

Motifs Discovery Using Profile HMM and Evolutionary Algorithms

Jader M. Caldonazzo Garbelini
Andre Y. Kashiwabara
Danilo Sipoli Sanches

Universidade Tecnológica Federal do Paraná

Abstract

The study on shared patterns of biological sequences called motifs, is important to help researchers understand how organisms work. Several inferences can be made through these patterns, as for example, function, secondary and tertiary structure of proteins and location promoters and regulatory regions of genes. Experimental methods, such as DNase Footprint, gel-shift, reporter construct assays, ChIP has been used to determine the position of motifs with relative success. However, find hundreds or thousands of likely candidate sites using experimental techniques demand a lot of time and money. This makes computational methods of aligning an excellent way in search for motifs. They consist basically in perform multiple sequence alignment (MSA) of target sequences based on similarity of the residues. Has already been proven that the MSA is a computational complexity problem Non-Polynomial (NP). Problems of this class cannot be solved in polynomial time, i.e., depending on the number of input sequences, the alignment can be a prohibitive time. Although brute-force computational techniques can find always a optimum alignment, they cannot be employed because of the restrictions imposed by the MSA. Therefore, heuristic methods need to be applied. Profile Hidden Markov Model (PHMM) has been used with much success in Bioinformatics in the solution of the MSA. Although PHMM is a very robust method, in the course of its implementation, it can be pressure on local maximum or minimum. Evolutionary Algorithms (EAs) has been shown to be effective in solving large problems that have a complex search space. The EAs are nondeterministic algorithms, i.e., there is no certainty that they will find the optimal solution. However, you can find most of the time very close to the great solutions at relatively low computational time compared with mathematical programming techniques. In view of the above, this work is being developed a new methodology using PHMM and EAs, aiming to solve the problem of the MSA more efficiently.

Keywords: Motifs, Multiple Sequence Alignment, Evolutionary Algorithms, Profile Hidden Markov Model

IDENTIFICAÇÃO TAXONÔMICA MOLECULAR DE LEVEDURAS ASSOCIADAS A PLANTA MEDICINAL ALECRIM-DO-CAMPO (*Baccharis dracunculifolia* DC -Asteraceae)

Gilberto de Aguiar Pereira¹; Renan Cantanti Marques¹; Giovanna Capello Batista¹; Elisete Pains Rodrigues¹; Fernando Gomes Barcellos¹.

1 – Departamento de Biologia Geral, Laboratório de Genética de Microrganismos, Universidade Estadual de Londrina, 86057-970, Londrina, Paraná. E-mail: gilbertopperera@gmail.com.

Leveduras são habitantes comuns em muitas espécies de plantas, no entanto, o conhecimento sobre a diversidade de leveduras associadas (endofíticas e epifíticas) às plantas, ainda encontra-se em um estágio inicial; considerando as plantas medicinais, este conhecimento é ainda mais limitado. As plantas medicinais, devido às suas propriedades terapêuticas são fontes promissoras para o isolamento, a identificação e estudos da diversidade de microrganismos associados, como as leveduras, possibilitando estudos de bioprospecção de microrganismos com potencial biotecnológico em diversas áreas de interesse. Desta forma, o presente estudo, teve como objetivo, isolar e identificar taxonomicamente (gênero e espécie) leveduras associadas a folhas, caules e flores da planta medicinal alecrim-do-campo (*Baccharis dracunculifolia* DC- Asteraceae), uma planta reconhecida como a principal fonte de exsudatos utilizados pelas abelhas (*Apis mellifera*) para produzir o própolis verde. Para realizar o desenvolvimento deste estudo, amostras de folhas, caule e flores de *B. dracunculifolia* DC (Asteraceae) foram coletadas na fazenda escola da UEL. O isolamento e purificação das leveduras ocorreu nos meios sólidos BDA e TSA a partir dos quais colônias selecionadas como representantes de seus morfotipos tiveram suas sequências de ITS1-5,8S-ITS2 do rDNA amplificadas e sequenciadas. As sequências obtidas foram analisadas utilizando os programas Phred/Phrap, editadas com o programa BioEdit, depositadas no banco de dados de nucleotídeos *GenBank* e identificadas taxonomicamente (gênero e espécie) a partir de análises de similaridade utilizando a ferramenta blastN (NCBI). Adicionalmente, uma árvore filogenética utilizando o software MEGA, foi construída para determinar e confirmar as relações evolutivas entre os nossos isolados com as sequências de nucleotídeos de espécies tipo ou de referência. Como resultados obtivemos que a amplificação da região ITS1-5,8S-ITS2 do rDNA gerou fragmentos de aproximadamente 550-700 pb; a análise por similaridade das sequências de nucleotídeos da região ITS1-5,8S-ITS2 do rDNA revelaram a ocorrência de cinco espécies de leveduras na planta medicinal *B. dracunculifolia* DC (Asteraceae): *Starmerella bombicola*, *Aureobasidium leucospermi*, *Aureobasidium pullulans*, *Rhodotorulla mucilaginosa* e *Occultifur externus* e a árvore filogenética construída com o uso do programa MEGA possibilitou a confirmação das identificações taxonômicas realizadas, ainda possibilitou que os isolados de *Aureobasidium pullulans* pudessem ser identificados até o nível de subespécie. Em conclusão, os nossos resultados revelaram que a comunidade de leveduras associadas as folhas, caules e flores da planta medicinal alecrim-do-campo é diversificada, sugerindo que os segmentos aéreos de plantas dessa espécie, representam locais em potencial para a bioprospecção de microrganismos com possíveis aplicações em diversas áreas de interesse biotecnológico.

