisadas via SMART para confirmação da estrutura do domínio e intervalo, sendo posteriormente submetidas ao MEGA5 para alinhamento múltiplo, bem como ao SignalP para identificação de peptídeo sinal. Finalmente foram confrontadas contra o GenBank, usando o BLASTp. A partir do tBLASTn foram obtidos 15.843 alinhamentos com nodulinas precoces, dos quais, após retirada de redundância, restaram 4.589, sendo 2.292 com domínio completo. As nodulinas-candidatas foram classificadas em 49 superfamílias, destacando-se a RWP-RK cujo motivo é conservado na região C-terminal, considerado típico de proteínas vegetais envolvidas no desenvolvimento controlado por nitrogênio, tais como as proteínas do tipo NIN (*Nodule Inception*). Foi observado que seis sequências apresentaram entre 409 e 875 aminoácidos, com domínio conservado flanqueando o motivo RWP-RK exibindo 48 aminoácidos, sendo este motivo verificado nas seis sequências candidatas desta superfamília. O resultado do BLASTp reportou para as seis sequências a proteína do tipo NIN (NLP), nos acessos XP_015935378.1 (*Arachis duranensis*), XP_015967403.1 (*A. duranensis*), XP_020966948.1 (*A. ipaensis*), XP_020216193.1 (*Cajanus cajan*), XP_015934088.1 (*A. duranensis*) e XP_016193058.1 (*A. ipaensis*), todas com *e-value* 0.0. Em nenhuma das seis sequências foi detectado peptídeo sinal. Os resultados *in silico* confirmam que as sequências caracterizadas podem estar envolvidas no processo de nodulação em *S. scabra*, o que posteriormente deverá ser validado experimentalmente.



Proteínas e proteômica

ANÁLISE IN SILICO DA PROTEÍNA EDR2-X1 (ENHANCED DISEASE RESISTANCE 2) DA PALMA FORRAGEIRA (*OPUNTIA STRICTA*) SOB ATAQUE DE COCHONILHA-DE-ESCAMAS (*DIASPIS ECHINOACTI*)

Autores: Suelen Cristina Lourenço de Barros¹; Mara Danielle Silva do Carmo Santana¹; Rayssa Guedes Gomes da Silva¹; Fabiana Aparecida Cavalcante Silva¹; Tercílio Calsa Junior¹;

E-mail para correspondência: suelen_barros@outlook.com.br

Instituições: ¹Laboratório de Genômica e Proteômica de Plantas, Universidade Federal de Pernambuco (UFPE), Av. Prof. Moraes Rego s/n, Cidade Universitária, CEP: 50670-901, Recife, PE, Brasil.;

Palavras-chave: Palma forrageira; cochonilha-de-escama; Proteína EDR2-X1

Apoio: CNPq, IPA, Facepe

As espécies do gênero *Opuntia*, apesar de exóticas, apresentam grande adaptabilidade ao nordeste brasileiro, servindo como importante fonte alternativa de alimentação e economia de água para rebanhos de caprinos e bovinos. No entanto, o isento cochonilha-de-escamas (*Diaspis echinocacti* Bouché, 1833) tem atacado a espécie e prejudicado sua produtividade, o que tem levado a diversos estudos para melhor compreender esta interação planta-patógeno. O Trabalho anteriormente realizado avaliou o proteoma da palma forrageira sob ataque da cochonilha-de-escamas e dentre as centenas de proteínas identificadas destaca-se a EDR2- X1 (enhanced disease resistance 2-like isoform X1), ela atua como regulador negativo da morte celular, especificamente a morte celular provocada pelo ataque do patógeno e mediada pela via de defesa do ácido salicílico. O objetivo deste trabalho foi caracterizar *in silico* a proteína EDR2-X1 como uma forma de validação de análise proteômica. Uma vez que a proteína foi identificada a partir de similaridade com proteína de *Nelumbo nucifera*, esta foi utilizada como sonda. Foram obtidas três sequências mais similares à sonda a partir do Genbank/NCBI, das quais se realizou tBLASTn (Genbank/NCBI) para obtenção da sequência proteica. Foi utilizada a ferramenta ORF Finder para verificar o quadro de leitura, onde, seguiu-se com análises: 1) CD-Search para buscar o domínio e confirmar sua localização; 2) Signal P 4.1 em identificação de peptídeo sinal; 3) DISULFIND usado em observação de pontes dissulfeto; 4) PROTCOMP 9.0 identificação da localização sub-celular da proteína; 5) Philius para identificar a forma da proteína transmembrana; 6) O EXPASY permite o cálculo de vários parâmetros físicos e químicos para uma determinada proteína; 7) SWISS-MODEL na modelagem por homologia. Foram identificados dois domínios, o PH e START. Não foram observados peptídeo sinal para sequência, incluindo a sonda, assim como pontes dissulfeto. Foi observado no meio extracelular um score com média de 2,4, a proteína foi identificada como globular e os valores médios de PI e Massa Molecular foram de 7,7 e 79827,72, respectivamente. A sequência desta forma foi computacionalmente confirmada como provável ENHANCED DISEASE RESISTANCE 2-like isoform X1 de palma forrageira. Os resultados obtidos são bastante promissores e servirão como base para análises em nível transcricional posterior, para uma melhor avaliação em resposta ao estresse aplicado.



Proteínas e proteômica

HISTONA H2B.9 DE ALGODÃO COLORIDO: UMA ABORDAGEM BIOINFORMÁTICA

Autores: Tainara Fernandes Dantas¹; Geisenilma Maria Gonçalves da Rocha¹; Sheyla Carla Barbosa da Silva Lima¹; Rayssa Guedes Gomes da Silva¹; Fabiana Aparecida Cavalcante Silve¹; Tercilio Calsa Junior¹;

E-mail para correspondência: tainarafd85@gmail.com

Instituições: ¹Laboratório de Genômica e Proteômica de Plantas, Universidade Federal de Pernambuco (UFPE), Av. Prof. Moraes Rego, Cidade Universitária, CEP: 50670-901, Recife, PE, Brasil.;

Palavras-chave: Proteína; Validação ; Gossypium

Apoio: CNPq

O algodão (*Gossypium hirsutum*) é um produto influente na economia mundial principalmente por ser a mais importante fonte natural de fibras e por sua ampla forma de comércio e aplicação essencialmente na indústria têxtil e na extração de óleo das sementes para produtos alimentícios, lubrificantes e produção de biodiesel. A análise proteômica preliminar da cultivar BRS Rubi possibilitou a anotação presumível da proteína histona H2B.9 (A2Y7R3) em condições de estresse por radiação UVB. Como principal componente do nucleossomo, as histonas são responsáveis por regular a acessibilidade ao DNA. Assim, as histonas desempenham papel fundamental na regulação da transcrição, no reparo e replicação do DNA, além de promover a estabilidade cromossômica. O objetivo desse trabalho foi caracterizar *in silico* a sequência da proteína histona H2B.9 de algodão. A sequência do acesso A2Y7R3 foi utilizada como sonda para a realização de tBLASTn no Genbank NCBI e considerou-se o alinhamento apenas para a família Malvaceae. Em seguida foram utilizadas as ferramentas ORF Finder para verificação do quadro de leitura; CD-Search para observação de domínios conservados; SignalP 4.1 para identificar a presença de peptídeo sinal; DISULFIND para identificar as pontes dissulfeto; PHILIUS para predição transmembrana; ProtComp para localização subcelular; ProtParam para verificar o peso molecular e o ponto isoelétrico teórico da proteína; por fim, o Swiss-Model para modelagem por homologia de sequências da proteína H2B.9 com o banco. Com os resultados do tBLASTn foi observado para os alinhamentos com *Gossypium hirsutum* (AF025667.1) e *Gossypium raimondii* (XM_012619037.1) identidades de 85% e E-value de 2×10^{-48} e 3×10^{-48} respectivamente. Nas ferramentas DISULFIND e SignalP 4.1 foi possível observar que a proteína não apresenta pontes dissulfeto e peptídeo sinal, nessa ordem. O resultado do PHILIUS demonstrou que a proteína apresenta característica do tipo globular e a localização subcelular da proteína analisada pelo ProtComp a definiu como proteína nuclear, corroborando com a literatura existente. Os valores da modelagem no Swiss-Model mostraram que a proteína possui um grau de identidade acima de 73%. Dessa forma, o presente trabalho refere-se a um estudo preliminar *in silico* sobre a proteína histona H2B.9. Subsequentemente, essa proteína pode contribuir para estudos aprofundados sobre processos de regulação transcricional.



Proteínas e proteômica

SURFACEOMA BACTERIANO DO TRATO GASTROINTESTINAL

Autores: Tayná da Silva Fiúza¹; Gustavo Antonio De Souza¹; João Paulo Matos Santos Lima¹;

E-mail para correspondência: taynafiuza2@gmail.com

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: surfaceoma; trato gastrointestinal; microbiota

Apoio: CAPES

A maior parte dos microorganismos que reside no corpo humano se encontra no trato gastrointestinal (TGI), sendo ele uma das mais amplas interfaces entre indivíduo, microorganismos, fatores ambientais e antígenos. Esses microorganismos realizam funções na manutenção da integridade da barreira mucosa, provisão de vitaminas, proteção contra patógenos, auxílio em etapas da digestão e outras atividades essenciais. Distúrbios na composição dessa microbiota podem causar desregulação em células da imunidade adaptativa, posteriormente levando a desordens inflamatórias. Neste trabalho buscou-se produzir um rascunho do surfaceoma bacteriano do trato gastrointestinal utilizando espécies já cultivadas em laboratório. Os dados acerca das espécies presentes na microbiota gastrointestinal humana foram obtidos a partir do trabalho de RAJILI?-STOJANOVI? e DE VOS (2014) e posteriormente utilizados para obtenção do proteoma de cada uma das linhagens das referidas espécies bacterianas através da plataforma NCBI (<ftp.ncbi.nlm.nih.gov/>). Os proteomas foram tratados para a remoção de redundâncias e em seguida foi utilizado o software TMHMM, para predição de domínios transmembranares em proteínas. Proteínas transmembranares cuja presença em membranas de organelas já era sabida foram retiradas do grupo de análise utilizando o banco Gene Ontology. De acordo com a previsão do TMHMM, das 12304281 proteínas analisadas, 1929699 possuem domínios transmembranares (15.68%) e após a filtragem feita com o Gene Ontology restaram 1859706 proteínas que compõem então a atual proposta de surfaceoma microbiano do trato gastrointestinal. O surfaceoma sugerido conta com 124 gêneros bacterianos e será utilizado para melhor compreensão de estados de saúde e doença em seres humanos, sendo explorada no próximo trabalho uma possível relação entre sua composição, seu potencial imunogênico e o desencadeamento das doenças inflamatórias do trato gastrointestinal como doença de Chron e doença celíaca.



Proteínas e proteômica

EVIDÊNCIA IN-SILICO DO POTENCIAL DE DESINTOXICAÇÃO DO HERBICIDA BISPIRIBAQUE-SÓDICO POR UM MODELO TEÓRICO DA PROTEÍNA GLUTATIONA S-TRANSFERASE CLASSE TAU 5 EM ORYZA SATIVA

Autores: Vinícius Costa Amador¹; Ravenna Lins Rodrigues²; Felipe França de Oliveira²; Rafael Trindade Maia²;

E-mail para correspondência: vinicius.costa.amador@gmail.com

Instituições: ¹Universidade Federal Rural de Pernambuco; ²Universidade Federal de Campina Grande;

Palavras-chave: Arroz; Bioinformática; Biologia Computacional

Apoio: BCT (Biologia Computacional e Teórica) CDSA (Centro de Desenvolvimento Sustentável do Semiárido) CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) NemPe (Núcleo de estudos em melhoramento genético de plantas de Pernambuco)

O arroz (*Oryza sativa* L.) é considerado o cereal mais importante do mundo, sendo alimento primário de mais da metade da população mundial, principalmente de países asiáticos em desenvolvimento. A principal limitação do seu cultivo é a competição com plantas daninhas por água, luz e nutrientes, cuja solução tem sido o uso de herbicidas, que podem promover decréscimo produtivo da cultura. Sabe-se que a superfamília de enzimas Glutathione S-transferase confere em arroz proteção a estresse biótico e abiótico e resistência à alguns herbicidas, através da conjugação do tripeptídeo glutathione a um composto hidrofóbico, tornando-o mais solúvel e passivo de desintoxicação. Neste âmbito, a bioinformática é usada para alcançar informações acerca dessa interação molecular. Este trabalho objetiva estudar o atracamento molecular entre um modelo estrutural teórico de uma proteína Glutathione S-transferase Tau 5 de arroz (OsGSTU5) ao herbicida bispiribaque-sódico, verificando a possibilidade da proteína em questão se apresentar como agente de desintoxicação. A sequência primária da OsGSTU5 foi obtida na base de dados do NCBI (ID: AAG32470.1) alinhada contra a base de dados do PDB (*Protein Data Bank*) através do BLAST (*Basic Local Alignment Tool*), obtendo-se o *template*, PDB ID: 5j4u com identidade de 55% e resolução de 1.249Å. Por meio do servidor SWISS-MODEL foi gerado o modelo tridimensional por homologia e validado por análise gráfica do programa PROCHECK para avaliar sua qualidade estereoquímica, em seguida por cálculos de energia, ANOLEA e GROMOS. O *docking* semi-rígido foi executado no programa Autodock 1.5.6 sob o campo de forças AD4, através do algoritmo genético lamarckiano. O grid estabelecido foi 126x126x126 Å, centrado no receptor. Os parâmetros usados na simulação foram: 10.000 réplicas, análises de energia a cada 1.500.000 e 27.000 gerações, tamanho populacional de 150 e taxas de mutação e crossing-over de 0,02 e 0,08 respectivamente. Foram analisados as 10 possíveis conformações ranqueadas pelo programa no visualizador Visual Molecular Dynamics. O modelo teórico desenvolvido, apresentou-se adequado para a simulação de atracamento. Foram identificadas uma afinidade do ligante por Lisinas como resíduos âncora e uma possível evidência de desintoxicação, por parte da OsGSTU5 quanto ao bispiribaque-sódico. Os resultados revelam um potencial alvo para o processo de melhoramento genético, visando a tolerância ou resistência à herbicidas em arroz.

RNA e transcriptômica

MIRNOMA REVELA NOVOS INSIGHTS SOBRE A BIOLOGIA MOLECULAR DO CAMPO DE CANCERIZAÇÃO GÁSTRICO

Autores: Adenilson Pereira¹; Fabiano Moreira^{1,2}; Tatiana Vinasco-Sandoval¹; Amanda Vidal¹; André Ribeiro-dos-Santos¹;

E-mail para correspondência: adenilson.leao@hotmail.com

Instituições: ¹Programa de Pós-Graduação em Genética e Biologia Molecular, Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Avenida Augusto Corrêa, 66075-110, Belém, Pará, Brasil.; ²Programa de Pós-Graduação em Oncologia e Ciências Médicas, Núcleo de Pesquisa em Oncologia, Instituto de Ciências da Saúde, Universidade Federal do Pará, Rua dos Mundurucus, 66073-000, Belém, Pará, Brasil.;

Palavras-chave: miRNome; Gastric Cancer; Field Cancerization

Apoio: CAPES/Biologia Computacional: No. 3381/2013/CAPES (Rede de Pesquisa em Genômica Populacional Humana); UNIVERSIDADE FEDERAL DO PARÁ - UFPA

MicroRNAs play an important role in gastric carcinogenesis and have been associated with field cancerization in gastric cancer; however, the role of these molecules is not fully understood in this process. We performed miRNome sequencing in MiSeq System (Illumina) of gastric samples: non-cancer (n=15), paired gastric cancer (n=15) and adjacent to cancer (n=15), to understand the involvement of miRNAs in gastric field cancerization. Differential expression analysis was performed using the bioconductor-DESeq2 package and consider $|\log_2(\text{fold change})| > 2$ and adjusted P -value < 0.05 by FDR as parameter. We also evaluated the biomarkers potential of the studied miRNAs by using receiver operating characteristic (ROC) and the area under curve (AUC > 0.85). The identification of differentially expressed miRNA target genes was performed using miRTargetLink Human tool and the functional annotation was made with DAVID v.6.8. We found fifteen miRNAs exclusively deregulated in gastric cancer, of which six have potential diagnostic value for advanced disease. Furthermore, we found eleven miRNAs exclusively deregulated in adjacent tissue, with ten up-regulated tumor suppressors miRNAs (TS-miRs). Of these, five have potential diagnostic value for the early stages of gastric cancer. The functional analysis of these TS-miRs demonstrates that they regulate important cellular signaling pathways (PI3K-Akt, HIF-1, Ras, Rap1, ErbB and MAPK signaling pathways), which are involved in gastric carcinogenesis. Five miRNAs were differentially expressed in both gastric cancer and adjacent regarding to non-cancer tissues; among them, *hsa-miR-200a-3p* and *hsa-miR-873-5p* have potential diagnostic value for early and advanced stages of the disease. Only *hsa-miR-196-5p* was differentially expressed between adjacent to cancer and gastric cancer tissues. In addition, the other miRNAs identified in this study were not differentially expressed between adjacent to cancer and gastric cancer, suggesting that these tissues share these changes. Our results show that gastric cancer and adjacent tissues have an expression profile for miRNAs analyzed very similar, indicating that miRNAs studied are intimately associated with field cancerization in gastric cancer. The overexpression of TS-miRs in adjacent tissue is likely an antitumor mechanism. Additionally, these miRNAs have great accuracy in discriminating non-cancer from other tissues and can be potentially useful as biomarkers for gastric cancer.

