

COTEGRAMLA (CONVERSOR DE TEXTO PARA GRAPH MARKUP LANGUAGE)

**MARIN, Gabriel Cabral; WERHLI, Adriano V.
MACHADO, Karina
gabrielcmsistemas@outlook.com**

**Evento: 14ª Mostra da Produção Universitária
Área do conhecimento: Ciências Exatas e da Terra**

Palavras-chave: Bioinformática, Proteínas

1 INTRODUÇÃO

Cada vez mais, a computação tem sido utilizada para acelerar procedimentos que utilizam proteínas. Os testes *in vitro* são demorados e custosos, enquanto os testes *in silico* produzem resultados mais rápidos e que demandam menos recursos. Já existem programas para tais testes, no entanto, são poucas as opções quando o foco é a visibilidade gráfica das interações/relações de similaridade entre proteínas. Este trabalho tem por objetivo desenvolver um programa para a conversão de arquivos no formato TXT genéricos gerados por programas que calculam relações de similaridade entre proteínas, para arquivos no formato GML, que produzem uma visualização mais rica e customizável, compatível com o programa de interação e análise de redes Cytoscape (Shannon et al.; 2003).

2 REFERENCIAL TEÓRICO

O Protein Data Bank (Berman et al., 2000) é um banco de dados biológico de estrutura tridimensional de proteínas. Atualmente esse banco de dados tem mais de 100.000 estruturas depositadas. A similaridade entre as estruturas de 2 ou mais proteínas é um tema de pesquisa com diferentes abordagens. Uma alternativa é relacionar a similaridade por meio das sequências. Dessa forma, é necessária uma forma de visualizar a relação entre as proteínas de acordo com valores de similaridade que são obtidos em outras ferramentas como por exemplo a partir do resultado de um alinhamento de sequências.

3 MATERIAIS E MÉTODOS (ou PROCEDIMENTO METODOLÓGICO)

O programa Cytoscape foi escolhido para ser utilizado como principal ferramenta de visualização gráfica. Foi utilizado um arquivo TXT fornecido pelo próprio site da ferramenta. O conversor foi criado com a linguagem PHP, por ser leve e de fácil acesso.

Em um primeiro momento, a ferramenta de conversão lê o arquivo TXT e o transforma em um vetor de linhas. Após, cada linha é quebrada novamente, para que os nodos sejam arquivados em um vetor e as interações em outro. Este arquivamento dos nodos é necessário, já que o formato GML não permite a repetição dos mesmos. Em loop, cada nodo é colocado no vetor e então tratado para se adequar ao novo formato requerido pelo arquivo GML. Cada linha convertida é escrita diretamente no novo arquivo. O mesmo procedimento é feito com as ligações

entre os nodos. Quando o conversor não encontra mais nenhuma linha a ser tratada, o arquivo é gravado na máquina local.

4 RESULTADOS e DISCUSSÃO

O arquivo TXT (Figura 1) é tratado pelo conversor. Quando o arquivo GML é gerado com sucesso (Figura 2), o mesmo pode ser importado para o programa Cytoscape para visualização no formato de rede. Se a rede em questão for muito grande e/ou tiver muitas interações, é possível aproximar-se de um pedaço da mesma, possibilitando melhor visão das ligações, bem como os nomes das proteínas/nodos.

Figura 1 – Exemplo de TXT genérico

```
4R79.1a F44D12.4      0.394
AC3.10 AC3.4      0.537
AC3.3  AC3.4      0.537
AC3.3  C03A7.14     0.537
AC3.3  C03A7.4     0.999
AC3.3  C05D2.1a.1  0.537
AC3.3  C37C3.6b.1  0.537
AC3.3  F29G6.3b.1  0.538
AC3.3  H04J21.3b   0.998
```

Figura 2 – Exemplo de GML gerado

```
graph [
  node [ id 5
    label "AC3.3"
  ]
  node [ id 6
    label "C03A7.4"]
  edge [ label "0.999"
    source 5
    target 6
    graphics [ type "line"
      fill "#00FF00"
    ]
  ]
]
```

5 CONSIDERAÇÕES FINAIS

O trabalho desenvolvido facilita a utilização da ferramenta Cytoscape e todos seus recursos para análise de relações de similaridade entre proteínas obtidas por meio de outras ferramentas. Como trabalho futuro pretende-se desenvolver uma melhor interface para o programa de conversão proposto.

REFERÊNCIAS

- Berman, H.M. et al. (2000) The Protein Data Bank. *Nucleic Acids Research*, 28: 235-242.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, chwikowski B, Ideker T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* Nov; 13(11):2498-504