Suporte Financeiro: Capes, CNPq e Fundação Araucária.

Annotation of Transposable Elements in the Coffee Genome (*Coffea canephora*)

Stéphanie Bocs³, Alexis Dereeper², Thomas Gayraud¹, Véronique Jamilloux⁴, Philippe Lashermes² and Romain Guyot¹

¹ IRD UMR DIADE, EVODYN, BP 64501, 34394 Montpellier Cedex 5, France

² IRD UMR RPB, DIVA, BP 64501, 34394 Montpellier Cedex 5, France

³ CIRAD (Centre de coopération internationale en recherche agronomique pour le développement, UMR AGAP, F-34398 Montpellier, France

⁴ INRA, Unité de Recherches en Génomique-Info (UR INRA 1164), Centre de recherche de Versailles, 78026 Versailles Cedex, France

Key words: Coffee genome, Transposable elements, Annotation

Recently, 54.4 million of Roche 454 sequences, 131,412 Sanger BAC-end sequences and 60X Illumina coverage of the 710 Mb genome of a *C. canephora* Double Haploid accession (DH200-94) were generated, assembled and anchored to a genetic map by the COFFEE GENOME CONSORTIUM. The *C. canephora* genome sequence represents a formidable resource to understand the chromosome structure and the genome evolution in the *Coffea* genus. It is now well established that plant genomes are dynamic structures submitted to a wide range of modifications through the activity of Transposable Elements (TEs). TEs are mobile sequences that share several key properties such as the ability to move from one chromosome location to another one, to amplify their copy number within the host genome by burst. TEs activity contributes to the chromosome structure, organization and evolution. Particularly, TEs play a major role in creating structural variation and genetic diversity in plant genomes.

Here we present the identification and classification of TEs in the 568 Mb of genomic sequences from the *C. canephora* using a combination of *ab initio*, similarity and structure search approaches. Particularly, we used the REPET package V.2.1-RC developed by URGI (Flutre et al., 2011) to identify, classify and annotate TEs. We found that about half of the genomic sequences produced are composed of TEs similarly to other sequenced crop species such as banana, papaya, castor bean and soybean. Class I LTR retrotransposons represent the vast majority of identified elements, accounting to 42% of the genome assembly. *Gypsy* elements clearly outnumbering *Copia* elements since *Ty3-Gypsy* family covers 24.1% of the genome. Interestingly active non-autonomous LTR retrotransposons elements were detected and classified into a new subgroup of non-autonomous elements containing a *capsid* domain but lacking the region involved in their mobility. Finally in an attempt to study conservation of LTR retrotransposons between coffee and Angiosperm genomes we identified an outstanding conservation of several families across distantly related plant species, suggesting that conservation of such elements or horizontal transfer events might be more frequent than recognized actually.