RNA e transcriptômica

PERFIL DE EXPRESSÃO DIFERENCIAL E CARACTERIZAÇÃO ESTRUTURAL DE FATORES DE TRANSCRIÇÃO WRKY EM VIDEIRA SOB ESTRESSE BIÓTICO

Autores: Andreia Cristiny Cezarino de Araújo¹; Mireli de Santana Rêgo¹; Roberta Lane de Oliveira Silva¹; João Pacífico Bezerra Neto¹; Flávia Figueira Aburjaile¹;

E-mail para correspondência: andreiacristinyac@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Vitis; Xanthomonas campestris; Defesa vegetal

Apoio: CAPES, CNPq e FACEPE.

As proteínas WRKY representam uma família de fatores de transcrição envolvidos na regulação de diversos processos fisiológicos e moleculares, como desenvolvimento, senescência e resposta ao estresse. Apesar do número crescente de estudos sobre genômica e transcriptômica envolvendo a família WRKY em diversas espécies, ainda há pouca informação sobre sua regulação na videira quando infectada pela bactéria *Xanthomonas campestris* pv. *viticola*. Assim, o objetivo desse estudo foi identificar e caracterizar transcritos WRKY diferencialmente expressos em acessos de videira contrastantes quanto à resistência à *X. campestris* pv. *viticola*. Inicialmente, sequências sondas de WRKY foram selecionadas utilizando bancos de dados públicos (NCBI e UniProt) aplicando as palavras chaves “WRKY” e “Vitis”. Tais candidatos foram alinhados contra o banco de RNA-Seq de videira através da ferramenta tBLASTn. As sequências identificadas como WRKY foram traduzidas utilizando o Transdecoder e seus domínios funcionais foram identificados por meio da ferramenta Batch CD-search e alinhadas com auxílio de MEGA 7. Em seguida, foi realizada uma busca por motivos conservados na literatura e no programa MEME suíte 4.12.0, bem como a predição do ponto isoelétrico (pI) e do peso molecular (MW) via JVirGel 2.0. Posteriormente, foram identificados os *loci* associados aos WRKY identificados, utilizando para tal o genoma de *Vitis vinifera* disponível no Phytozome v12.1. Foram identificados 89 candidatos apresentando o domínio WRKYGQK completo e cinco motivos conservados (DGYN, PRSYYKCT, DG, YYRC e TYEG) desta família no transcriptoma de videira. Com a análise do MEGA 7, as sequências candidatas foram classificadas em três grupos principais de acordo com a quantidade de domínios WRKY e a composição do dedo de zinco (C-C ou C-H) na árvore fenética. Tais sequências apresentaram ponto isoelétrico entre 4,42 e 10,42, peso molecular entre 16,19 e 44,36 kDa e ancoraram em sua maioria nos cromossomos 4, 19 e 21. Três transcritos foram considerados diferencialmente expressos no transcriptoma da videira, sendo dois induzidos em ambos os acessos e um transcrito apresentando indução apenas no acesso resistente. Assim, a identificação de genes WRKY atuantes nas vias de tolerância a bactéria *X. campestris* compreende uma importante fonte de candidatos para validação via RT-qPCR e posterior aplicação no melhoramento genético da videira.



RNA e transcriptômica

CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE β -1,3-GLUCANASES EM FEIJÃO-CAUPI INFECTADO PELO VÍRUS DO MOSAICO SEVERO

Autores: Ayug Bezerra Lemos¹; Roberta Lane de Oliveira Silva¹; João Pacífico Bezerra Neto¹; Ana Maria Benko Iseppon¹;

E-mail para correspondência: ayug.lemoos@gmail.com

Instituições: ¹Departamento de Genética-Centro de Biociências-Universidade Federal de Pernambuco;

Palavras-chave: *Vigna unguiculata*; PR-2; CPSMV

Apoio: Apoio financeiro: CAPES, CNPq, Fapepe.

O feijão-caupi [*Vigna unguiculata* (L.) Walp.] tem grande importância socioeconômica no Brasil devido ao seu valor nutricional, sendo bastante relevante na agricultura familiar nas regiões Norte e Nordeste. No entanto, sua produtividade tem sido avariada devido à ação do vírus do mosaico severo (CPSMV), um dos principais patógenos que acometem a cultura. As proteínas relacionadas à patogênese do tipo 2 (PR-2) ou β -1,3-Glucanases são de extrema importância para a defesa vegetal, destacando-se por atuar na resposta hipersensível (HR) do vegetal, além de possuir atividade antifúngica. O presente trabalho visou caracterizar funcionalmente e estruturalmente transcritos da família gênica PR-2 em acessos contrastantes de feijão-caupi infectados por CPSMV. Para tanto, foram realizadas buscas pelas palavras-chave “*Beta 1,3 glucanase AND Plant*” no banco de dados UNIPROT (www.uniprot.org) para selecionar sequências sonda. Tais sequências foram alinhadas via tBLASTn contra o banco de RNA-seq do feijão-caupi (NordEST) adotando *cut-off* de e^{-05} . As sequências candidatas foram anotadas, traduzidas pelo Transdecoder e tiveram seus domínios identificados pelo CD-search. Alinhamentos múltiplos foram gerados com o auxílio do programa MEGA v.7, enquanto que a predição do ponto isoelétrico (pI) e peso molecular (p.M) foram realizadas a partir do JVirGel 2.0. Por fim, os transcritos foram submetidos a uma análise de ontologia gênica (GO) visando elucidar a função molecular, processo biológico e componente celular. Foram localizadas 75 sequências caracterizadas como PR-2. Após a tradução e análise de domínio, foram identificadas 69 sequências com domínios completos, confirmando a função dos transcritos alvos. O alinhamento múltiplo permitiu identificar 46 resíduos conservados, bem como os motivos Y-Tirosina, L-Leucina, D-Aspartato, G-Glicina, W-Triptofano, P-Prolina em posições similares às relatadas na literatura. Seu ponto isoelétrico variou entre 4,36 e 10,27 e o peso molecular entre 28,77 e 83,13. Em relação aos dados GO, os transcritos mais representativos apresentaram atividade de hidrólise acerca de o-glicosilo, envolvimento no processo metabólico de carboidratos, estando presentes em maior quantidade na região extracelular. Até o momento, as análises permitiram identificar e caracterizar glucanases envolvidas no processo de defesa do feijão-caupi frente ao CPSMV, contribuindo para o desenho de *primers* específicos para PR-2 visando sua validação por RT-qPCR.



RNA e transcriptômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO ESTRUTURAL DE GENES CODIFICADORES DE QUITINASES NO GENOMA DO FEIJÃO-GUANDU [CAJANUS CAJAN (L.) MILLSP.]

Autores: Beatriz Chagas Silva¹; Flávia Tadeu de Araújo²; Mitalle Matos²; Carolline de Jesús Pires²; Flavia Figueira Aburjaile²; Ana Maria Benko Iseppon²; João Pacifico Bezerra Neto²;

E-mail para correspondência: beatrizchagas1999@gmail.com

Instituições: ¹Colégio Militar do Recife, Recife, PE, Brasil.; ²Universidade Federal de Pernambuco, Recife, PE, Brasil.;

Palavras-chave: Bioinformática; Leguminosas; Estresse biótico

Apoio: CNPq, CAPES, FACEPE.

As leguminosas (Fabaceae), representam uma das principais famílias vegetais em termos agrícolas. Entre as espécies que compõe esta família, destaca-se o feijão-guandu (*Cajanus cajan*), espécie amplamente cultivada na Ásia, África e América, sendo uma importante fonte de nutrientes para a alimentação humana e animal, além de ampla empregabilidade como fitorremediadora. Por ser rústico, o feijão-guandu pode ser cultivado em condições adversas de clima e solo, apesar de sofrer severas perdas devido à ocorrência de patógenos. Com isso, ressalta-se a necessidade da compreensão do mecanismo de defesa dos vegetais em termos da resposta de defesa do hospedeiro contra a diversidade de fitopatógenos, como no caso das proteínas PR (*Pathogenesis-Related*), responsáveis pelas maiores variações quantitativas nos teores de proteína solúvel frente a infecção, podendo estar correlacionadas com a resistência não específica do hospedeiro ao patógeno. Dentre essas, encontram-se as enzimas quitinases, representadas por quatro famílias PR-3; PR-4; PR-8 e PR-11, sendo amplamente encontradas na natureza, em uma ampla diversidade de organismos (procariotos e eucariotos). Visto a importância das quitinases na resposta de defesa do vegetal frente a estresses bióticos, o presente trabalho buscou identificar genes codificantes de quitinases no genoma de *C. cajan*, a fim de realizar a caracterização estrutural e diversidade dos mesmos. Para isto, 72 sequências sondas (*seed sequence*) correspondentes a genes de quitinases já caracterizadas, foram utilizadas em um tBLASTn contra o genoma do feijão-guandu, disponível no GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), adotando-se *cut-off* de e^{-05} . Após a busca 99 candidatos a quitinases foram identificados, como o XM_020382083.1 (1191pb; *endochitinase* CH5B). Destas, 34 apresentaram o domínio *GH18_chitinase-like superfamily*, associados ao grupo das quitinases, com 270 aminoácidos de comprimento, em média. A partir destas sequências, foi detectado ponto isoelétrico variando de 4.22 a 9.19 e peso molecular variando de 31.23 a 41.12 kDa. Já sua localização subcelular é relacionada a exportação para o ambiente extracelular, corroborando com sua atividade de defesa contra patógenos. Esta diversidade de transcritos codificantes para quitinase e, associados a defesa contra patógenos surgem como uma ferramenta promissora para o melhoramento do guandu, sendo necessárias novas análises para compreensão deste mecanismo na espécie.

RNA e transcriptômica

ANÁLISE DE TRANSCRITOS HIPERINDUZIDOS EM TUBÉRCULOS DE SOLANUM TUBEROSUM APÓS INFECÇÃO COM PHYTOPHTHORA INFESTANS

Autores: Carolline de Jesús Pires¹; José Ribamar Costa Ferreira-Neto¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: carollinejpires@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: BDVA; Expressão gênica; Estresse biótico

Apoio: CAPES, CNPq e FACEPE

Diversos grupos de pesquisa têm utilizado bancos de dados de valor agregado (BDVA) para comparar seus dados com outros e/ou minerar situações específicas não investigadas em dados publicados. O presente trabalho recai sobre a segunda alternativa. Para tanto, minerou-se informações contidas nos transcriptomas de batata, estudados no manuscrito "*Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-Phytophthora infestans interactions*" (PMID:23702331), disponível na BDVA *Expression Atlas* (código "E-MTAB-4301"). Dessa forma, objetivou-se analisar transcritos hiperinduzidos (*Adjusted p-value* ≤ 0.05 e *Log2-fold change* ≥ 3.0) em duas linhagens de batata (não transformada 'WT' e transgênica '+RB', tubérculos) com um número de cópias no mínimo dez vezes maior após a infecção por *Phytophthora infestans*, quando comparado à situação controle (*mock*). Os transcritos que apresentaram tal modulação podem estar associados à resposta ao referido patógeno. Assim, foram encontrados 160 transcritos hiperinduzidos, na linhagem '+RB', e 50 transcritos hiperinduzidos, na linhagem 'WT', contidos nas bibliotecas RNA-Seq analisadas. Adotou-se a estratégia de "*Gene Set/Pathway Enrichment Analysis*" (GSEA; FDR < 0,01) para identificar os termos GO enriquecidos na situação estudada. Dos dez termos GO mais enriquecidos, em 48 horas após a infecção, o processo biológico "reconhecimento de pólen" (GO: 0048544) foi o único termo GO enriquecido em comum nos transcriptomas analisados ('WT' e '+RB'). A linhagem selvagem 'WT' apresentou os termos relacionados à função molecular "ligação à calmodulina" (GO: 0005516) e "atividade oxidoreductase" (GO: 0016717) e processo biológico "fotossíntese" (GO: 0015979) dentre os dez GO mais enriquecidos. Já a linhagem transgênica '+RB' apresentou os processos biológicos "processo biossintético" (GO: 0009058), "resposta de defesa à bactéria" (GO: 0042742) e "resposta a estímulo biótico" (GO: 0009607) e função molecular "atividade transferase" (GO: 0016747), dentre os dez termos GO mais enriquecidos, além do componente celular "parede celular" (GO: 0005618). Tal fato seria esperado, uma vez que a referida linhagem possui o gene que confere resistência ao patógeno. Assim, é possível verificar uma especificidade da linhagem transgênica na resposta à *P. infestans* em nível de transcritos hiperinduzidos, tanto em relação ao seu quantitativo, tanto em relação aos processos enriquecidos representados por esses.



RNA e transcriptômica

DINÂMICA DA EXPRESSÃO DE GENES ENVOLVIDOS NA BIOSÍNTESE DO ASCORBATO DURANTE O DESENVOLVIMENTO DE FRUTOS DE ACEROLA

Autores: Clesivan Pereira dos Santos¹; Mathias Coelho Batista¹; Dayane Manfrine Bastos Rocha¹; André Luiz Maia Roque¹; Rachel Alves Maia¹;

E-mail para correspondência: clesivanpereira@gmail.com

Instituições: ¹Departamento de Bioquímica e Biologia Molecular ? DBBM, Universidade Federal do Ceará, Fortaleza, CE, Brasil;

Palavras-chave: Acerola; biossíntese de ascorbato; transcriptoma

Apoio: Os autores agradecem ao CNPq, CAPES e FUNCAP pelo suporte financeiro a esta pesquisa.

