



Uma Metodologia para Identificação de Módulos Formadores de Sequências de Proteínas Mosaicas do *Trypanosoma cruzi* a partir do Transcriptoma do Parasito Utilizando a Ferramenta BLAST

Elisa Boari de Lima – elisa@comp.ufla.br

Orientador: Thiago de Souza Rodrigues – thiago@dcc.ufla.br

Novembro/2008

Roteiro

- **Introdução**
- **Expressão Genômica**
- **Proteínas Mosaicas**
- *Trypanosoma cruzi*
- **Alinhamento de Sequências**
- **BLAST**
 - Parâmetros do BLAST
 - Casamentos Curtos
- **Materiais e Métodos**
- **Resultados**
- **Conclusão**

Introdução

- *T. cruzi* → Doença de Chagas
- Sequenciamento do Genoma do *T. cruzi*
 - Dados estruturais como dados sobre módulos de proteínas mosaicas
- Módulos
 - Diferentes funções enzimáticas, sinalizadoras, regulatórias ou estruturais
 - Identificação abre caminho para novas estratégias terapêuticas

Roteiro

- Introdução
- **Expressão Genômica**
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- Materiais e Métodos
- Resultados
- Conclusão

Expressão Genômica

- Genoma → depósito de informação biológica
- Dogma Central da Biologia Molecular



Expressão Genômica

- RNA – cadeia de nucleotídeos.
- Nucleotídeos – base nitrogenada (A, T, G, C) + açúcar + fosfato

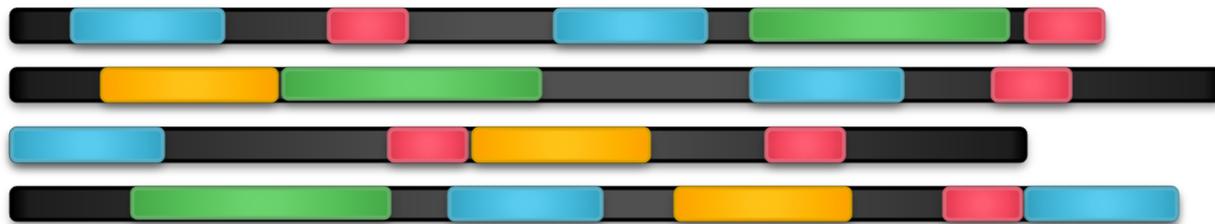
...GATGATGATGACTGGCCGTGTGC...

Roteiro

- Introdução
- Expressão Genômica
- **Proteínas Mosaicas**
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- Materiais e Métodos
- Resultados
- Conclusão

Proteínas Mosaicas

- Justaposição de módulos
- Papel essencial nas reações químicas da biologia extracelular
- Papel importante na multicelularidade



Roteiro

- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- Materiais e Métodos
- Resultados
- Conclusão

Trypanosoma cruzi

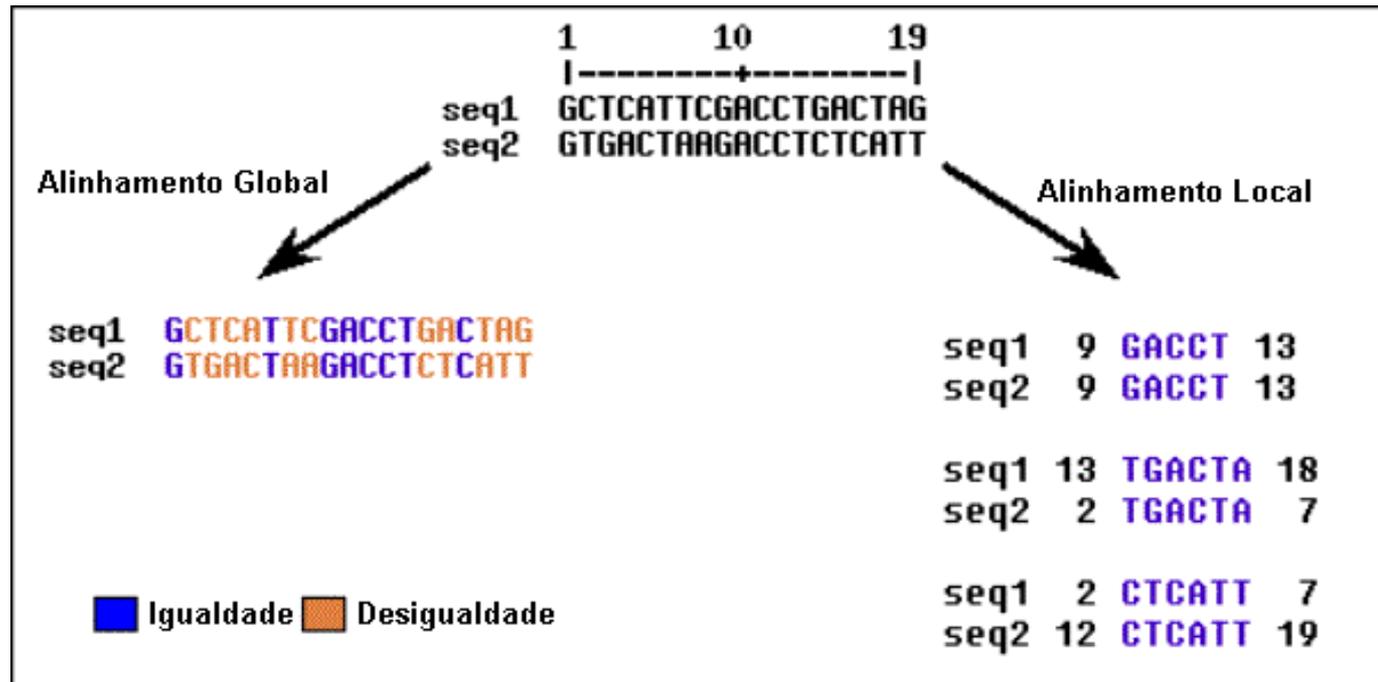
- 50% do genoma → sequências repetitivas principalmente de proteínas de superfície
- Família MASP
 - Proteínas de Superfície Associadas a Mucinas
 - Regiões centrais altamente variáveis
 - Sequências repetitivas