Supported by This research was supported by Agropolis Fondation through the “Investissement d’avenir” program (ANR-10-LABX-0001-01) under the reference ID 1102-006 (Retro-cof)

Perfis de expressão em *Streptococcus agalactiae* indicam sub expressão de ncRNAs.

Ivan Rodrigo Wolf¹, Alexandre Rossi Paschoal², Douglas Silva Domingues³, Rogério Fernandes de Souza¹, Laurival Antônio Vilas-Boas¹

¹ Universidade Estadual de Londrina

² Universidade Tecnológica Federal do Paraná

³ Universidade Estadual Paulista - Campus de Rio Claro

Streptococcus agalactiae (GBS), são responsáveis por infecções em humanos, cachorros, gatos, cavalos, macacos, gado e peixes, assim são responsáveis por perdas econômicas na indústria e constituem um problema de saúde pública. Análises genéticas mostram grupos de genes de virulência presentes em regiões conservadas no genoma desta bactéria. Embora RNAs não codificantes (ncRNAs) já tenham sido identificados participando como moduladores de expressão de fatores de virulência em outras bactérias, somente um estudo de ncRNAs está disponível para *S. agalactiae*. Deste modo desenvolvemos uma análise híbrida para identificação e caracterização de ncRNAs candidatos no genoma da linhagem NEM316. Em adição o perfil de expressão para os ncRNAs candidatos foi estabelecido com auxílio de transcriptomas disponíveis em bancos de dados para a linhagem selecionada em dois meios: Caldo Nutriente e Leite, este último um ambiente favorável a manifestação de sistemas de virulência. Os resultados mostraram 82 ncRNAs candidatos e observou-se que entre estes 7 possuíam como prováveis alvos genes associados a processos de virulência e outros 13 possuíam como alvos genes preditos como sendo associados a fatores de virulência. Os perfis de expressão avaliados permitiram observar que a maioria dos candidatos é sub expresso no meio de leite, e também mostraram dois padrões de relações entre candidato e provável alvo (I) inversamente proporcionais (e.g. ncRNA_31, ncRNA_32, ncRNA_81), quando o ncRNA é sobre expresso a maioria dos alvos é sub expresso e vice-versa, e (II) proporcionais, quando ncRNA e prováveis alvos tem o perfil de expressão semelhante (e.g. ncRNA_25, ncRNA_26, ncRNA_36, ncRNA_58). Ainda não é claro o padrão esperado quando os ncRNAs estão em contato com seus prováveis alvos, entretanto em situações onde ncRNAs de ação negativa estão em maior quantidade que os mRNAs alvo, a expressão gênica tende a ser fortemente reprimida, porém, quando o inverso é observado existe pouco efeito sobre a expressão. Dessa forma, no presente estudo, os ncRNAs que exibiram o padrão (I) podem estar reprimindo a ação de seus alvos, quando em meio de cultura, e aqueles com padrão (II) podem alternativamente estimular a expressão quando em leite. Finalmente, mais estudos irão permitir uma melhor compreensão do papel de ncRNAs nesse importante microrganismo.

Palavras-Chave: *Streptococcus agalactiae*, ncRNAs, Virulência, Expressão Diferencial

Agência de fomento: CNPq.

PLANTE-MIR DB: UM REPOSITÓRIO PARA MICRORNAS RELACIONADOS A ELEMENTOS TRANSPONÍVEIS

APR Lorenzetti¹, GYA de Antonio², AR Paschoal² e DS Domingues^{1,3}

¹ Programa de Pós-graduação em Genética e Biologia Molecular, Universidade Estadual de Londrina, Londrina, Paraná.

² Laboratório de Bioinformática, Universidade Tecnológica Federal do Paraná, Cornélio Procopio, Paraná.

³ Departamento de Botânica, Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Rio Claro, São Paulo.