A acerola é um fruto tropical de destaque no cenário brasileiro, apresentando grande importância econômica, nutricional e social, em especial nos Estados nordestinos. Entretanto, existem poucos estudos moleculares com acerola, principalmente devido à ausência de dados genômicos disponíveis. A acerola destaca-se pelo elevado teor de ácido ascórbico (AAs), contudo, seu balanço em frutos é um processo complexo que pode envolver a participação de pelo menos quatro vias de biossíntese (L-galactose, L-gulose, Mio-inositol oxigenase e D-galacturonato). Entretanto, a provável contribuição de cada via é desconhecida e as seguintes indagações ainda permanecem: Quais vias estão ativas para a biossíntese de AsA em acerola? Qual é a provável contribuição de cada via? Para responder a essas perguntas, objetivou-se investigar o perfil transcricional de frutos verde, intermediário e maduro de acerola. As concentrações de ascorbato foram determinadas por Cromatografia à Líquido de Alta Eficiência. Para a montagem do transcriptoma de frutos de acerola usou-se o programa Trinity. A análise de expressão diferencial foi conduzida utilizando o pacote DESeq2. Foram filtrados e considerados como TDEs, os transcritos com taxa de descoberta falsa (FDR < 0,05) e \log_2 fold-change (≥ 1). A anotação dos TDEs foi feita pelos bancos *Non-redundant* (NCBI) e KEGG. O teor de AsA reduzido variou de 2.975 mg.100 g⁻¹ MF nos frutos verdes para 1.324,7 mg.100 g⁻¹ MF nos frutos maduros. Um total de 48.203 *contigs* (≥ 200 pb) foram obtidos. A análise de expressão revelou um total de 4.326 transcritos diferencialmente expressos (TDEs) durante o amadurecimento dos frutos. Do estágio verde para o maduro, 1.591 transcritos tiveram a expressão aumentada, à medida que 2.318 diminuíram o nível de expressão. Verificou-se a presença de transcritos codificando para todas as vias de biossíntese de ascorbato em acerola. Notavelmente, a predominância da via da L-galactose e a contribuição da via do mio-inositol oxigenase correlacionaram-se com o acúmulo de AsA em frutos verdes, destacando os genes *PM11*, *PMM*, *GMP1* e 3, *GME1* e 2 e *GGP1* e 2 (via da L-galactose) e *MIOX4* (via do mio-inositol oxigenase) como cruciais na regulação desse processo. Em contrapartida, os resultados sugerem que as vias da L-gulose e do D-galacturonato não contribuem significativamente. Estudos futuros serão necessários para verificar a aplicabilidade desses genes no aumento do conteúdo de AsA em vegetais e outras espécies frutíferas.



RNA e transcriptômica

CHUMBO GROSSO - TRANSCRIPTOGRAMER ANALISA O EFEITO PROGRESSIVO DA INTOXICAÇÃO POR CHUMBO NOS SISTEMAS GENÉTICOS DE CÉLULAS PROGENITORAS NEURAIAS.

Autores: Clovis ferreira dos Reis¹; Iara Dantas de Souza¹; Diego Arthur de Azevedo Morais¹; Raffael Azevedo De Carvalho Oliveira¹; Danilo Oliveira Imparato¹;

E-mail para correspondência: cfreis@ufm.edu.br

Instituições: ¹Universidade Federal do Rio Grande do Norte;

Palavras-chave: exposição ao chumbo; análise de transcriptograma; RNA-seq

Apoio: Núcleo de Processamento de Alto Desempenho da UFRN (NPAD/UFRN)

As implicações do envenenamento por chumbo são importantes na saúde humana. Atingindo todos os sistemas orgânicos, afeta principalmente o sistema nervoso, com implicações graves e irreversíveis do neurodesenvolvimento, consolidação de memória e processos de aprendizagem em crianças. Sua interação com componentes celulares dá-se de muitas formas, afetando proteínas de ligação a íons, proteínas de sinalização de transdução, canais iônicos transmembrana e fatores de transcrição. No presente trabalho, aplicamos o pipeline do pacote *transcriptograder R/Bioconductor* com a finalidade de avaliar o perfil transcricional de células progenitoras neurais humanas (NPCs) tratadas com acetato de chumbo 30µM. O algoritmo do *transcriptograder* é dotado de um método não supervisionado, baseado em biologia de sistemas, projetado para identificar, em experimentos do tipo caso-controle, grupos de genes funcionalmente associados e diferencialmente expressos. Tal pipeline foi capaz de identificar 8 clusters diferencialmente expressos nos primeiros dias do tratamento com chumbo, que tiveram o seu número de genes envolvidos ampliados com o passar do tempo de tratamento. A análise de enriquecimento de tais clusters revelou que no início do tratamento ocorreram modificações nas vias de glicosilação, biossíntese lipídica e de sinalização da proteína G, além de alterações detectadas nas vias de organização do citoesqueleto, do metabolismo dos canais iônicos e do processo de divisão celular. No tempo final do tratamento, ouve a incorporação dos nós próximos com consequente expansão daqueles clusters, bem como ampliação do número de termos do Gene Ontology (GO) envolvidos. O pipeline do *transcriptograder* foi capaz de identificar diversos sistemas com transcrição alterada, envolvendo termos relacionados ao zinco e ao cálcio, além de vias bem descritas como sendo afetadas quando da exposição a esse metal. Desta forma, conclui-se que o chumbo induz modificações transcricionais importantes nas NPCs as quais podem ser correlacionadas à danos e/ou adaptações de diversos sistemas, todos decorrentes da intoxicação por este metal pesado.

RNA e transcriptômica

BIOINFORMATICS ANALYSIS REVEALS CIRCULATING MICRORNAS REGULATING BIOLOGICAL PATHWAYS IN THYROID CANCER

Autores: Diego Marques da Costa Santos^{1,2}; Ana Beatriz Vezerra de Oliveira²; Lara Clarisse de Lima²; Amanda Ferreira Vidal¹; Raul Hernandes Bortolin²; Ândrea Ribeiro-dos-Santos^{1,3}; Vivian Nogueira Silbiger²;

E-mail para correspondência: diegomarquescs@outlook.com

Instituições: ¹Programa de Pós-Graduação em Genética e Biologia Molecular, Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Avenida Augusto Corrêa, 66075-110, Belém, Pará, Brasil.; ²Laboratório de Bioanálise e Biotecnologia Molecular, Departamento de Análises Clínicas e Toxicológicas, Faculdade de Farmácia, Universidade Federal do Rio Grande do Norte, Natal, RN, CEP: 59014-520, Brasil.; ³Programa de Pós-Graduação em Oncologia e Ciências Médicas, Núcleo de Pesquisa em Oncologia, Instituto de Ciências da Saúde, Universidade Federal do Pará, Rua dos Mundurucus, 66073-000, Belém, Pará, Brasil.;

Palavras-chave: Thyroid cancer; microRNA; non-invasive biomarker

Apoio: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), No. 483031/2013-5; Rede de Pesquisa em Genômica Populacional Humana, No. Biocomputacional/CAPES-051/2013; Fundação de Amparo a Pesquisa do Estado do Pará, No. 155/2014

Thyroid cancer (TC) is one of the most common endocrine malignancies and its global incidence seems to be rising. Several studies have searching for lower invasive biomarkers to improve the diagnostic, follow-up and treatment, and the circulating microRNAs (miRNAs) has been an alternative. Thus, we collected the current published data concerning to circulating miRNA expression with TC diagnosis, TNM-staging risk, postoperative follow-up, and radioiodine therapy. We also discuss some issues that need to be further explored and standardized. The miRNA profiles were used to predict pathways and biological process using bioinformatics tools. In the present review, it was observed 47 circulating miRNAs differentially expressed in TC, which might play in the main pathways associated with carcinogenesis, acting synergistically. Regarding to TNM staging risk, five miRNAs were described as associated with higher metastasis risk, playing in pathways already correlated with tumor aggressiveness. Regarding to postoperative follow-up, the present study observed that 16 miRNAs had reduced expression after surgery, which have a vital role in the main pathways that might favor cell proliferation and invasion, showing to be a potential tool to predict recurrence. In relation to radioiodine therapy profile, the present study observed that 13 miRNAs had altered serum level expression in patients whose lung metastases are not avoid by radioiodine therapy, which had a worse prognosis. These data show that the circulating miRNAs might be a potential alternative to be applied on clinical. However, more studies need to be conducted considering that the preanalytical and analytical variables might influence in the results, and the choice of the control group might result in conflict results.



RNA e transcriptômica

PERFIL TRANSCRIPCIONAL DE PACIENTES INFECTADOS PELO VÍRUS CHIKUNGUNYA COM DIFERENTES FENÓTIPOS DA DOENÇA

Autores: Diego Gomes Teixeira^{1,2}; João Paulo Matos Santos Lima^{1,2,4}; Selma Maria Bezerra Jeronimo^{1,2,3};

E-mail para correspondência: diego.go.tex@gmail.com

Instituições: ¹UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE; ²Departamento de Bioquímica; ³Instituto de Medicina Tropical; ⁴BioME;

Palavras-chave: Surto; Transcriptoma; Auto-imune

Apoio: CNPQ, CAPES

O vírus Chikungunya (CHIKV) é um alfavírus transmitido através da picada dos mosquitos do gênero *Aedes*. Como assinatura da infecção, os pacientes apresentam febre, manchas vermelhas na pele e fortes dores nas articulações. A etapa aguda se dá até o 21º dia após o aparecimento dos sintomas, enquanto as etapas seguintes, mediante a persistência dos sintomas, são caracterizadas como pós-aguda e crônica, essa última tem início após o 3º mês. Há também os casos em que a infecção ocorre de forma assintomática. Um surto de CHIKV ocorreu no Brasil em 2016, onde foram reportados mais de 305,585 casos, abrangendo cerca de 40% dos municípios brasileiros. Durante esse período, na região de Natal-RN, foram coletadas amostras de sangue total de 12 indivíduos. Desses, 4 estavam na fase aguda da doença (grupo Agudo), 6 já se encontravam recuperados ou com leves sintomas (grupo Não-agudo), enquanto 2 indivíduos, mesmo não apresentando sintomas, foram positivos para PCR (grupo Assintomático). Além das amostras dos infectados, foi coletado um grupo controle composto por 6 doadores brasileiros que estavam morando fora do Brasil durante o surto. As amostras tiveram seus RNAs isolados e sequenciados, gerando um total de 18 bibliotecas. Todos os fragmentos sequenciados foram mapeados ao genoma de *Homo sapiens* e ao de CHIKV. A média de mapeamento foi de 96% para humano e apenas 3 bibliotecas da fase Aguda apresentaram sequências para CHIKV. A identificação dos genes diferencialmente expressos (DGEs) entre cada um dos grupos e o controle mostrou que o Assintomático, Agudo e Não-agudo apresentaram 8384, 7538 e 6272 GDEs, respectivamente. Além disso, por meio de uma abordagem de enriquecimento com as vias do KEGG, observamos alta associação dos GDEs do grupo Agudo com a via de lúpus eritematoso sistêmico. Uma análise de enriquecimento por meio dos termos de Gene Ontology revelou que as ontologias enriquecidas para os grupos Agudo e Assintomático estão essencialmente voltadas para o controle da infecção. Nossas observações demonstram que a mudança no perfil transcricional de pacientes infectados por CHIKV, possui uma leve característica de doença auto-imune e que suas semelhanças não estão presentes somente em seus sintomas.



RNA e transcriptômica

ANOTAÇÃO DE UNITAGS SUPERSAGE DE VIGNA UNGUICULATA SOB ESTRESSE DE SECA A PARTIR DE SIMILARIDADES COM VIGNA RADIATA

Autores: Elvson Wallacy da Silva¹; Éderson Akio Kido¹; Valquíria da Silva¹; Vinicius Torres Guerra;

E-mail para correspondência: wallacyelvson@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: data-mining; leguminosa; bioinformática

Apoio: CNPq; Embrapa-Soja

O feijão-caupi (*V. unguiculata* (L.) Walp) é uma cultura de importância econômica na região Nordeste, a qual é afetada por períodos de seca. Diminuir os efeitos desse estresse é interessante para a cultura. O gênero *Vigna* encontra-se representado no NCBI, principal depósito de biomoléculas, sendo que dos 37.183 transcritos de *V. radiata*, 74% apresentam anotação informativa (gene e/ou função), tornando-as úteis para *V. unguiculata*, que apesar de dispor de maior número, estes carecem de anotações adequadas. Em contraste, *V. unguiculata* dispõe de 149.305 *unitags* SuperSAGE (26pb), oriundas de dois acessos sob desidratação radicular (até 150 minutos). Este trabalho objetivou ancorar *unitags* de *V. unguiculata* em transcritos de *V. radiata*, anotá-las funcionalmente e mapeá-las no genoma disponível. Para tanto, alinhamentos válidos (BLASTn, e-value < e⁻⁴) e aceitos envolveram 23.851 *unitags* de *V. unguiculata* e 18.437 transcritos de *V. radiata*, sendo que 83 % dessas *unitags* (19.694) puderam ser anotadas funcionalmente. Das *unitags* anotadas, 141 foram relacionadas com proteínas aquaporinas, responsáveis em parte pelo controle hídrico e osmótico das células, e que podem ser afetadas em situações de estresse como o aplicado. Assim, as anotações das *unitags* derivadas dos transcritos de *V. radiata*, bem como os locos identificados, em associação com as expressões diferenciais de cada acesso, deverão ser úteis na caracterização da resposta tolerante ao estresse em questão.



RNA e transcriptômica

ANÁLISE IN SILICO DE MICRORNAS ALVOS DE RNAs CIRCULARES DIFERENCIALMENTE EXPRESSOS EM CÂNCER GÁSTRICO

Autores: Gloria Tatiana Vinasco Sandoval¹; Amanda Ferreira Vidal¹; Katia de Paiva Lopes¹; André Mauricio Ribeiro dos Santos¹; Samia Demachki²; Paulo Pimentel de Assumpção²; Sidney Emanuel Batista dos Santos^{1,2}; Ândrea Kelly Ribeiro-dos-Santos^{1,2};

E-mail para correspondência: gtvinascos@gmail.com

Instituições: ¹Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Belém, Pará, Brasil. ; ²Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Belém, Pará, Brasil. ;

Palavras-chave: Circular RNAs; microRNAs; Gastric cancer

Apoio: This project was supported by Rede de Pesquisa em Genômica Populacional Humana (RPGPH) - 3381/2013 CAPES- BioComputacional, CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Circular RNAs (circRNAs) have recently emerged as new class of long non-coding RNA and some of these circRNAs were found to be microRNA-sponges. Several studies have shown that miRNA expression is associated with tumorigenesis, tumor progression and prognosis of gastric cancer. Previously, we published the global expression of circular RNAs in gastric cancer, comparing samples without cancer with both tumor-adjacent and gastric cancer samples using RNA-Seq on ribosomal-depleted total RNA. Head-to-tail back-spliced junctions for circRNAs were detected by using two combined prediction algorithm and the differential expression analysis was performed by comparing of the three types of samples. We found five differentially expressed circRNAs (*hsa_circ_0001136*, *hsa_circ_0000284*, *hsa_circ_0000211*, *hsa_circ_0004771*, *hsa_circ_0000524*). The aim of this study was to indicate the target miRNAs of these circRNAs by using an *in silico* analysis. The prediction of target miRNAs was performed by using the miRanda algorithm [v3,3a], applying strict parameters such as alignment score ≥ 170 , free energy ≤ -30 kcal/mol. We selected only the miRNAs that presented at least two complementary *seed* sequences in the circRNA sequence. Then, we identified the target genes of the selected miRNAs by using Target Compare and miRTargetLink Human tools. To describe the biological processes in which the genes are involved, functional analysis was performed by DAVID Bioinformatics Resources [v 6.8]. We found seven microRNAs (*hsa-miR-4646-5p*, *hsa-miR-5095*, *hsa-miR-5096*, *hsa-miR-619-5p*, *hsa-miR-1273a*, *hsa-miR-4459*, *hsa-miR-8485*) potentially regulated by these circRNAs. The functional analysis of the target genes of these microRNAs was enriched in gene expression, transcription, regulation of Hippo signaling, protein import, negative regulation of TOR signaling and cell proliferation. Our results indicate that the epigenetic mechanisms of gene regulation made by the different types of non-coding RNAs, such as circular RNAs, microRNAs and piwi-interacting RNAs, are synergistic and simultaneous. The elucidation of these epigenetic mechanisms may promote the chance of using these non-coding RNAs as biomarkers in gastric cancer.