Roteiro

- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- **Alinhamento de Sequências**
- **BLAST**
 - Parâmetros do BLAST
 - Casamentos Curtos
- Materiais e Métodos
- Resultados
- Conclusão

Alinhamento de Sequências

- Comparar sequências buscando caracteres na mesma ordem
- Objetivo: encontrar sequências similares



Roteiro

- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- **BLAST**
 - Parâmetros do BLAST
 - Casamentos Curtos
- Materiais e Métodos
- Resultados
- Conclusão

BLAST – *Basic Local Alignment Search Tool*

- Método heurístico para alinhamento de uma sequência contra um banco de dados
- Conjunto de programas
 - BLASTP
 - BLASTX
 - TBLASTN
 - TBLASTX
 - **BLASTN**
- Gera relatório com informações de cada alinhamento
 - Pontuação
 - *E-value*
 - Identidade
 - Posições Alinhadas

BLAST – Parâmetros

- *E-value*
 - Probabilidade de alinhamento ser aleatório
- Tamanho inicial de palavra
 - Sensibilidade das buscas do BLAST
- Filtro de regiões de baixa complexidade
 - Elimina regiões biologicamente desinteressantes
- Sistema de pontuação para nucleotídeos
 - Recompensa para igualdade, penalização para desigualdade

BLAST – Casamentos Curtos

- Sequências menores que 20 nucleotídeos
- Dificuldades Encontradas
 - *E-value* padrão muito baixo
 - Sequências curtas têm *e-value* alto
 - Tamanho de palavra muito alto
 - Filtro de complexidade remove boa parte da sequência

Roteiro

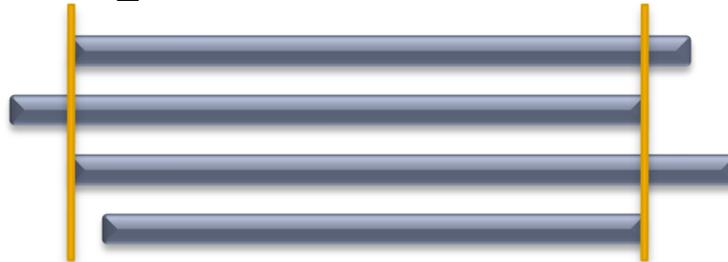
- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- **Materiais e Métodos**
- Resultados
- Conclusão

Materiais e Métodos

- 810 sequências de nucleotídeos da MASP
- Ferramenta BLAST versão 2.2.17
 - BLASTN – alinhamentos
 - *formatdb* – banco de dados de sequências

Materiais e Métodos

- Alinhamento das sequências originais (S_1)
- Filtragem dos resultados
- Regiões comuns às sequências
 - Estratégia de Corte → Extração para novo conjunto S_2



- Transitividade nos resultados
 - Redundância em S_2 → Alinhamento $S_2 \times S_2$

Materiais e Métodos

- Filtragem dos novos resultados
- Separação em grupos de sequências semelhantes
- Mesma sequência em mais de um grupo
 - Estratégia de corte nos grupos gerando S_3

Materiais e Métodos

- Necessidade de processo iterativo
- Condição de Parada
 - Todos os grupos unitários
 - Grupo unitário \rightarrow sequência é possível módulo
- Definição de módulos a partir de módulos candidatos

Roteiro

- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- Material
- Métodos
- **Resultados**
- Conclusão

Resultados – Metodologia

- $i = 1$
- **Loop:**
 - Alinhamento S_i contra S_i
 - Filtragem dos resultados
 - **Se $i > 1$**
 - Separação em grupos
 - **Se** todos os grupos forem unitários
 - Interromper o loop
 - **Senão**
 - Estratégia de corte nos grupos gerando S_{i+1}
 - **Senão**
 - Estratégia de corte nos alinhamentos filtrados gerando S_{i+1}
 - $i = i + 1$
- Fim do Loop
- Alinhamento S_i contra S_1
- Definição dos módulos