Elementos transponíveis (ETs) constituem grandes porções de genomas vegetais e são conhecidos por desempenhar papel crucial em sua organização e evolução. Diversos estudos apontam que estes elementos repetitivos podem ser responsáveis pela origem de regiões geradoras de RNAs não codificantes, como por exemplo, alguns *loci* de precursores de microRNAs (pre-miRNAs), agentes fundamentais no mecanismo de regulação pós-transcricional. A formação de *loci* de microRNAs (miRNAs) a partir de ETs já foi descrita em plantas e animais, sobretudo em espécies modelo. No entanto não existe uma plataforma *web* de fácil utilização que disponibilize estes dados para plantas. Neste estudo, identificamos 152 pre-miRNAs que apresentam sobreposição com ETs em 10 espécies vegetais. Um repositório, denominado PlanTE-MIR DB, foi concebido para reunir esses dados e disponibilizá-los para a comunidade científica interessada no estudo das vias de regulação, origem e evolução de miRNAs. Os usuários podem explorar o banco de dados por meio de um *website* e obter a anotação de interesse utilizando um dos critérios de busca disponíveis: nome da sequência de referência do ET ou miRNA, posição no genoma, ou mesmo por um filtro hierárquico com a classificação do ET. Além disso, esta base de dados apresenta referência cruzada com os principais repositórios de ETs e miRNAs (*Repbase Update* e *miRBase*, respectivamente), a fim de promover a integração dos dados e facilitar o acesso pelo usuário que precisar de mais informações. Com o intuito de favorecer uma boa compreensão e usabilidade dos dados, estes estão publicamente disponíveis para *download* em mais de um formato de arquivo. O banco de dados pode ser acessado por meio do seguinte endereço: <http://bioinfo-tool.cp.utfpr.edu.br/plantemirdb/>.

Palavras-chave: Elementos transponíveis, microRNAs, genomas vegetais, banco de dados

Apoio financeiro: CAPES e Fundação Araucária (Edital 02/2014 - PIBIC).

Uma ferramenta versátil para análise de dados de RNA-Seq

Juliana Costa Silva¹, Douglas Silva Domingues², Luiz Filipe Protasio Pereira³,
Mariangela Hungria da Cunha⁴ and Fabrício Martins Lopes¹

¹ Programa de pós-graduação em Bioinformática (PPGBIOINFO), Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Brasil

² Departamento de Botânica, Instituto de Biociências, Universidade Estadual Paulista, UNESP, Rio Claro, SP, Brasil

³ Embrapa Café/IAPAR, Laboratório de Biotecnologia Vegetal, Instituto Agrônomo do Paraná, Londrina, PR, Brasil

⁴ Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Londrina, Brasil

Aplicação de tecnologias de sequenciamento de nova geração no sequenciamento de cDNA (RNA-Seq), em estudos com transcritos está se tornando amplamente acessível. RNA-Seq é aplicável a descoberta de transcritos, descrição de mecanismos de regulação de genes e análises de expressão diferencial. Comparado com *microarray*, RNA-Seq possui algumas vantagens: uma das vantagens mais importantes, é o fato de não ser obrigatório possuir o genoma de referência, ou qualquer outro conhecimento a prévio, permitindo resultados confiáveis até em organismos que não são modelo.

Um grande número de métodos foi desenvolvido para análises de expressão diferencial em RNA-Seq. Entretanto, não existe um consenso sobre a melhor abordagem se baseando no *design* do estudo, e a melhor escolha de *software* depende do objetivo experimental. Abordagens típicas utilizam Poisson ou distribuição negativa binomial para o modelo de perfis de expressão e vários processos de normalização.

Este trabalho propõe o desenvolvimento de um *software* interativo de análises de dados de RNA-Seq. Este *software* criará um fluxo pra análises de expressão diferencial garantindo a ordem de execução e o padrão do processo.

Além disso, será possível o usuário escolher dentre diferentes técnicas de análises para cada etapa do processo, resultando em uma caixa de ferramentas versátil para análises de dados de RNA-Seq. Será utilizada a tecnologia Java para o desenvolvimento, para possibilitar a utilização em diferentes sistemas operacionais. A

análise de dados proposta para essa abordagem é composta por três passos: i) Qualidade: exibição de informações sobre a qualidade dos reads e sobre a qualidade de mapeamento contra o genoma de referência (se houver), utilizando Bowtie2 ou BWA. Sequências com baixa qualidade de mapeamento poderão ser removidas utilizando o software PicardTools. ii) Mapeamento e quantificação: Reconstrução de transcritos poderá ser executada com Cufflinks ou Trinity (*de novo*); no caso de análises de quantificação baseadas em referência BaySeq, Cutdiff, edgeR ou HTSeq poderá ser utilizado; iii) Enriquecimento: disponibilizará anotação em bancos dados públicos, como Uniprot.