RNA e transcriptômica

PROSPECÇÃO IN SILICO DE PROTEÍNAS RELACIONADAS À PATOGÊNESE (PR-1 E PR-6) NO TRANSCRIPTOMA DO FEIJÃO-CAUPI

Autores: Gabriel Lucas Araújo Lima¹; João Pacífico Bezerra Neto¹; Amanda Regueira Fernandes¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: gabriel53bio@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: PR-1; PR-6; Feijão-Caupi

Apoio: CNPq, CAPES

A produção agrícola enfrenta condições adversas, tais como seca e ataque de patógenos. Estudos que visam o melhoramento genético vegetal e a melhor compreensão da tolerância aos estresses bióticos e abióticos são necessários. Nossa pesquisa buscou entender, através da prospecção *in silico*, o papel de proteínas PR-1 e PR-6 (pathogenesis related) em Feijão-Caupi (*Vigna unguiculata*). Para tal, foi montado um banco de sequências-sonda através do NCBI, em seguida com a realização de um tBLASTn contra o NATBios. Anotaram-se os domínios conservados (DC) através do CD-search, Uniprot e Pfam, a localização subcelular através do ProtComp, a estrutura secundária pelo Pssfinder, e o peso e ponto isoelétrico (PI) através do JVirgel. Adicionalmente, foi analisada a expressão diferencial dessas proteínas frente estresse hídrico em cultivares tolerante e sensível, por meio da construção de heatmaps via Clustal 3.0 e JavaTreeview, juntamente com a construção de árvores fenéticas por meio do Mega 7, com o método de Neighbor-Joining e bootstrap de 1000 repetições. Foram obtidos 163 candidatos de PR-1 e 17 de PR-6, com 28 sequências apresentando o domínio SCP em PR-1, enquanto que para PR-6, não foram identificados domínios. O domínio SCP é atribuído à família PR1 e possui atividade antimicrobiana. O PI, estrutura secundária e peso molecular variaram, existindo similaridade apenas com as proteínas possuidoras do mesmo domínio conservado. A árvore fenética de PR-1 não apresentou agrupamentos coerentes entre seus membros, o oposto ocorreu para PR-6. O heatmap de PR-1 mostrou sequências diferencialmente expressas (97 na variedade susceptível e 211 na tolerante), enquanto o de PR-6 teve todas as sequências diferencialmente expressas na comparação controle *vs.* estressado. Nas PR-1 as proteínas com domínio SCP foram mais expressas na raiz (variedade tolerante e susceptível), enquanto as com domínio GILT (Gamma inducible lysosomal thiol reductase), estavam mais expressas na cultivar tolerante. O fato dessas PR serem mais expressas na raiz (órgão com maior estresse especialmente na condição de seca), pode indicar uma prevenção a infecções. As proteínas com GILT tiveram alta expressão na variedade tolerante, reiterando sua importância para a tolerância abiótica, existindo a probabilidade de PR-1 possuir atividade frente estresses abióticos. São necessários estudos adicionais da família PR-6, para compreensão de sua estrutura.



RNA e transcriptômica

ANÁLISE IN SILICO DA VIA SOS EM *CENOSTIGMA PYRAMIDALE* (FABACEAE) SUBMETIDA À CONDIÇÃO SALINA

Autores: Gabriella Frosi Albuquerque Figueiroa Faria¹; José Ribamar Costa Ferreira Neto²¹; João Pacífico Bezerra Neto¹; Valesca Pandolfi¹; Mauro Guida Santos¹;

E-mail para correspondência: gabriella.frosi@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Estresse abiótico; Bioinformática; RNA Seq

Apoio: FACEPE

Cenostigma pyramidale (Tul.) Gagnon & G.P. Lewis é uma espécie nativa do semiárido nordestino brasileiro, com potencial para recuperação de áreas degradadas devido à sua capacidade de suportar os estresses abióticos da região, como a salinidade. Essa condição causa perdas da produtividade vegetal. O objetivo do estudo foi analisar a expressão dos três principais genes (SOS1, SOS2 e SOS3) da via *Salt Overly Sensitive* (SOS) no transcriptoma radicular de *C. pyramidale* submetida à salinidade (NaCl 100 mM) em casa de vegetação. O transcriptoma foi obtido nos tempos 30 min, 2h e 11 dias após imposição da salinidade, cada tempo com seu respectivo controle. Para a identificação dos alvos, sondas proteicas foram obtidas no UNIPROT, sendo realizado um tBLASTn no banco de dados local para recuperação dos transcritos alvos e análise da expressão diferencial ($\log_2 FC > 1$ e < -1). Considerando todos os tempos analisados, foram encontrados 19, 162 e 79 transcritos para SOS1, SOS2 e SOS3, respectivamente, em relação ao experimento como um todo. SOS1 apresentou 58% dos transcritos diferencialmente expressos (DE), sendo a maioria induzida após 30 min de rega salina. Para SOS2, 60% dos transcritos apresentaram DE, também com maior indução associada ao tempo de 30min. Para os genes SOS3, 73% dos transcritos foram DE, sendo a maioria induzida após 2 h de NaCl. Os resultados obtidos corroboram com os dados da literatura, sugerindo que os genes analisados, apesar de pertencerem a um mesmo processo metabólico, são heterogêneos quanto a sua expressão temporal e espacial. É válido ressaltar que SOS1, SOS2 e SOS3 também apresentaram transcritos reprimidos nos diferentes tempos de tratamento. A atividade da SOS1 é regulada pela SOS2 (serina/treonina proteína quinase) e SOS3 (proteína de ligação ao cálcio). SOS3 recruta SOS2 para a membrana plasmática, e o complexo formado pelas duas ativam a SOS1 por fosforilação, promovendo o efluxo de Na⁺ das células. Entretanto, a repressão representativa de transcritos em todos os tempos indica que outras vias devem ter sido utilizadas pela planta, no sentido de tolerar o estresse imposto. Além disso, como SOS1 participa do carregamento de Na⁺ no xilema, a repressão pode ser uma estratégia para evitar o transporte desse íon das raízes para a parte aérea. Tal fato indica que isoformas transcricionais específicas são recrutadas na regulação da via SOS em *C. pyramidale* desde os tempos iniciais de estresse salino.

RNA e transcriptômica

IDENTIFICAÇÃO DE CORREGULADORES EXPRESSOS EM FEIJÃO-CAUPI SOB INFECÇÃO DO VÍRUS DO MOSAICO DO CAUPI ASSOCIADO AO AFÍDEO (CABMV)

Autores: Gizele de Andrade Luz¹; Maria Fernanda da Costa Gomes¹; Vinicius Torres Guerra¹; Manassés Daniel da Silva¹; José Ribamar Costa Ferreira Neto¹; Ana Maria Benko Iseppon¹; Antônio Félix da Costa²; Ederson Akio Kido¹;

E-mail para correspondência: gizeledeandrade@gmail.com

Instituições: ¹Universidade Federal de Pernambuco; ²Instituto Agronômico de Pernambuco;

Palavras-chave: Regulação gênica; estresse biótico; transcriptoma

Apoio: FACEPE

O feijão-caupi (*Vigna unguiculata*) é uma das principais espécies cultivadas em regiões áridas e semiáridas do globo, sendo as infecções virais um dos principais fatores limitantes a sua produtividade. A aclimatação, adaptação e resistência a estresses bióticos são dependentes da ativação de respostas moleculares, primariamente iniciadas por meio da regulação de genes específicos. Uma classe de proteínas essenciais na regulação gênica são os correguladores. Essas regulam a expressão gênica de forma indireta, interagindo com o aparato de transcrição basal ou alterando a acessibilidade dos fatores de transcrição ao DNA. Assim, este trabalho objetivou identificar transcritos de correguladores em feijão-caupi submetido à infecção pelo *Cowpea Aphid-Borne Mosaic Virus* (CABMV), nos tempos de 60 min e 16 h, bem como analisar a conservação de seus *loci* codificadores em feijão-comum (*Phaseolus vulgaris*) e soja (*Glycine max*). O sequenciamento foi realizado com o HiSeq 2500 (Illumina) e a montagem do transcriptoma *de novo*. A identificação dos transcritos (RNA-Seq) de feijão-caupi foi feita via BLASTn ($e\text{-value} < e^{-20}$) contra 388 sequências, correspondentes a 23 famílias de correguladores de *V. angularis*, alocadas na base de dados ITAK. A conservação gênica foi analisada via BLASTx ($e\text{-value} < e^{-10}$) dos transcritos de feijão-caupi, anotados como correguladores, contra os proteomas putativos de feijão-comum e soja, depositados no NCBI. Assim, foram anotados 1594 transcritos de feijão-caupi associados a correguladores. Esses foram ancorados em 490 genes da referida espécie e são representantes das 23 famílias mencionadas. As famílias mais representadas foram: AUX/IAA (124), GNAT (153), PHD (211), SET (213) e SNF2 (210). No total, 96 transcritos foram diferencialmente expressos ($P\text{-value} < 0,05$ e $\text{Log}_2\text{FC} \geq 1$), sendo 72 (29 induzidos/43 reprimidos) e 24 (12 induzidos/12 reprimidos) para os tempos de 60 min e 16 h, respectivamente. A família SNF2 apresentou o maior número de transcritos induzidos, com 4. Dos 490 genes minerados, 444 (90,6%) também estiveram presentes nos genomas de feijão-comum e soja. Essa significativa conservação gênica sugere a importância dessas proteínas na regulação da expressão gênica em legumes. Assim, este estudo lança luz na genômica comparativa de correguladores entre leguminosas, bem como na sua orquestração transcricional em feijão-caupi submetido à infecção pelo CABMV, adicionando informações à fisiologia molecular dessa cultura sob estresses.

RNA e transcriptômica

IDENTIFICAÇÃO E ANÁLISE FENÉTICA DE AQUAPORINAS NO TRANSCRIPTOMA DE [STYLOSANTHES SCABRA VOGEL]

Autores: Grazielly Beserra Calixto Venancio³; João Pacifico Bezerra Neto¹; Luana Taís Bezerra Silva²; Flávia Tadeu de Araújo¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: graziellyvenancio.gv@gmail.com

Instituições: ¹Universidade Federal de Pernambuco, Recife, PE, Brasil.; ²Universidade de Pernambuco, Recife, PE, Brasil.; ³Centro Universitário UniFBV | Wyden, Recife, PE, Brasil.;

Palavras-chave: Bioinformática; Estresse abiótico; Leguminosas

Apoio: CAPES, CNPq, FACEPE.

A estilosantes (*Stylosanthes scabra*) é uma leguminosa forrageira, nativa das regiões tropicais e subtropicais, com grande potencial para a produção de massa verde, além de apresentar tolerância à seca e alto valor nutricional, possuindo ampla distribuição no nordeste do Brasil. Diante disto, a identificação de aquaporinas (AQPs) surge como uma importante abordagem, pois estas são os principais transportadores de membrana, associados ao transporte de água e outras substâncias, essenciais para o controle do balanço hídrico e, por consequência, do crescimento e desenvolvimento vegetal. Assim, este estudo buscou identificar e classificar as AQPs expressas em *S. scabra*. Para isso, sequências de AQPs já caracterizadas foram submetidas a um tBLASTn (*cut-off* de $1e^{-05}$) contra o genoma expresso da espécie, visando identificar possíveis ortólogos. Os transcritos identificados foram então anotados contra o Uniprot e posteriormente traduzidos para a identificação de domínios. As sequências que possuíam domínio completo foram submetidas ao MEGA7 para a construção de uma árvore pelo método de Neighbor-Joining (*bootstrap* = 1000). Após o BLAST, 237 transcritos foram anotados como possíveis AQPs em estilosantes, onde após anotação contra o Uniprot (banco revisado), retornou 124 transcritos associados com aquaporinas. Na identificação de domínios conservados, verificamos que 48 sequências possuíam o domínio ‘Major Intrinsic Protein (MIP)’ íntegro, sendo submetidas à análise fenética, juntamente com 10 sequências classificadas em *Arabidopsis thaliana*. Classificando qualitativamente esse conjunto, obteve-se: 16 proteínas intrínsecas de membranas plasmáticas (PIPs); 17 proteínas intrínsecas de tonoplasto (TIPs); sete proteínas intrínsecas de membrana de nódulos (NIPs); e seis pequenas proteínas básicas intrínsecas (SIPs), com outras duas sequências associadas ao grupo, mas não inseridas nos ramos SIP1 e SIP2. A partir do dendrograma, os transcritos foram classificados em quatro subfamílias, como observado para plantas, reforçando a elevada conservação de sua estrutura/função/afinidade pelo soluto transportado. Apesar de identificados 48 transcritos, análises adicionais serão conduzidas para maximização do número de aquaporinas em *S. scabra*, permitindo assim, a aplicação destes genes em iniciativas de biotecnologia, tanto para a espécie, como para outras leguminosas de interesse agrônomo.

RNA e transcriptômica

IDENTIFICAÇÃO MOLECULAR DE *BACILLUS* SPP. COM POTENCIAL PARA O COMBATE DE PATÓGENOS EM SEMENTES DE ARROZ.

Autores: Ivaneide de Oliveira Nascimento^{1,2}; Antonia Alice Costa Rodrigues; Leonardo de Jesus Machado Gois de Oliveira²;

E-mail para correspondência: ivaneide_agro@yahoo.com.br

Instituições: ¹Universidade Estadual da Região Tocantina do Maranhão; ²Programa de pós-graduação em Agroecologia - Universidade Estadual do Maranhão;

Palavras-chave: Controle Biológico; fitopatógenos de semente; arroz

Apoio: Fundação de Apoio e Amparo a Pesquisa no Maranhão - FAPEMA Universidade estadual da Região Tocantina do Maranhão Universidade Estadual do Maranhão

A preocupação com as questões ambientais, aponta as bactérias do gênero *Bacillus* como uma ferramenta promissora para a sustentabilidade dos agroecossistemas. Com o presente trabalho objetivou-se identificar e avaliar espécies de *Bacillus* spp. do filoplano de arroz, com potencial para o combate de fitopatógenos em sementes de arroz. Foram realizadas coletas de folhas de plantas de arroz sadias jovens, em campos comerciais, em dez municípios maranhenses e posterior isolamento e identificação molecular dos *Bacillus*. O isolamento de *Bacillus* foi realizado de acordo com Mariano e Silveira (2005) e Bettiol (1995), com modificações. Para a identificação molecular das bactérias, o gene codificador do RNA ribossomal 16S foi amplificado pela reação em cadeia pela polimerase (PCR) e, posteriormente, sequenciado. A verificação da amplificação foi realizada por meio de eletroforese em gel de agarose 0,8%, acrescido de brometo de etídio (100 ng/mL), e registrada em fotodocumentador acoplado a transiluminador UV. O sequenciamento foi realizado em sequenciador capilar 3500XL Genetic Analyzer (Applied Biosystems). As sequências obtidas foram comparadas com sequências de espécies tipo depositadas no Ribosomal Database Project, release 10 (<http://rdp.cme.msu.edu>) (COLE et al., 2009). A árvore filogenética foi construída utilizando o programa MEGA 6.3. Para a avaliação da redução dos fitopatógenos, as sementes foram microbiolizadas com a suspensão de bactéria em solução salina (NaCl 0,85 %), na concentração de OD₅₄₀ = 0,5 para cada isolado, as sementes foram plaqueadas em papel de filtro, em experimentos conduzidos em delineamento inteiramente casualizado. Os parâmetros avaliados foram incidência e redução dos patógenos em sementes aos sete dias após o plaqueamento. Foram obtidos e identificados 12 isolados *Bacillus* spp., distribuídos em três espécies: *B. thuringiensis*, *B. methylotrophicus* e *B. amyloliquefaciens*. A microbiolização das sementes com os isolados de *Bacillus* spp. controlou a incidência de *C. oryzae*, com destaque para as espécies *B. methylotrophicus* (B22') e *B. thuringiensis* (B35) com 69,00 % e 61,57 % de controle, respectivamente. Os patógenos *Fusarium* sp., *Rhizoctonia* sp. e *R. stolonifer* não foram controlados pela maioria dos *Bacillus* spp. Os *Bacillus* spp apresentaram bom desempenho no combate a *C. oryzae*.