Resultados – Aplicação na MASP

- Diferentes conjuntos de valores de parâmetros
- Mapeamento de módulos nas sequências
- Para comparação: número de módulos e frequências de ocorrências

Resultados – Aplicação na MASP

- Filtro de Regiões de Baixa Complexidade
 - Ativado (Padrão) x Desativado

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
Ativado	496	29	14
Desativado	527	28	16

Resultados – Aplicação na MASP

- *E-value*
 - 10 (Padrão) x 100

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
10	527	28	16
100	1300	17	24

Resultados – Aplicação na MASP

- *E-value*
 - 100 x 500 x 1000 (Indicado Sequências Curtas)

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
100	1300	17	24
500	1045	24	28
1000	1043	24	28

Resultados – Aplicação na MASP

- Tamanho Inicial de Palavra
 - 11 (Padrão) x 10 x 9 x 8

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
11	1045	24	28
10	530	21	12
9	169	25	4
8	24	32	0

Resultados – Aplicação na MASP

- Tamanho Inicial de Palavra
 - 11 (Padrão) x 12 x 13

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
11	1045	24	28
12	2464	16	42
13	1235	17	23

Resultados – Aplicação na MASP

- Sistema de Pontuação de Nucleotídeos
 - 1/-3 (Padrão), 1/-2, 1/-1

Valor	Total de Módulos	Média de Ocorrências	Média de Módulos por Sequência
1/-3	2464	16	42
1/-2	2364	17	42
1/-1	1368	21	30

Resultados – Aplicação na MASP

- Visualização parcial da sequência com mais módulos (183)

ATGGCGATGATGATGAGTGGCCGTGTGCTGCTGGTGTGTGCCCTCTGCGTGC
TGTGGTCCGTTGCGGCCGATGGAGATGTTGTTGTITCTGGTGGGGAAGACAA
CAGTCTGAAAGAATTATTATTCCAGTTGCGAGATTGCAGGAAAGACAAGA
ACAAAGAGCAGTAGAAGCAACAGCTGATGCAAAAGGCAGCAGCAGAAGCAG
CAGAAACAGCAACAGCAAAAGCAGAGGAAGCAGAGGCAGCAGCAACAGA
AGCAAAGGCGGCTGCAAGAGACAGCAGCAGAAGCAGCAAAGGCAGCAGCAG
AGGCAGCAGCAACGGCAGCAGAAGCGGCAGCAGCAGAAGCAAAAACAGCA
GCCACAGCAGCAAAGGCAGTAGACACCGAGGCAAAGCAAAGCAGCAGC
AGCAGCAGCTGAATCAGCAGCAACAAAAGCAACAACAGCATCAGAAGCAG
CAACAAAAGCAAAGCAAACAGCATCAGCAGCAAAAGGCAGCGACAGAGGCA
GCAGCAGCAAAGGCAGCAGCAGCAGCAGCAGCAGCAAAGCAGAAGAAGCAG
AAGCAGAAAGCAGCAGCAGAAGCAGCAAAGGCAGCGGCAAAGGCGGCAGCC

Roteiro

- Introdução
- Expressão Genômica
- Proteínas Mosaicas
- *Trypanosoma cruzi*
- Alinhamento de Sequências
- BLAST
 - Parâmetros do BLAST
 - Casamentos Curtos
- Metodologia
 - Estratégias Utilizadas
- Resultados
- **Conclusão**

Conclusão

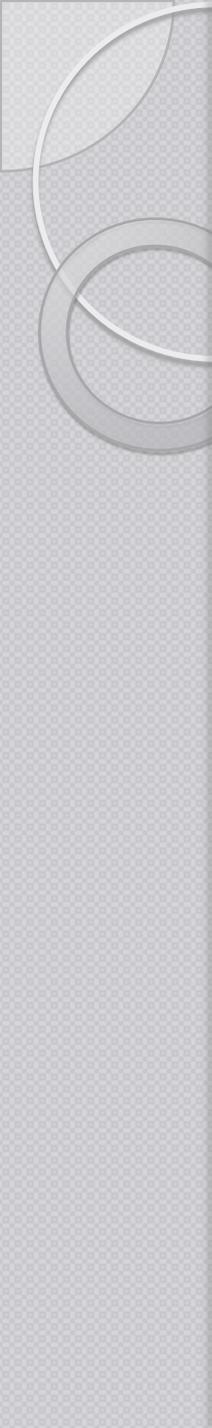
- Metodologia desenvolvida é eficaz para detecção de módulos de proteínas
- Configuração com melhores resultados:
 - Filtro de baixa complexidade desativado
 - *E-value* alto
 - Tamanho de palavra maior
 - Sistema de pontuação padrão
- Confirmação da estrutura mosaica da família MASP do *T. cruzi*
- Aceitação no Rocky 2008

Trabalhos Futuros

- Comparação com resultados baseados no proteoma
- Tratamento das sobreposições
 - Ocorrência condicional de módulos
 - Refinamento dos resultados obtidos



Dúvidas?



Obrigada!