Nós esperamos produzir um software que facilite estudos comparativos de transcritos e ajude a simplificar o uso em áreas de bioinformática de análises de RNA-Seq.

Keywords: Bioinformática, análise de expressão, expressão diferencial, Transcriptoma, Reconhecimento de padrões;

Apoio financeiro: CAPES.

Reconhecimento de padrões e mineração de dados: uma nova abordagem para anotação da seqüência genômica do fungo *Phakopsora pachyrhizi*

Cynara Leão GARCIA¹, Carlos Nascimento Silla JUNIOR¹,

Francismar Corrêa MARCELINO-GUIMARAES²

¹ Programa de Pós-graduação em Bioinformática (PPGBIOINFO), Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Brasil

² Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Londrina, Brasil

A ferrugem da soja causada pelo fungo biotrófico *Phakopsora pachyrhizi* é o estresse biótico mais importante em campos de soja que levam a 80% de perdas, desde o ano de 2001 no Brasil. Estudos permitiram o monitoramento simultâneo da expressão do gene na interação planta-patógeno, ampliando o conhecimento sobre mecanismos moleculares subjacentes das respostas compatíveis e incompatíveis da soja à *P. pachyrhizi*. Em um trabalho anterior, a microdissecção e captura por laser (LCM) foi usado para isolar as células do mesófilo foliar dos locais de infecção da ferrugem e realizar processos específicos no local de acesso aos reguladores tolerantes (interação compatível) (BRS231) e resistentes (interação incompatível) (PI561356) do genótipo da soja. O RNA foi extraído a partir de células isoladas, amplificadas, e sequenciadas com a plataforma Solexa. As seqüências geradas *paired-end* (54 pb) foram mapeadas com o genoma da soja e com modelos de genes para identificar genes expressos e variantes de *splicing*. Um total de 28,572 e 30,743 genes (RPKM > 3) foram identificados por BRS231 e PI561356, respectivamente. Os *reads* restantes foram utilizados para executar uma montagem De Novo dos transcritos *P. pachyrhizi* expressos em 10 dpi na planta, uma vez que o genoma do fungo não está disponível. Para melhorar a qualidade da montagem *P. pachyrhizi*, ESTs Sanger com *reads* sequenciados e disponíveis no NCBI, foram trimados e reconstruídos *contigs* e *singlets*. As duas montagens foram combinadas para formar PPGC1.0 resultando em 36,350 *sequencia unica P. pachyrhizi* (unisequencia), expressos em 10 dpi na planta. Ao combinar LCM com um sequenciamento de alto desempenho (RNA-seq) fomos capazes de acessar os potenciais genes expressos do patógeno durante a infecção do hospedeiro e previu-se um *secretome* potencial enriquecido de genes essenciais para a infecção e patogenicidade. No entanto, um grande número de

agentes do patógeno, contigs resultantes da montagem *ab initio* não apresentaram semelhanças nas bases de dados BLAST. Considerando a importância do agente patógeno da ferrugem da soja, é importante desenvolver uma ferramenta computacional baseado em reconhecimento de padrões e mineração de dados, para realizar o mapeamento das seções (*no hits*), que não foram identificadas.

Palavras-chaves: *Phakopsora pachyrhizi*, soja, reconhecimento de padrões, mineração de dados, bioinformática, alinhamento de sequência.