RNA e transcriptômica

ANALYSIS OF LNCRNAs IN NPCs OF GOLDEN SYRIAN HAMSTER (*MESOCRICETUS AURATUS*) NEWBORNS INFECTED BY ZIKA VIRUS

Autores: Jardel Fabio Lopes Ferreira¹; Samir Manour Moraes Casseb^{1,2}; Carlos Alberto Marques de Carvalho³; Karla Fabiane Lopes de Melo^{1,2}; Suellen de Almeida Machado^{1,2}; Francisco Canindé Ferreira de Luna^{1,4}; Paloma Daguer Ewerton dos Santos^{1,2}; Walter Felix Franco Neto¹; Paula Katharine de Pontes Spada^{1,4}; Livia Carício Martins¹; Ana Cecilia Ribeiro Cruz¹; Pedro Fernando da Costa Vasconcelos¹;

E-mail para correspondência: jardell.fe@gmail.com

Instituições: ¹Instituto Evandro Chagas; ²Escola Superior da Amazônia; ³Universidade Estadual do Pará; ⁴Universidade Federal do Pará;

Palavras-chave: Transcriptoma; Zika; lncRNAs

Apoio: CNPq, IEC, SVS e MS

The paradigm of the predominant role of RNA as a messenger in the cellular process of translating genetic information (DNA) into functional molecules (proteins) has completely changed since the discovery of the mechanism of RNA interference (RNAi) more than a decade ago. Since then, it has become increasingly clear that RNAs, large and small, are key intracellular molecules that execute vital, direct roles in e.g. organismal development, cell survival, cell to cell communication and innate immune responses against invading pathogens. Regulation of these processes takes place at multiple levels involving an expanding range of newly discovered functional RNAs, e.g. small interfering (si)RNA, micro (mi)RNA, PIWI-interacting (pi) RNAs, but also longer non-coding RNA and circular RNAs. Non-coding RNA molecules have many, mainly undiscovered functions, and are involved in a diverse range of intimately regulated pathways, including those related to viral diseases. Zika virus (ZIKV), which is a mosquito-borne RNA virus of the family Flaviviridae (genus Flavivirus), has emerged as a medically important pathogen since the start of a large-scale outbreak in Brazil in 2015. ZIKV infections now threaten human health in the Americas, large parts of Asia and Africa, with severe symptoms including congenital microcephaly and Guillain-Barré syndrome. Male hamsters (*Mesocricetus auratus*) infected with Zika Virus were used, where RNA was extracted from the neural tissue of animals with 5 days post infection and uninfected animals (Controls) with the Maxwell 16 Tissue LEV RNA Purification Kit (Promega, USA). This RNA was used for a transcriptome adopting the NextSeq 550 platform (Illumina, USA), the data from this sequencing were analyzed using the Tuxedo protocol (Trapnell, 2012) and statistical analysis using the R project platform with the Dynamic Transcriptome Analysis (DTA) (Schwalb, 2012). By accounting for the experimental design and estimation of technical and inferential variance we were able to pinpoint Zika infection affected pathways that highlight Zika's neural tropism. The examination of differential genes reveals cases of isoform divergence. Transcriptome analysis of Zika has the potential to identify the molecular signatures of Zika infected neural cells. These signatures may be useful for diagnostics and for the resolution of infection pathways that can be used to harvest specific targets for further study.



RNA e transcriptômica

INTERAÇÃO DE MICRORNAS REGULATÓRIOS NO GENE HLA-G: ANÁLISE IN SILICO

Autores: Jessyca Kalynne Farias Rodrigues¹; Diego Santana Jerônimo da Silva¹; Kleodoaldo Oliveira de Lima³; Dalila Bernardes Leandro²; Everton Neri da Silva²; Ronaldo Celerino da Silva¹; Sergio Crovella^{1,2};

E-mail para correspondência: jessyca_kalynne@hotmail.com

Instituições: ¹Laboratório de Imunopatologia Keiso Asami (LIKA), Universidade Federal de Pernambuco, Recife, Brasil; ²Departamento de Genética, Universidade Federal de Pernambuco, Recife, Brasil; ³Hospital das Clínicas, Universidade Federal de Pernambuco, Recife, Brasil;

Palavras-chave: microRNAs ; HLA-G; regulação pós-transcricional

Apoio: FACEPE

O gene HLA-G codifica uma molécula não-clássica do complexo principal de histocompatibilidade (MHC) com propriedades imunomoduladoras importantes. Polimorfismos na 3' UTR do HLA-G tem sido relacionado com a estabilidade do mRNA e interação com microRNAs (miRNAs). Os miRNAs são pequenos RNAs não codificantes envolvidos na regulação gênica pós-transcricional por meio da degradação da molécula de RNA mensageiro ou da inibição da tradução. Alguns miRNAs são relatados como potenciais reguladores da expressão de HLA-G, entre eles o hsa-miR-365a-3p. O objetivo deste trabalho foi avaliar a possível interação entre o hsa-miR-365a-3p na 3'UTR do mRNA do *HLA-G* através de ferramentas de bioinformática. Foi realizada a predição de alvos do hsa-miR-365a-3p através do banco miRWalk 2.0, assim como anotação funcional dos genes alvos pela ferramenta DAVID v6.8 e avaliação da interação entre o miRNA e a 3'UTR do mRNA do *HLA-G* pelo RNAhybrid. Pares de miRNA-mRNA com FDR <0,20 foram considerados notáveis. Através da predição pelo miWalk e DAVID, 150 genes foram selecionados como possíveis alvos de interferência pelo hsa-miR-365a-3p, atuando em diversas vias biológicas. A melhor via predita foi a "*Sprouty regulation of tyrosine kinase signals*" com FDR: 0.42, p-value 3.85E-04. Em relação à interação miRNA/mRNA, o mde encontrado para o hsa-miR-365a-3p foi -39.59 kcal/mol. Na análise individual de Hits, foi encontrada uma interação microRNA/mRNA com mfe: -24.3 kcal/mol, p-value: 1.00, na posição 988. Portanto, a utilização de algoritmos diferentes na inferência de alvos de miRNAs e a união/comparação de seus resultados são ferramentas úteis, gerando ao final das análises um painel de miRNAs com alto potencial de regulação do alvo em questão. Estes resultados sugerem que a regulação pós-transcricional do HLA-G pode desempenhar um papel importante no desenvolvimento de vários outros contextos patológicos, e esta regulação pode ser dependente de fatores genéticos e do microambiente celular.



RNA e transcriptômica

IDENTIFICAÇÃO DE GENES CO-REGULADORES (COREGS) EM *J. CURCAS* L. SOB ESTRESSE SALINO.

Autores: Jorge Luís Bandeira da Silva Filho¹; Manassés Daniel da Silva¹; Marislane Carvalho Paz de Souza¹; Éderson Akio Kido¹;

E-mail para correspondência: jorge08luis94@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Transcriptômica; Epigenética; Pinhão-manso

Apoio: Facepe e CAPES

Jatropha curcas L. é uma planta da família Euphorbiaceae com potencial para gerar biodiesel. Apesar de a cultura ser de relativa tolerância à seca, ela é sensível à salinidade dos solos, a qual pode ocorrer em regiões nordestinas semiáridas. Melhorar essa tolerância poderá incluir a cultura como opção em plantios nordestinos. Logo, identificar genes relacionados com o caráter e o modo como estes são regulados é fundamental. Dos agentes controladores da expressão gênica, os quais podem influenciar nas respostas das plantas aos estresses, um grupo importante não interage diretamente com o DNA. Estes são conhecidos como co-reguladores (CoREGs). Este estudo visou identificar famílias de CoREGs expressas em *J. curcas* sob estresse salino de 150 mM de NaCl, com potencial expressão diferencial em resposta ao estímulo. Para tanto, transcritos CoREGs de mandioca [*Manihot esculenta* L. (520)] e mamona [*Ricinus communis* L. (415)], da base de dados ITAK, e de *Arabidopsis thaliana* L. (337), da base PTFDB, foram alinhados (BlastN, e-value < e⁻³⁰) com sequências RNAseq (145.422) de 12 bibliotecas (raízes) de *J. curcas* sob o referido estresse salino. Os alinhamentos aceitos envolveram 716 transcritos relativos a 736 CoREGs. Quase todos os potenciais CoREGs de *J. curcas* foram identificados a partir de *M. esculenta* e *R. communis*, com exceção de sete exclusivamente similares aos CoREGs de *A. thaliana*. Além dos CoREGs identificados com base em mais de uma espécie, *M. esculenta* foi mais eficiente ao identificar 132 transcritos exclusivos desta espécie; a contribuição de *R. communis* foi de 51 exclusivos. As famílias CoREGs mais representadas foram: SET (76), PHD (70) e SNF2 (63). Dos 44 representantes disponíveis da família SET de *M. esculenta*, 41 foram identificados em *J. curcas*, e dos 33 disponíveis para *R. communis*, 29. Para as famílias PHD e SNF2, dos disponíveis para *M. esculenta* (41 e 33, respectivamente), foram identificados 36 e 32 transcritos de *J. curcas*. Dos representantes de *R. communis* para as mesmas famílias, daqueles disponíveis (37 e 31, respectivamente) foram identificados 32 e 25 em *J. curcas*. Destes potenciais CoREGs identificados, aqueles diferencialmente expressos (*p-value* < 0,001 e FDR < 0,05) de cada família foram 15 (PHD), quatro (SNF2) e quatro (SET). Estes resultados apontam possível expressão diferencial dos genes CoREGs destas famílias em resposta ao estresse. Os resultados estão sendo validados quanto à expressão relativa em ensaios de qPCR.



RNA e transcriptômica

RESISTÊNCIA A *P. VITICOLA* ATIVADA POR *T. HARZIANUM* EM ACESSO SUSCETÍVEL DE UVA: UM ENFOQUE NOS TRANSCRITOS HIPERINDUZIDOS

Autores: Jéssica Barbara Vieira Viana¹; José Ribamar Costa Ferreira Neto¹; Ana Maria Benko Iseppon¹;

E-mail para correspondência: jessbarbara.jb@gmail.com

Instituições: ¹Universidade Federal do Pernambuco;

Palavras-chave: BDVA; resveratrol; termos enriquecidos

Apoio: CAPES, CNPq e FACEPE.

Em transcriptômica, bancos dados de valor agregado (BDVA) têm ganhado relevância por alocarem dados públicos, processando-os e disponibilizando em versão mais acessível e informativa à comunidade científica. Assim, outros grupos de pesquisa podem minerar situações específicas não investigadas em dados já publicados. Dessa forma, foi realizada mineração de informações contidas nos transcriptomas de uva analisados no manuscrito “*Downy mildew resistance induced by Trichoderma harzianum T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotypes*” (PMID:23173562), disponível na BDVA Expression Atlas (código “E-MTAB-4324”). Teve-se como intuito analisar, em cultivar suscetível de videira (*Vitis vinifera*), somente os transcritos hiperinduzidos ($p\text{-value} \leq 0.05$ e $\text{Log}_2\text{-fold change} \geq 3.0$) associados à resistência ao patógeno *Plasmopora viticola*, sendo essa ativada por tratamento prévio com o microrganismo benéfico *Trichoderma harzianum* T39 (T39). Nesse contexto, foram encontrados 12 transcritos para o acesso tratado com T39 (transcriptoma ‘A’); 26, para o acesso tratado com *P. viticola* (transcriptoma ‘B’); e 22, para o acesso tratado com T39 e, posteriormente com *P. viticola* (transcriptoma ‘C’). Adotou-se a estratégia de “Gene Set/Pathway Enrichment Analysis” (GSEA; $\text{FDR} < 0,01$) com o objetivo de agregar significado biológico aos transcriptomas escrutinados. Constatou-se os termos GO:0008017 e GO:0000413, como os mais enriquecidos no transcriptoma ‘A’, os quais se referem, respectivamente, à interação de microtúbulos e isomerização de proteínas. Para o transcriptoma ‘B’, os termos mais enriquecidos foram: GO:0031425 e GO:0050350, relacionados, respectivamente, ao processamento de RNA e síntese de resveratrol, sugerindo a atividade de estilbeno na resposta ao fungo *P. viticola*. Para o transcriptoma ‘C’, os termos mais enriquecidos, GO:0009535 e GO:0009579, estão associados aos tilacóides do cloroplasto, sugerindo que plantas previamente tratadas com T39 investem fortemente na proteção do aparato fotossintético, evitando que o mesmo seja danificado em decorrência do estresse aplicado. Em suma, foi possível identificar diferenças no enriquecimento de termos associados aos transcritos hiperinduzidos nos três transcriptomas escrutinados. Adicionalmente, a estratégia de GSEA sugere que o tratamento com o microrganismo benéfico T39 pode auxiliar na manutenção da estabilidade do aparato fotossintético em uva sob infecção de *P. viticola*.



RNA e transcriptômica

IDENTIFICAÇÃO DE REGIÃO ALVO PARA DIAGNÓSTICO MOLECULAR ESPECÍFICO DO PAPILOMA VÍRUS HUMANO 16

Autores: Ketly Rodrigues Barbosa dos Anjos¹; Anderson Alves da Silva Bezerra¹; Ana Lisa do Vale Gomes¹;

E-mail para correspondência: ketly.rodrigues@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Genoma viral; bioinformática; diagnóstico

Introdução: Os vírus tendem a se adaptar ao ambiente com alterações em seu genoma, o que pode resultar em uma maior patogenicidade e dano ao hospedeiro. Assim, é necessário identificar as variações no genoma viral e elaborar um primer in silico espécie-específico com as sequências nucleotídicas atuais, o que possibilita uma forma mais eficaz de diagnóstico laboratorial. A região E6 do *Human papillomavirus (HPV)* é essencial para o diagnóstico específico pois é responsável pela interação oncogênica com as células do colo uterino. **Objetivo:** Identificar uma região alvo para diagnóstico molecular a partir da região E6 do Human papillomavirus. **Método:** Trata-se de uma pesquisa descritiva com abordagem quantitativa dos dados. A amostra é composta de sequências selecionadas no banco de dados eletrônico do National Center for Biotechnology Information, do HPV subtipo 16, região E6, utilizando os seguintes filtros de pesquisa: viruses, genomic DNA/RNA. Foram obtidas 3227 sequências do subtipo 16 as quais foram separadas de acordo com o país da amostra. Em países que possuíam mais de uma sequência, foi realizado um alinhamento múltiplo utilizando o programa Clustal X 2.1 e selecionada a sequência com maior similaridade entre as outras sequências como representante do país. A amostra total, após seleção, é composta de 22 sequências que foram alinhadas no programa Clustal Ômega, disponível na plataforma online no site: <https://www.ebi.ac.uk/Tools/msa/clustalo/> e a leitura dos resultados através do programa Jalview para visualização das áreas em consenso. Para o desenho do primer foi utilizado o programa FastPCR e os dados foram anexados em uma planilha no Microsoft Excel 2010. **Resultados:** A sequência consenso entre os países foi selecionada de acordo com a região em consenso para o desenho do primer, denominada região alvo, com 95 nucleotídeos. O FastPCR analisou a sequência e definiu 31 primers, 13 forward e 18 reverse. Os primers foram selecionados de acordo com a porcentagem de relação CG/AT (60%) e a temperatura específica de 58° C, resultando em 3 pares de primers. Os pares foram analisados no programa BLAST, obtendo uma identificação e query score de 100%, identificando outras sequências de HPV 16, região E6. **Conclusão:** A utilização de um primer com maior especificidade levará ao diagnóstico específico do vírus, possibilitando o diagnóstico específico precoce e o tratamento mais eficaz.



RNA e transcriptômica

PERFIL DE EXPRESSÃO DIFERENCIAL DE MICRORNAS NO CÂNCER GÁSTRICO SUGERE A EXISTÊNCIA DE CAMPO DE CANCERIZAÇÃO

Autores: Laís Reis das Mercês¹; Adenilson Leão Pereira¹; Leandro Lopes de Magalhães¹; Fabiano Cordeiro Moreira²; Amanda Ferreira Vidal¹;

E-mail para correspondência: laiscdreis@gmail.com

Instituições: ¹Programa de Pós-Graduação em Genética e Biologia Molecular, Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Avenida Augusto Corrêa, 66075-110, Belém, Pará, Brasil.; ²Programa de Pós-Graduação em Oncologia e Ciências Médicas, Núcleo de Pesquisa em Oncologia, Instituto de Ciências da Saúde, Universidade Federal do Pará, Rua dos Mundurucus, 66073-000, Belém, Pará, Brasil.;

Palavras-chave: microRNA; Field Cancerization; Biomarker

Apoio: Rede de Pesquisa em Genômica Populacional Humana (CAPES/Biologia Computacional: N°. 3381/2013/CAPES), UFPA

The biological role of microRNAs (miRNAs) in field cancerization is unknown. To investigate the involvement of these in gastric field cancerization, we evaluated the expression profile of ten miRNAs. Furthermore, the diagnostic value of these miRNAs for gastric cancer was tested. We used three groups of gastric tissues samples: non-cancer (NC, n=30), gastric cancer (GC, n=20), and matched tumor-adjacent (ADJ, n=11). The expression profiles of *hsa-miR-10a*, *hsa-miR-21*, *hsa-miR29c*, *hsa-miR-135b*, *hsa-miR-148a*, *hsa-miR-150*, *hsa-miR-204*, *hsa-miR-215*, *hsa-miR-483* and *hsa-miR-664a* were investigated using qRT-PCR. The statistical significance among the groups of samples was analyzed by One-way ANOVA and the Bonferroni's correction was applied for post hoc analysis. Furthermore, was performed an unsupervised hierarchical clustering analysis to evaluate the similarity among the groups. The potential biomarker was evaluated using a receiver operating characteristic (ROC) curve and the derived area under the curve (AUC>0.85) analysis. The statistical analyses and graphs were developed in the packaged R. The search for target driver genes of the studied miRNAs was performed in the miRTarBase public database, using experimentally validated interactions with strong evidence. The miRNAs *hsa-miR-10a*, *hsa-miR-21* and *hsa-miR-135b* were up-regulated in ADJ and GC compared to NC; and up-regulated in GC compared to ADJ. *hsa-miR-148a*, *hsa-miR-150*, *hsa-miR-204*, *hsa-miR-215*, *hsa-miR-483* and *hsa-miR-664a* were up-regulated in ADJ and GC compared to NC. These miRNAs were not differentially expressed between GC and ADJ, suggesting that both share similar changes. *hsa-miR-29c* was up-regulated in ADJ compared to NC and CG; we did not observe a significant difference in the expression of this miRNA between NC and GC. The clustering analysis using all miRNAs studied was able to separate non-cancer from tumor-adjacent and gastric cancer samples. *hsa-miR-10a*, *hsa-miR-21*, *hsa-miR-135b*, *hsa-miR-148a*, *hsa-miR-150*, *hsa-miR-215*, *hsa-miR-483* and *hsa-miR-664a* were able to discriminate NC from other tissues with great accuracy (AUC>0.85). Eighteen tumor suppressor genes are targets of the miRNAs studied. The studied miRNAs are closely related to field cancerization, regulating genes involved in gastric carcinogenesis, and can be potentially useful as biomarkers in GC.



RNA e transcriptômica

ANÁLISE DE TRANSCRITOS HIPERMODULADOS EM BRASSICA RAPA SOB INFECÇÃO POR PSEUDOMONAS SYRINGAE E INFILTRAÇÃO POR EPÍTOPO DE FLAGELINA BACTERIANA

Autores: Manassés Daniel da Silva¹; José Ribamar Costa Ferreira Neto¹; Rahisa Helena da Silva¹; Éderson Akio Kido¹;

E-mail para correspondência: manassésdaniel@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: patógeno; transcriptômica; couve-china

Apoio: CAPES e CNPq

Experimentos de sequenciamento a partir de RNAs obtidos de sequenciamento de alto desempenho têm gerado uma quantidade substancial de dados para análises da expressão gênica, onde boa parte desses dados é depositada em bancos de dados públicos. Assim, outros grupos de pesquisa podem comparar informações obtidas em seus ensaios com os dados depositados nesses bancos de dados ou minerar situações específicas não investigadas em dados já publicados. Com este intuito, foi realizada mineração específica de dados oriundos de transcriptomas de couve-china, enfocados no manuscrito “*Two cytochromes P450 catalyze S-heterocyclizations in cabbage phytoalexin biosynthesis.*” (PMID: 26389737), os quais se encontram disponíveis na base de dados *Expression Atlas* (código “E-GEOD-69785”). Teve-se como intuito analisar transcritos hipermodulados ($\text{Log}_2\text{-fold change} \geq 3.0$ e $p\text{-value} \leq 0.05$) associados à resposta de *Brassica rapa* à infecção pelo patógeno *Pseudomonas syringae* pv. *maculicola* (Psm) e à infiltração do epítipo de flagelina bacteriana (flg22). Nesse contexto, foram encontrados 1.617 transcritos (1.026 hiperinduzidos e 591 hiper-reprimidos) para o tratamento com Psm e 804 (516 hiperinduzidos e 288 hiper-reprimidos) para o tratamento com flg22. Os transcritos que apresentaram tais modulações podem estar associados à resposta ao referido patógeno e/ou ao epítipo de flagelina. Em paralelo, adotou-se a estratégia de “*Gene Ontology Enrichment Analysis*” (GOEA; $\text{FDR} < 0,01$) com o objetivo de agregar significado biológico aos transcriptomas analisados. Para o tratamento flg22 os termos mais enriquecidos foram GO:0008652, GO:0009409 e GO:0004222, os quais se referem, respectivamente, ao processo de biossíntese de aminoácidos, resposta ao frio e a atividade de metaloendopeptidase. Para o tratamento Psm o termo GO:0008652 também foi o mais enriquecido, em conjunto com GO:0016874, e GO:0046686, relacionados, respectivamente, com atividade de ligase e resposta ao íon cádmium. A biossíntese de aminoácidos apresenta-se, aparentemente, como um fator relevante na resposta vegetal frente à infecção bacteriana, tornando os componentes dessa via candidatos para ensaios futuros de biologia molecular para o entendimento da resposta de *Brassica rapa* à patógenos.

RNA e transcriptômica

ANÁLISE IN SILICO DO QUINOMA DE *JATROPHA CURCAS* EM RESPOSTA À SALINIDADE

Autores: Maria Fernanda da Costa Gomes¹; José Ribamar Costa Ferreira Neto¹; Gizele de Andrade Luz¹; Vinícius Torres Guerra¹; Manassés Daniel da Silva¹;

E-mail para correspondência: fernanda_mf2@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco (UFPE);

Palavras-chave: Estresse abiótico; pinhão-manso; quinases

Apoio: CAPES

Jatropha curcas, pertencente à família Euphorbiaceae, é conhecida popularmente como pinhão-manso, sendo considerada uma oleaginosa com grande potencial para a produção de biodiesel. Em sua diversidade genética, há indicações de um número reduzido de acessos tolerantes à salinidade. Enzimas do tipo quinase desempenham um importante papel na tolerância ao referido estresse, catalisando reações de fosforilação e ativando, por exemplo, a atividade de proteínas envolvidas na via de sinalização SOS (*Salt Overly-Sensitive*). Assim, objetivou-se identificar e classificar o quinoma (grupos e famílias de proteínas quinases) expresso (via RNA-Seq) em acesso tolerante de pinhão-manso sob estresse salino (NaCl, 150 mM). Adicionalmente, mapeou-se esses transcritos candidatos no genoma de mandioca (*Manihot esculenta*), espécie filogeneticamente próxima e com genoma disponível. Por meio de alinhamentos BLASTx (e-value < e^{-10}), foram identificados 3015 transcritos quinases de pinhão-manso similares a quinases ortólogas de *M. esculenta*, alocadas na base de dados iTAK. Esses transcritos compreenderam as nove grandes classes descritas para tal grupo proteico, bem como todas as 121 famílias contidas na base de dados escrutinada. No que tange à modulação transcricional sob estresse salino, 386 potenciais quinases (cerca de 13%) de pinhão-manso foram diferencialmente expressas ($p\text{-value} \leq 0,00001$; $FDR \leq 0,005$ e FC de $\log_2\text{ratio} \geq 1$), sendo 30 induzidas e 356 reprimidas. Dentre as induzidas, a família CAMK_CDPK apresentou o maior quantitativo (08). Essa família desempenha papel chave na percepção e transdução de sinais das plantas sob estresse. Um total de 2144 transcritos expressos foram associados (via BLASTx; e-value < e^{-10}) ao proteoma putativo de *M. esculenta*, o qual encontra-se devidamente mapeado em genoma da espécie. O resultado demonstra conservação de cerca de 71% dos alvos analisados entre as duas espécies. Os dois cromossomos de *M. esculenta* que apresentaram o maior número de loci codificadores das proteínas acima mencionadas foram LG1 (178) e LG6 (167), sendo este último ancorador do maior número de loci diferencialmente expressos (04 induzidos e 17 reprimidos). Os resultados aqui apresentados servirão de plataforma para estudos associados ao quinoma de pinhão-manso expresso sob estresse salino. Isso contribuirá para adicionar mais informações à fisiologia molecular da referida espécie sob condições estressantes.



RNA e transcriptômica

PROSPECÇÃO IN SILICO DE PEPTÍDEOS DA CLASSE CATELICIDINA EM VIDEIRA SOB ESTRESSE BIÓTICO

Autores: Marianne Firmino de Oliveira¹; Roberta Lane de Oliveira Silva¹; João Pacífico Bezerra Neto¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: marianneoliveira93@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Vitis; AMP; Xanthomonas campestris

Apoio: CNPq e CAPES.

Peptídeos antimicrobianos (*AntiMicrobial Peptides* - AMPs) fazem parte da imunidade inata nos vegetais e compreendem moléculas ancestrais que desempenharam papel fundamental no sucesso evolutivo de organismos multicelulares. AMPs estão presentes em todos os grupos de organismos, desde procariontes até eucariontes. A catelicidina compreende um grupo AMP com atividade antibacteriana, antifúngica, imunestimulatória e imunomoduladora, atuando no combate a microrganismos e como ponte entre as imunidades inata e tardia. Diante disso, buscamos identificar e caracterizar candidatos a catelicidina em acessos de *Vitis* spp. infectados com *Xanthomonas campestris* pv. *viticola*. A análise tomou como modelo referente à catelicidina, depositado no banco de dados CAMP (*Collection of Anti-Microbial Peptides*; www.camp.bicnirrh.res.in/), específico para peptídeos antimicrobianos, as sequências foram submetidas à ferramenta *AMP-identifier* visando identificar possíveis homólogos a esta família no transcriptoma de videira por meio de alinhamentos baseados em modelos probabilísticos do HMMER. As sequências foram anotadas, traduzidas e tiveram seus domínios identificados com o auxílio do CD-search. A predição do ponto isoeletrico (pI) e do peso molecular foi realizada utilizando o JVirGel 2.0, localização subcelular através do Cell-PLoc 2.0 e as pontes dissulfeto por meio do Dianna. Foram identificadas 5 sequências candidatas para catelicidinas com domínios completos em *Vitis*. Tais sequências possuem pI entre 4,18 e 8,24, peso molecular variando entre 18,08 a 99,42 kDa, com localização subcelular predita para o citoplasma ou núcleo, apresentando de três a cinco pontes dissulfeto. No entanto, apesar de apresentar domínios completos, as candidatas não apresentaram todas as características estruturais de uma catelicidina. Assim, é provável que o modelo utilizado como *seed* na ferramenta *AMP-identifier* deve ser incrementado, visando caracterizar de forma mais abrangente este grupo AMPs em videira, com vistas à identificação de candidatos que se enquadrem em termos da estrutura na família para futuras caracterizações de catelicidinas em videira.



RNA e transcriptômica

ANÁLISE ESTRUTURAL E FUNCIONAL DA FAMÍLIA SNAKIN/GASA EM SOJA (GLYCINE MAX) E DESCRIÇÃO DE UM NOVO MEMBRO (GMSN2).

Autores: Marx Oliveira Lima¹; Valesca pandolfi¹; William Farias Porto²; Octávio Luiz Franco²; Jose Ribamar Costa Ferreira Neto¹; Lidiane Lindinalva Barbosa Amorim¹; Francismar Correa Marcelino Guimarães³; Ricardo Vilela Abdelnoor³; Ederson Akio Kido¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: olima.marx@gmail.com

Instituições: ¹Departamento de Genética, Centro de Biociências, Universidade Federal de Pernambuco.; ²Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília; ³Empresa Brasileira de Pesquisa Agropecuária, Embrapa Soja;

Palavras-chave: peptídeos antimicrobianos; defesa vegetal; anotação

Apoio: CAPES, FACEPE, CnPQ, INTERIS

Esnaquinas são peptídeos antimicrobianos (AMPs) da família Snakin /GASA formados por três regiões distintas: um peptídeo sinal N-terminal, um sítio variável e o domínio GASA na região C-terminal composto por doze resíduos de cisteína conservados que contribuem para a estabilidade bioquímica da molécula. Sabe-se que esses peptídeos desempenham papéis diferentes em resposta a uma variedade de estresse biótico (induzidos por bactérias, fungos e nematódeos) e abióticos (salinidade, seca e ROS), bem como na resposta cruzada ‘crosstalk’ promovida por hormônios vegetais, principalmente ácido abscísico e salicílico. Neste trabalho, sequências sonda de 10 famílias vegetais foram utilizadas na busca de sequências não redundantes homólogas de esnaquinas no transcriptoma (ESTs) de soja. Das 33 sequências identificadas, 20 apresentaram domínio GASA completo, com ponto isoelétrico de 4.3 - 9.58 (peptídeo sinal) e de 5.53 - 9.32 (peptídeo maduro) e peso molecular variando de 2.37 - 2.99 KDa (peptídeo sinal) e de 6.88 - 18.04 KDa (peptídeo maduro), sendo todas as sequências endereçadas para o meio extracelular. A partir do isolamento de uma sequência via PCR (e confirmação via sequenciamento), foi possível predizer a estrutura completa do gene, com aproximadamente 374 pb, 60 resíduos de amino ácidos, pI 8.52 e 6.32 KDa. Este peptídeo (nomeado de ‘GmSN2’) apresenta estrutura similar ao modelo experimental da *Snakin-1*, estabilizada por seis pontes dissulfeto. A validação da expressão por RT-qPCR mostrou que *GmSN2* é constitutivamente expressa, mesmo após a inoculação do fungo *Phakopsora pachyrhizi* quando comparada com o controle negativo (plantas não inoculadas). O mapeamento genômico *in silico* revelou 37 homólogos de *GmSN2* distribuídos entre 15 dos 20 cromossomos da soja. Todos os resultados aqui apresentados corroboram o isolamento de um novo peptídeo com características bastante similares a outros previamente isolados em outras espécies vegetais.



RNA e transcriptômica

MONTAGEM E ANOTAÇÃO DOS TRANSCRIPTOMAS DE FOLHAS, FLORES E FRUTOS DE ACEROLEIRA

Autores: Mathias Coelho Batista¹; Clesivan Pereira dos Santos¹; Susan Karen Silva das Chagas¹; Antonio Edson Rocha Oliveira¹; Rachel Alves Maia¹;

E-mail para correspondência: mathiascoelho@alu.ufc.br

Instituições: ¹Departamento de Bioquímica e Biologia Molecular - DBBM, Universidade Federal do Ceará, Fortaleza, CE, Brasil;

Palavras-chave: Aceroleira; amadurecimento; transcriptoma

Apoio: Os autores agradecem ao CNPq, CAPES e FUNCAP pelo suporte financeiro a esta pesquisa.