StreptoRNAdb: Sistema e Base de Dados integrado de ncRNAs de Streptococcus

Autores: Tatianne da Costa Negri¹, Ivan Rodrigo Wolf², Rogério Fernandes de Souza³, Laurival Antonio Vilas-Boas², Alexandre Rossi Paschoal¹

¹Universidade Tecnológica Federal do Paraná

² Universidade Estadual de Londrina

Resumo:

A bactéria *Streptococcus agalactiae* Grupo B (GBS), é responsável por provocar grandes perdas econômicas na aquicultura e produção de leite, entre outros e ainda infecções bacterianas graves em seres humanos. Essa bactéria têm sido isolada de diferentes fontes como animais domésticos, seres humanos, cavalos, macacos, gado e peixes, mostrando a diversidade de hospedeiros. Apesar de pesquisas mostrarem que existem RNAs não-codificantes (ncRNAs) que atuam como reguladores de expressão de genes, apenas alguns candidatos têm sido descritos para importante patógeno. Para preencher esta lacuna, propõem-se o StreptoRNAdb - um sistema e banco de dados que reúne informações de ncRNAs obtidos por análises em bioinformática por nosso grupo de pesquisa.

Abordagem híbrida com ferramentas de bioinformática foram utilizadas para a identificação e caracterização de ncRNAs em nove genomas GBS de diferentes fontes de isolamento: humano, leite e peixe. Estas ferramentas são: INFERNAL, nocoRNAC, sRNAscanner, pesquisa de similaridade contra bases de dados público de sequências de ncRNA e as descritas no NRDR (<http://www.ncrnadatabases.org>).

Os resultados anteriores mostraram uma média de 56 candidatos por genoma sendo que 38 são encontrados em regiões conservadas de todas as linhagens e 2 em regiões únicas de cada genoma. A análise revelou ainda que genes de virulência estão entre alvos desses candidatos.

Os ncRNAs são resultados de nove genomas GBS (NEM316, 2603V/R, A909, 09mas018883, ILRI005, ILRI112, SA20-06, 2-22, GD201008-001), serão integrados em um único e fácil sistema de banco de dados.

O banco de dados será implementadas em MySQL. O sistema está agora na etapa de desenvolvimento em linguagem PHP e na importação dos dados para o SGBD.

O StreptoRNAdb terá uma interface amigável para o usuário, com informações sobre ncRNAs, genes-alvo e expressão de dados públicos de RNA-Seq. O site permite que o usuário pesquise de forma simples ou avançada, pela localização genômica e contém também uma visualização gráfica baseada em circos.

Todo sistema de banco de dados é público, oferecendo aos pesquisadores dados seguros, curados e organizado, permitindo que a comunidade científica explore os RNAs não-codificantes localizados na pesquisa realizada pelo nosso grupo.

CRIAÇÃO E ESTRUTURAÇÃO DE UMA BASE DE DADOS DE FAMÍLIAS DE SEQUÊNCIAS DE GENES CRY DE BACILLUS THURINGIENSIS E A IMPLEMENTAÇÃO DE FERRAMENTA DE IDENTIFICAÇÃO

Erinaldo Sanches Nascimento, Laurival Antonio Vilas-Boas, Katia Brumatti Gonçalves, Gislayne Trindade Vilas-Bôas, Alessandro Botelho Bovo

O *Bacillus thuringiensis* (Bt) é uma bactéria formadora de esporos que produz a toxina Cry como cristais de paraesporais, que tem se mostrado efetiva no controle de pragas agrícolas, no controle de mosquitos transmissores de doenças e no desenvolvimento de plantas transgênicas resistentes a insetos. O bioinseticida Bt tem resultado em uma significativa redução no uso de inseticidas químicos, fazendo um controle biológico de insetos, sem representar risco ou dano à saúde humana. No entanto, existem muitas pragas de insetos não susceptíveis às toxinas Cry identificadas até agora. Já existe a identificação de mais de 700 sequências de gene cry classificadas em pelo menos 70 grupos. Uma alternativa às pragas que não se mostram susceptíveis às toxinas Cry parentais é o isolamento de outras estirpes de Bt com novos genes cry com toxicidade melhorada, além da identificação das moléculas receptoras e epitopos de ligação, a triagem de novas proteínas Cry com novos insetos e receptores específicos. Outra alternativa é a evolução genética in vitro de tais toxinas. No apoio a atividade de realizar uma análise filogenética e a classificação dos grupos de gene cry, bem como a otimização para predição de alvos susceptíveis às toxinas Cry, esse estudo propõe a criação e a estruturação de uma base de dados acurada de genes cry de Bt e a implementação de uma ferramenta capaz de realizar a classificação e a identificação da família de gene cry de uma sequência de nucleotídeos isolada ou um conjunto de contigs. O processo de análise e classificação filogenética resultará na utilização de métodos de mineração de dados e reconhecimento de padrões nos dados biológicos. A ferramenta computacional consiste em uma estratégia de descoberta de prováveis candidatos a ser um gene cry com base em score e um método de caracterização que faz a análise filogenética por similaridade. Apresentando como saída o alinhamento das sequências e a árvore filogenética desse gene cry. A tarefa desafiadora nesse projeto é amparar o biólogo, usando uma interface web e uma base de dados online, na identificação e classificação de proteínas Cry, com base na sequência de DNA, visando a descrição de produtos biotecnológicos com toxicidade melhorada que possam atuar no controle de pragas de insetos e vetores de doenças, num espectro de ação mais amplo alcançando aquelas que atualmente não são susceptíveis às toxinas Cry ou que são controlados ineficientemente.