Atualmente, o sequenciamento de RNA (RNA-seq) tornou-se a estratégia mais adequada para a caracterização estrutural e da expressão em larga escala de transcriptomas. A acerola é um fruto de importância socioeconômica que vem aumentando o seu consumo, mercado e aplicabilidade na indústria de alimentos. Em vista da ausência de dados transcriptômicos disponíveis para a acerola, nosso grupo de pesquisa priorizou a obtenção do perfil transcricional com a construção e o sequenciamento de nove bibliotecas de frutos, folhas e flores de aceroleira, na intenção de estabelecer uma base para estudos moleculares. Para a montagem dos transcriptomas usou-se o programa Trinity. A anotação dos transcritos foi feita com os bancos de proteínas *Non-redundant* (NCBI), *Swiss-Prot* e KEGG através da ferramenta BLAST considerando um E-valor $\leq 10^{-5}$. O sequenciamento gerou mais de 600 milhões de *reads* de 100 pb. Após a remoção de adaptadores e sequências de baixa qualidade, 547.089.334 *reads* (89,22%) foram selecionados com *phred score* $Q > 30$. Posteriormente, os *reads* foram montados em 40.830 *contigs* (> 300 pb) não redundantes. Um total de 25.298 transcritos únicos (aproximadamente 62%) foram anotados pelo BLASTx através de bancos de dados públicos de proteínas. Destes, praticamente todos os transcritos (25.272) foram anotados com o Nr-NCBI. Além disso, 18.608 e 10.789 transcritos foram também anotados pelos bancos *Swiss-Prot* e KEGG. De acordo com a anotação fornecida pelo Nr-NCBI, a maior quantidade de transcritos com os melhores E-valores foram atribuídos a *Hevea brasiliensis* (4.564), *J. curcas* (2.622), *R. communis* (2.620), *M. esculenta* (2.570), *P. trichocarpa* (1.934) e *P. Euphratica* (1.814) da mesma ordem taxonômica (Malpighiales) da acerola. Juntas, essas espécies possibilitaram a anotação de 53,6% dos transcritos únicos anotados. Este estudo reporta o primeiro sequenciamento de transcriptomas de aceroleira e fornece bases estruturais (milhares de sequências de transcritos) e dados de expressão valiosos para estudos moleculares e funcionais posteriores visando a aplicação no melhoramento da qualidade dos frutos.



RNA e transcriptômica

R-PERIDOT: UMA FERRAMENTA PARA SIMPLIFICAR E APERFEIÇOAR A ANÁLISE DE EXPRESSÃO DIFERENCIAL

Autores: Pitágoras de Azevedo Alves Sobrinho¹; Danilo Lopes Martins¹; André Luis Fonseca Faustino¹; Rayson Carvalho Barbosa³; Emily Vanessa Nascimento Araújo⁴; Sandro José de Souza²; Jorge Estefano Santana de Souza¹;

E-mail para correspondência: alves.pitagoras@gmail.com

Instituições: ¹IMD/UFRN; ²ICe/UFRN; ³UFMG; ⁴UFPA;

Palavras-chave: expressão diferencial; Graphical User Interface; Análise de RNA-Seq

Apoio: Apoio do CNPq, através de duas bolsas de iniciação científica.

Diversos software estão disponíveis para a análise de dados de RNA-Seq e Microarray. Cada um usando métodos estatísticos diversos e analisando a expressão diferencial sob diferentes condições para cada gene. No entanto, eles podem gerar resultados diferentes e muitos pesquisadores têm apresentado dificuldades quanto ao seu manuseio, devido a alta curva de aprendizado, pois requerem conceitos de áreas distintas como estatística e linguagens de programação, conceitos tais não presentes nos currículos acadêmicos da maioria dos cursos de ciências da vida e da saúde. Em luz dessas dificuldades, nós desenvolvemos o R-Peridot. Uma ferramenta que faz uso de pacotes disponíveis na linguagem de programação R, para permitir a análise de dados de RNA-Seq e Microarray usando processamento concorrente, com uma interface gráfica intuitiva. Nossa ferramenta abstrai o uso dos pacotes do R transformando-os em módulos que podem ser usados sem que o usuário precise se preocupar com detalhes de implementação e execução de scripts em R. O usuário também pode modificar os módulos existentes ou até adicionar novos módulos caso seja necessário. R-Peridot faz análises dos dados de contagem de reads do usuário através dos pacotes DESeq, DESeq2, sSeq, EBSeq e edgeR. Cada um desses usa seus próprios métodos estatísticos para gerar uma tabela com um subconjunto dos genes dos dados de entrada (os quais seriam diferencialmente expressos) e gráficos de M. A. plot, Volcano plot e Histograma. Então, para encontrar os genes que sempre são indicados como diferencialmente expressos, independentemente dos módulos usados, o R-Peridot define um consenso entre os resultados dos diversos módulos, o qual é usado para realizar novas análises como a criação de um HeatMap, um dendrograma, um gráfico PCA e gráficos de ontologia dos genes usando o pacote clusterProfiler do R. Através desses recursos, o projeto pretende ajudar pesquisadores, com ou sem conhecimentos prévios de programação, a realizar análises de expressão diferencial de forma simples e eficiente. A ferramenta vem sendo usada em estudos para encontrar miRNAs diferencialmente expressos em amostras tumorais e em pacientes com Leucemia Linfocítica Crônica (LLC) ou Leucemia Mielóide Aguda (LMA). Neste estudo sobre leucemia, foram encontrados 3 possíveis novos miRNAs relacionados à LLC.



RNA e transcriptômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO DAS PEROXIDASES NO TRANSCRIPTOMA DE VITIS SPP.

Autores: Raphael Gomes Paiva Silva¹; Carolline de Jesús Pires²; Artemisa Nazaré Costa Borges²; Flávia Tadeu de Araújo²; Mitalle Matos²; João Pacifico Bezerra-Neto²; Ana Maria Benko-Iseppon²; Flávia Figueira Aburjaile²;

E-mail para correspondência: raphaelgomes2002@hotmail.com

Instituições: ¹Colégio Militar do Recife, Recife, PE, Brasil; ²Universidade Federal de Pernambuco, Recife, PE, Brasil;

Palavras-chave: Estresse biótico; PR-9; Videira

Apoio: CAPES, FACEPE, CNPq

PR-9 ou peroxidases são enzimas-chave no processo de construção da parede celular. Além disso, tem-se sugerido que as peroxidases extracelulares ou de parede potencializem a resistência contra fitopatógenos em plantas por meio da construção de uma barreira na parede celular, prejudicando a penetração e o espalhamento do patógeno. Diante disso, este trabalho teve como objetivo identificar e caracterizar potenciais peroxidases no transcriptoma de *Vitis* spp. sob infecção de *Xanthomonas campestris*. Para isso, realizou-se uma busca pelas sequências sonda (*seed-sequences*) da família PR-9 no banco UniProt, as quais foram alinhadas (tBLASTn; *cut-off* ≤ e-05) contra o transcriptoma de *Vitis* spp. As sequências obtidas foram anotadas via BLASTx contra o banco não redundante (nr) do NCBI e UniProt. Posteriormente, essas sequências foram traduzidas por meio da ferramenta TransDecoder. A busca por domínios e motivos conservados proteicos foi realizada com o Batch CD-search e bancos do Pfam e InterProScan. A predição do peso molecular e o ponto isoelétrico das sequências foi realizada com o JVirGel 2.0. Já a presença de peptídeo sinal foi detectada por meio do SignalP 4.1 Server. Dessa forma, foi possível identificar 81 potenciais peroxidases expressas em *Vitis* spp. sob infecção de *X. campestris*, apresentando o domínio completo. A massa molecular das potenciais peroxidases variou de 12,05 a 73,43 kDa e o ponto isoelétrico variou de 4,14 a 10,43. Além disso, 26 potenciais peroxidases apresentaram peptídeo sinal, indicando que tal grupo proteico é direcionado para uma via secretora. O presente trabalho agregou informações à caracterização estrutural desse importante grupo proteico em *Vitis* spp. As características das peroxidases mineradas de videira são partilhadas com outros grupos vegetais, demonstrando a sua conservação. Adicionalmente, a constatação de sua expressão sob infecção de *X. campestris* sugere sua participação em processos de defesa vegetal. Entretanto, estudos posteriores de análise da expressão gênica serão realizados a fim de compreender como este grupo proteico é modulado frente aos estresses bióticos.



RNA e transcriptômica

PROSPECÇÃO IN SILICO DE GENES DE RESISTÊNCIA DA CLASSE V EM ACESSOS DE VITIS SPP. INOCULADOS COM XANTHOMONAS CAMPESTRIS PV. VITICOLA

Autores: Roberta Lane de Oliveira Silva¹; Jéssica Barboza da Silva¹; João Pacífico Bezerra Neto¹; Flávia Figueira Aburjaile¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: lane.roberta@gmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Videira; RPW8; Estresse biótico

Apoio: CAPES, CNPq e FACEPE

Os genes RPW8 pertencem à classe V de genes de resistência (*R*) e se destacam por não apresentar homologia significativa com outras proteínas, inclusive entre os próprios genes *R*, sendo caracterizados por possuir um domínio transmembrana N-terminal e um motivo de cauda espiralada ancorado à membrana. Embora essa classe de genes ainda permaneça pouco estudada, já se sabe que os RPW8 estão relacionados à regulação de vias de sinalização que promovem a resposta hipersensível (HR) e resistência a patógenos. Este trabalho objetivou selecionar e caracterizar transcritos RPW8 em acessos de videira contrastantes quanto a resistência a *Xanthomonas campestris* pv. *viticola*. Dessa forma, sequências proteicas relacionadas à RPW8 foram selecionadas no banco de dados do UniProt para alinhamento via tBLASTn no transcriptoma de videira. A anotação e tradução das sequências, bem como a busca por domínios conservados foi realizada a partir do ORFfinder e Batch CD-search, respectivamente. Adicionalmente, seu ponto isoelétrico (pI) e o peso molecular (pM) foram preditos por meio do software JVirGel 2.0, a localização subcelular utilizando o Cell-PLoc 2.0 e a ancoragem no genoma de *Vitis* através do banco Phytozome. Foram identificadas quatro sequências candidatas a membros do grupo RPW8. Dessas, três apresentaram domínios conservados e uma estava com quebra na região C-terminal. As proteínas traduzidas apresentaram ORFs variando entre 164 a 222 aa. O pI variou de 6,93 e 8,86 e o pM de 18,92 a 24,92 kDa, corroborando ao verificado na literatura. Dois transcritos apresentaram localização nuclear e um estava direcionado tanto para o núcleo quanto citoplasma. A ancoragem permitiu identificar dois transcritos ancorados no cromossomo 7. Apesar de iniciais, os resultados encontrados neste estudo disponibilizam novas informações acerca do número e características básicas dos RPW8, família possivelmente ligada ao processo de defesa de videira frente à bactéria *X. campestris*, as quais poderão ser confirmadas em análises posteriores de expressão gênica via RT-qPCR.



RNA e transcriptômica

IDENTIFICAÇÃO E CARACTERIZAÇÃO IN SILICO DE MICRORNAS CONSERVADOS DE PLANTAS EM ESTS DE *JATROPHA CURCAS* L.

Autores: Samara Alves da Silva¹; Éderson Akio Kido¹; Marislane Carvalho Paz de Souza¹; Elvson Wallacy da Silva¹; Vinicius Torres Guerra¹; Valquíria da Silva¹;

E-mail para correspondência: samara_alves_s@hotmail.com

Instituições: ¹Laboratório de genética molecular, Universidade Federal de Pernambuco;

Palavras-chave: pinhão manso; transcriptômica; microRNA

Apoio: CNPq, CAPES

O pinhão manso (*Jatropha curcas* L.) é uma planta rústica, perene e adaptável a uma gama de ambientes e condições edafoclimáticas, as suas sementes apresentam de 27-32% de óleo, com alta concentração de ácidos graxos insaturados. Os microRNAs (miRNAs) são pequenos RNAs endógenos não codantes que atuam na regulação gênica pós-transcricional. Então, aproveitando-se do crescente número de miRNAs identificados nas mais diversas plantas, da conservação destes em plantas evolutivamente relacionadas e da presença de bases de dados especializadas em miRNAs, que este trabalho teve como objetivo identificar e caracterizar *in silico* miRNAs maduros conservados de plantas em ESTs disponíveis de *J. curcas*. Para isso, executou-se um BlastN (e-value < 1e⁻³) entre as sequências de miRNAs e as ESTs de *J. curcas*, onde as ESTs foram o *database*. As sequências de ESTs que continham os miRNAs maduros foram alinhadas via BlastX com bancos de dados de proteínas das espécies *Manihot esculenta*, *Ricinus communis* e *Jatropha curcas*, permanecendo retidas somente as que não codificam proteínas, as quais foram utilizadas para predição das estruturas secundárias em forma de grampo a partir da ferramenta MFOLD. Através do banco de dados mirBASE, identificou-se os *targets* genes dos miRNAs encontrados, os que continham informação insuficiente foram submetidos ao psRNATarget. No total, correram 866 alinhamentos, onde 518 miRNAs alinharam com 187 ESTs de *J. curcas*. Entre os 187 alinhamentos exclusivos das ESTs, apenas 35 não codificam proteínas e estes foram submetidos a plataforma MFOLD. No total, foram encontrados 6 miRNAs. A família miR167 alinhou com 2 ESTs, GW881255.1 e GW879969.1, em 37 espécies diferentes e atingiu 407 alinhamentos dentre os 866, cerca de 47% do total; esta família tem importante papel no controle do crescimento e no desenvolvimento de órgãos vegetativos e florais em dicotiledôneas. 246 miRNAs da família miR16 alinharam com a EST GW880030.1, representando cerca de 28% do total de alinhamentos; isto indica que este, assim como a família miR167, é um grupo conservado. O miRNA gma-miR5368, que é sensível à seca e é reprimido quando está sob efeito deste estresse, alinhou com 2 ESTs diferentes, FM889530.1 e GT975674.1, cada uma com 1 alinhamento. Deste modo, é evidente que o estudo dos miRNAs e de seus alvos é uma fonte valiosa de informação para um melhor entendimento da expressão gênica de *Jatropha curcas* e para outras espécies vegetais de interesse econômico.



RNA e transcriptômica

USO DO R-PERIDOT PARA ANÁLISE DE EXPRESSÃO GÊNICA DIFERENCIAL A PARTIR DE DADOS DE RNA-SEQ

Autores: Stela Mirla da Silva Felipe¹; Christina Pacheco¹; Raquel Martins de Freitas¹; Jonathan Elias Rodrigues Martins¹; Danilo Lopes Martins²;

E-mail para correspondência: stelamirla@gmail.com

Instituições: ¹Universidade Estadual do Ceará - UECE; ²Universidade Federal do Rio Grande do Norte;

Palavras-chave: RNA-seq; Expressão gênica; R-peridot

Apoio: CNPQ CAPES

Introdução: Com o sequenciamento de nova geração (NGS) é possível obter o perfil transcriptômico em diferentes condições fisiológicas, tais como as geradas pela atividade física. Buscou-se produzir um transcriptoma – RNA-seq, relacionado à atividade física experimental em ratos e, para analisar os dados gerados, optou-se pela plataforma R-Peridot (BioME - Instituto Metrópole Digital - IMD/UFRN). O R-Peridot consiste de uma ferramenta que reúne pacotes da linguagem R para análises de expressão gênica diferencial com uma interface compatível para pesquisadores de áreas biomédicas. **Objetivo:** Utilizar a ferramenta R-Peridot para análise de expressão gênica diferencial a partir de dados de sequenciamento (RNA-seq) do músculo esquelético (sóleo) de *Rattus norvegicus* submetidos a uma sessão de exercício exaustivo em esteira adaptada a roedores. **Metodologia:** Foi feito RNA-seq de 8 amostras (n=4 controle, n=4 treinado) em músculo sóleo de ratos, com velocidade inicial de 0,5 Km/h e com aumento de 0,2 KM/h a cada três minutos, até a exaustão do animal (CEUA/UECE: 1592060/2014). O músculo sóleo foi dissecado e submetido aos protocolos bioquímicos e RNA-seq. Os dados obtidos entraram como input para o R-Peridot. As análises de expressão gênica diferencial foram realizadas a partir de uma lista de *count reads* (formato tsv) com a média de expressão de 12816 genes nas 8 amostras. Os pacotes escolhidos no R-peridot foram: *EBseq*, *Deseq2* e *EdgeR*. O *output* apresentou dados sobre os parâmetros: *Basemean*, *Fold change* (FC) 1.5, *False Discovery Rate* (FDR) 0.05, com *p-value* 0.05, além de gráficos de apresentação de dados. **Resultados e Discussão:** Após a análise FDR, foram obtidos 65 genes (FC 0,80 ±1,13) diferencialmente expressos com o pacote *EBseq*, 54 genes sob *up regulation* e 11 genes sob *down regulation*. Com o *EdgeR* foram obtidos 34 genes (FC 0,75 ±2,43), 18 *up* e 16 *down regulation*. O *DESeq2* apresentou 30 genes (FC -0.11±0.7), 13 *up* e 17 *down regulation*. Com os resultados de cada pacote, o R-Peridot produziu gráficos: MA plot, Volcano plot, Histograma, Heatmaps e Boxplots. Após a análise de expressão diferencial, a ferramenta gerou uma lista de consenso de 16 genes diferencialmente expressos nos 3 pacotes. A partir dos dados apresentados pela plataforma, foi possível dar continuidade às análises de ontologia gênica em bancos de dados específicos, de forma que a ferramenta R-Peridot pode ser um passo inicial facilitador de interesse, nas análises biomédicas.