Middleware para Análise in Silico de miRNAs

Rosana R. Aguiar¹, Leandro A. Ambrosio¹, Gonzalo Sepúlveda-Hermosilla²,
Vinicius Maracaja-Coutinho^{2,3,4}, Alexandre Rossi Paschoal¹ *

* paschoal@utfpr.edu.br

¹Universidade Tecnológica Federal do Paraná, Brasil.

²Centro de Genômica e Bioinformática, Universidad Mayor, Santiago, Chile.

³Beagle Bioinformatics, Santiago, Chile.

⁴Instituto Vandique, João Pessoa, Brasil.

miRNAs são pequenos RNAs, contendo cerca de 22 nucleotídeos (maduro) e que atuam como inibidores/silenciadores pós-transcricionais dos níveis de mRNA. Este interesse se deve pelo papel de controle e regulação das redes gênicas dentro das células, estando assim envolvidas em diversos processos biológicos e doenças como câncer ou em planta, por exemplo. O miRNA ou microRNA é uma classe de ncRNA que desperta maior interesse de pesquisa pela comunidade científica. E apesar de existir uma grande variedade de abordagens computacionais descritas para a identificação de microRNAs, em sua grande maioria, estas apresentam algum tipo de limitação (e.g. desatualizadas, não mais disponíveis). De modo a atender a uma demanda, este trabalho apresenta o miRQuest, um middleware, em uma plataforma web para a investigação miRNA que integra ferramentas de identificação de miRNA com duas funções principais: (i) Benchmark e (ii) Identificação dos miRNAs. Este sistema foi desenvolvido em Java com especificação J2EE, XML, Eclipse, Tomcat versão 8 (WebServer), os frameworks *PrimeFaces* e *Apache Shiro*, e a API *Commons E-mail*. Na verdade, o miRQuest não introduz um novo modelo computacional para predição de miRNAs, mas sim, uma nova metodologia que permite executar simultaneamente diferentes técnicas de identificação de miRNA. Para avaliar o sistema foram usados conjunto de dados extraídos do banco dados miRBase versão 19: **Animais, Plantas, Vírus e Protista**. Além desse grupo de sequencias, para demonstrar o uso da funcionalidade de identificação de miRNAs, foi utilizado o conjunto de dados do mosquito *Anopheles Gambiae* com o intuito foi de identificar miRNAs derivados de longos RNAs não-codificantes e sequenciamento "Illumina small RNA de células de neuroblastoma humano. Finalmente, a fim de otimizar os recursos de hardware e permitir o processamento paralelo, o sistema usa filas com aplicação do conceito FIFO (first-in first-out).

Agradecimentos

ARP agradece pelo suporte do projeto ao CNPq - Edital MCTI/CNPQ/Universal 14/2014 - Faixa A - até R\$ 30.000,00 Processo: 454505/2014-0. Esta pesquisa é parte do resultado de mestrado da aluna RRA do Programa de Pós-Graduação em Informática (Profissional) UTFPR Cornélio Procópio, PR, Brasil.