RNA e transcriptômica

PREDIÇÃO DE UNITAGS SUPERSAGE ESPERADAS NA EXPRESSÃO GÊNICA DE VIGNA RADIATA E SIMILARIDADE COM UNITAGS OBSERVADAS EM *V. UNGUICULATA* SOB ESTRESSE DE SECA

Autores: Valquíria da Silva¹; Elvson Wallacy da Silva¹; Ederson Akio Kido¹;

E-mail para correspondência: valquiria-dsilva@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco;

Palavras-chave: Bioinformática; Transcriptômica; Estresse abiótico

Apoio: Facepe, CNPq

O feijão-caupi exerce nas regiões Norte e Nordeste um papel de fundamental importância como fonte de alimento, emprego e renda. Entretanto a seca é um dos principais fatores limitante da produção desta leguminosa, sendo o uso de materiais de elite a melhor alternativa para minimizar o problema. O presente trabalho objetivou identificar uma população esperada de *unitags* (26 pb) SuperSAGE previstas em um potencial genoma expresso de feijão-mungo [*Vigna radiata* (L.) Wilczek], verificando a similaridade destas com *unitags* observadas em *V. unguiculata* sob estresse de seca, aproveitando-se da sintonia entre essas leguminosas e das anotações informativas de gene/ função relativas ao genoma, transcriptoma e proteoma de *V. radiata*. Assim, foram identificados em transcritos de *V. radiata* potenciais sítios de restrição da enzima *NlaIII* (CATG), junto aos quais foram extraídos fragmentos de 26 pb, simulando as possíveis *unitags* SuperSAGE. As possíveis *unitags* digitais (306.084), foram posteriormente alinhadas via BLASTn (*e-value cutoff* de e^4) contra *unitags* observadas de feijão-caupi, oriundas de ensaio com dois acessos contrastantes sob estresse de seca (desidratação radicular de até 150 minutos). Assim, 18.186 *unitags* digitais de *V. radiata* foram associadas a 23.006 *unitags* de *V. unguiculata*. A partir da ferramenta BLASTx, os transcritos relativos às 18.186 *unitags* foram associados aos respectivos produtos proteicos. Ao elege-se o melhor alinhamento para cada transcrito, o número de proteínas restringiu-se em 14.444, sendo estes também os locos do genoma de *V. radiata* que seriam expressos. O número total de proteínas compartilhadas, independente do estresse, pelas duas espécies, identificado via BLASTx, de *ESTs* de *V. unguiculata* (NCBI) com proteínas de *V. radiata*, foi de 16.624 proteínas, dos quais 10.178 foram comuns aos 14.444 observados anteriormente. Assim, o número total de proteínas similares deve conter além das 16.624 também as 4.266 proteínas identificadas somente através de transcritos de *V. radiata* contendo *unitags* de *V. unguiculata*, totalizando 20.890 proteínas similares. A similaridade total entre *V. radiata* e *V. unguiculata*, considerando o número total de proteínas compartilhadas pelas duas espécies correspondeu a 59% do total previsto para *V. radiata*; e do total de proteínas similares detectadas, a expressão de potenciais proteínas relacionadas com o estresse estudado foi de cerca de 70%.

RNA e transcriptômica

TRANSCRIPTÔMICA SUPERSAGE DO FEIJÃO-CAUPI E GENÔMICA COMPARATIVA PARA IDENTIFICAÇÃO DE POTENCIAIS QTLs DE INTERESSE AGRONÔMICO

Autores: Vinicius Torres Guerra¹; Juliana Roberta Muniz de Oliveira¹; Ana Maria Benko Iseppon¹; Éderson Akio Kido¹;

E-mail para correspondência: vini.guerra2@hotmail.com

Instituições: ¹Universidade Federal de Pernambuco (UFPE);

Palavras-chave: Bioinformática; Feijão-caupi; Estresse abiótico

Apoio: CNPq, Embrapa Meio-Norte

Déficit hídrico é um dos principais fatores responsáveis por restringir a produtividade, afetando diretamente o plantio de culturas importantes. O feijão-caupi [*Vigna unguiculata* (L.) Walp] é a maior fonte de proteína vegetal consumida na região Nordeste, a qual é amplamente castigada por estiagens. Como a resposta a estresses abióticos normalmente é poligênica e envolve um conhecimento multidisciplinar, identificar alelos contribuintes para a tolerância ao estresse em questão acaba tornando a cultura mais rentável e produtiva. Este trabalho visou identificar potenciais locos gênicos (e QTLs) atuantes na resposta ao estresse em estudo, se utilizando da genômica comparativa entre espécies de *Vigna* filogeneticamente próximas e genomicamente melhor elucidadas (*V. radiata* e *V. angularis*). Embora o feijão-caupi não disponha ainda de genoma devidamente montado e anotado ao nível das duas espécies de *Vigna* citadas, na base de dados *Genbank* do NCBI, a cultura dispõe de milhares de transcritos *unitags* SuperSAGE, com 26pb, obtidas pela rede NordEST (UFPE), a partir de RNAs de raízes de dois acessos contrastantes fenotipicamente para seca, conforme a Embrapa Meio-Norte. Para tanto, *unitags* foram devidamente ancoradas em ESTs (BlastN, $e\text{-value} < e^{-4}$), e as ESTs contendo *unitags* foram mapeadas nos 11 cromossomos dos genomas de *V. radiata* e de *V. angularis* (BlastX, $e\text{-value} < e^{-10}$). Em *V. angularis* 17.148 locos foram mapeados no acesso tolerante e 18.029 no sensível, já para *V. radiata*, esses números foram de 17.126 e 17.913, respectivamente. Com base nas ESTs de *V. unguiculata* compartilhadas pelos dois genomas, sete grupos foram considerados com relativa sintonia (acima de 50% de ESTs compartilhadas), e dentro de um desses grupos, 20 genes ligados foram identificados, compreendendo 308.721 pb em *V. radiata* e 357.461 pb em *V. angularis*, os quais se apresentaram colineares e de respostas divergentes: prioritariamente induzida pelo acesso tolerante (14 de 20), e reprimida ou n.s ($p < 0,05$ em teste Audic-Claverie) pelo acesso sensível (12 de 20). Este conjunto de genes ligados foi declarado um potencial QTL para seca. Os alvos compreendidos estão sendo validados quanto às suas expressões, por RTqPCR e se confirmadas as regulações, os alvos e a região genômica como um todo poderão ser úteis na seleção de novos materiais, em programas de melhoramento de feijões ou mesmo de outras leguminosas, além de servir de base para pesquisas futuras.



RNA e transcriptômica

PROTEÍNAS QUINASES DEPENDENTES DE CICLINA EM RESPOSTA A INFECÇÃO POR XANTHOMONAS CAMPESTRIS EM VIDEIRA

Autores: Jéssica Barboza da Silva¹; Roberta Lane de Oliveira Silva¹; João Pacífico Bezerra Neto¹; Flávia Figueira Aburjaile¹; Mireli de Santana Rêgo¹; Ana Maria Benko-Iseppon¹;

E-mail para correspondência: jessica.ufpebio@gmail.com

Instituições: ¹UNIVERSIDADE FEDERAL DE PERNAMBUCO - UFPE;

Palavras-chave: Vitis; CDK; Cancro bacteriano

Apoio: CAPES e CNPq.

As proteínas quinases estão entre as classes de genes envolvidos no processo de defesa vegetal e destacam-se por participarem de quase todas as vias de transdução de sinais. As quinases dependentes de ciclina (CDKs) possuem papel crítico no controle da progressão da mitose e citocinese, participam do controle da transcrição gênica e outros processos, auxiliando na manutenção dos processos fisiológicos normais. Diante disso, esse trabalho se propôs a identificar e caracterizar transcritos codificadores de proteínas CDK em acessos de *Vitis* infectados com *Xanthomonas campestris* pv. *viticola*. Inicialmente, foram selecionadas sequências sonda curadas da subfamília gênica CDK no banco de dados NCBI (<http://www.ncbi.nlm.nih.gov/>). As sequências foram alinhadas via tBLASTn contra o transcriptoma de videira, adotando um *cut-off* de e^{-5} . As sequências candidatas obtidas em videira, foram anotadas, traduzidas e tiveram seus domínios conservados identificados com o auxílio do Transdecoder e Batch CD-search, respectivamente. Por fim, foi realizada a predição do ponto isoelétrico (pI) e peso molecular (pM) utilizando o JVirGel 2.0, bem como a localização subcelular determinada através do Cell-PLoc 2.0. Adicionalmente, foi determinada a posição dos loci associados aos transcritos obtidos, utilizando para isto o genoma de *Vitis vinifera* depositado no banco *Phytozome*. Foram identificadas 34 sequências candidatas relacionadas com o gene CDK no transcriptoma de videira. As proteínas traduzidas apresentaram ORFs variando entre 273 a 741 aa, onde 22 transcritos apresentaram domínios conservados completos. Desse total, 15 possuíam o domínio CDK-9, quatro o domínio CDK-1, dois o domínio CDK-8 e um o domínio CDK-7. As proteínas candidatas possuem pI entre 6,87 e 10,13 e pM variando entre 31,34 e 76,6 kDa. Todas as proteínas apresentaram localização subcelular associados ao núcleo, sendo 18,2 % direcionadas para o citoplasma corroborando com os dados da literatura. Os transcritos ancoraram nos cromossomos 7, 8, 12, 14, 17, sendo a grande maioria ancorada no cromossomo 14. Os resultados obtidos contribuem para uma melhor compreensão das características moleculares da subfamília gênica CDK e sua participação no processo de defesa vegetal em videira, disponibilizando candidatos para validação via RT-qPCR.



RNA e transcriptômica

IDENTIFICAÇÃO DE AQUAPORINAS NO TRANSCRIPTOMA DA CATINGUEIRA [CENOSTIGMA PYRAMIDALE (TUL.) L.P. QUEIROZ]

Autores: Luana Taís Bezerra Silva¹; João Pacifico Bezerra Neto²; Grazielly Beserra Calixto Venancio³; Gabriella Frosi Albuquerque Figueiroa Faria⁴; Flavia Tadeu de Araújo⁵;

E-mail para correspondência: luastais16@gmail.com

Instituições: ¹Universidade de Pernambuco; ²Universidade Federal de Pernambuco; ³Centro Universitário UniFBV | Wyden; ⁴Universidade Federal de Pernambuco; ⁵Universidade Federal de Pernambuco;

Palavras-chave: Bioinformática; Fabaceae; proteína de membrana

Apoio: CAPES, CNPq, FACEPE.

As aquaporinas (AQPs) são proteínas de canais de água pertencentes à superfamília MIP (*Major Intrinsic Protein*), presente em todos os reinos. São geralmente descritas como tetraméricas, compreendendo quatro canais independentes. Apesar de sua estrutura geral ser bastante conservada, existem regiões específicas que determinam diferenças na seletividade do poro, o que permite classificar as AQPs em cinco subfamílias para a maioria dos vegetais. No presente trabalho, foi realizada a identificação e caracterização de AQPs no genoma expresso da catingueira, uma leguminosa arbórea da Caatinga nordestina. A busca ocorreu por meio de tBLASTn contra o transcriptoma da catingueira utilizando 'seed sequences' de AQPs vegetais do banco Uniprot (*cut-off* de 1e-05). Para confirmação das AQP-candidatas foi realizada uma busca por domínios no Pfam, além de um enriquecimento GO. Foram identificadas 289 AQP-candidatas expressas em *C. pyramidale*, sendo 198 com o domínio conservado PF00230.19 (MIP - *Major intrinsic protein*). Após a análise de clusters gênicos no COG, obtivemos a anotação COG0580 para 153 sequências, correspondendo a canais que permitem o movimento osmótico celular, enquanto que os termos GO mais prevalentes foram os GO:0016021 (*cellular component: integral component of membrane*); GO:0015250 (*molecular function: water channel activity*), confirmando nossa anotação para estas AQPs. As sequências identificadas, após alinhamento múltiplo, confirmam a presença dos motivos NPA, bem como filtros ar/R, responsáveis pela seletividade do poro destes canais de membrana. A caracterização destes canais em catingueira, permitirá uma maior compreensão do aparato vegetal para a regulação do fluxo de água em nível celular, fornecendo candidatos para análise mais detalhadas e possível aplicação na engenharia genética.

RNA e transcriptômica

ANÁLISE PRIMARIA DE TRANSCRIPTOMA DE TECIDO TESTICULAR DE HAMSTERS (MESOCRICETUS AURATUS) INFECTADOS COM VÍRUS ZIKA

Autores: Walter Felix Franco Neto¹; Samir Mansour Moraes Casseb¹; Karla Fabiane Lopes de Melo¹; Milene Silveira Ferreira¹; Gustavo Moraes Holanda¹;

E-mail para correspondência: walter96neto@gmail.com

Instituições: ¹Instituto Evandro Chagas; ²Universidade Federal do Pará;

Palavras-chave: Transcriptoma; vírus zika; MicroRNA

Apoio: CNPq, SVS, MS, IEC.

Introduction: The Zika Virus (ZIKV) belongs to the family Flaviviridae, genus Flavivirus. It was first isolated in 1947 in Zika Forest, Uganda, from a nonhuman primate that served as a sentinel for the study of the yellow fever virus (VFA). Its genome consists of single-stranded ribonucleic acid (RNA) with positive polarity, the messenger RNA itself (mRNA), or the translation of viral proteins. The mRNAs are fundamental for small non-coding RNAs conserved throughout evolution to perform their functions. These small RNAs are known as microRNAs and have the ability to regulate gene expression. This virus has been causing a major public health problem due to its connection with malformation of the nervous system of fetuses with mothers affected by this virus, as well as its sexual transmission. This situation is an important health problem and describes profiles of possible RNAs that may interfere in the replication in cells linked to the sexual transmission of Zika Virus, which is very important for better knowledge the viral infection. **Materials and Methods:** Male hamsters (*Mesocricetus auratus*) infected with Zika Virus were used, RNA was extracted from the testicles of animals with 3 days post infection and uninfected animals (Controls) with the Maxwell® 16 Tissue LEV RNA Purification Kit Kit (Promega, USA). This RNA was used for a transcriptome using the NextSeq 550 platform (Illumina, USA), the data from this sequencing were analyzed adopting the Tuxedo protocol (Trapnell, 2012) and statistical analysis applying the R project platform with the Dynamic Transcriptome Analysis (DTA) (Schwalb, 2012). **Result:** The total number of paired cleaned readings varied among the 8 sequenced samples. More than 96% of the readings mapped to the host genome with about 80% of the fragments counted and mapped to gene regions and 20% to intergenic areas of the host genome. Analysis and comparison of mRNA expression profiles, at different time points after VZIK infection, revealed that, in total, 1332 genes had changes of 2 or more times in either direction. **Conclusion:** In general, our results showed large changes in the anaerobic transcriptome after VZIK infection, both in coding and in non-coding RNAs.