

ALINHAMENTO DE SEQÜÊNCIAS EST-SSR DE GÉRBERA, UTILIZANDO BLASTX

BENEMANN, Daiane de Pinho^{1,2}, ARGE, Luis Willian Pacheco², NOGUEIRA, Luciana Rodrigues², MAIA, Luciano Carlos da³; PETERS, José Antonio².

¹*Doutoranda do Programa de Pós-Graduação em Biotecnologia, Universidade Federal de Pelotas (daiane_bio@yahoo.com.br)*

²*Laboratório de Cultura de Tecido de Plantas, Departamento de Botânica, Instituto de Biologia, Universidade Federal de Pelotas*

³*Deptº de Fitotecnia – Centro de Genômica e Fitomelhoramento - FAEM/UFPel*

1 INTRODUÇÃO

A Biologia Molecular é uma ciência que apresentou avanços muito significativos nas últimas décadas. Os pesquisadores frequentemente trabalham com uma grande quantidade de informações geradas a partir de experimentos em laboratório. Dada à necessidade de manipular essa informação, surgiu a bioinformática, que aplica técnicas computacionais, matemáticas e estatísticas para tratar aspectos referentes à biologia molecular. Um dos problemas principais dentro da biologia molecular é a comparação de seqüências. A técnica mais frequentemente usada para comparar duas ou mais seqüências de nucleotídeos ou aminoácidos consiste em sobrepor uma cadeia sobre outra e buscar semelhanças e diferenças, a qual é formalmente denominada alinhamento (Ticona 2003). O objetivo é realizar buscas por similaridades, que quando encontradas, indicam uma alta probabilidade de se encontrar funções análogas entres as seqüências. Com o grande volume de dados oriundos dos projetos de sequenciamento, ferramentas de alinhamento se tornaram muito úteis. As ferramentas hoje disponíveis utilizam duas formas de alinhamento: global e local. Na primeira, são comparadas seqüências totais, isto é, verificando se há similaridade entre seqüências inteiras. No alinhamento local, há uma busca por trechos de seqüência que sejam parecidos (Korf et al., 2003).

A ferramenta BLAST (Basic Local Alignment Search Tool) utiliza heurísticas e algoritmos de programação dinâmica para obter os melhores alinhamentos locais, com um tempo de execução bem reduzido em relação aos programas até então conhecidos. (Altschul et al., 1997).

O presente trabalho tem como objetivo o alinhamento de seqüências não redundantes de EST-SSR de Gérbera, presentes no NCBI, a fim de identificar as possíveis funções e utilização destas seqüências em trabalhos futuros.

2 MATERIAL E MÉTODOS

Todos os ESTs relacionadas com Gerbera listados no NCBI foram baixados em formato FASTA. Para evitar a redundância, foi realizada uma análise em cluster com o software Cap3 (Huang e Madan, 1999). Para identificação e localização dos microssatélites (SSRs) em ESTs foi utilizado o software *SSRLocator* (Maia *et al.*, 2008). Após identificar seqüências ESTs-SSR, estas foram alinhadas (BLASTx) contra o banco de dados nr. As seqüências foram consideradas homólogas às proteínas conhecidas no banco de dados quando o E-Value do

BLAST foi menor do que $1e^{-10}$ e a identidade superior a 80% e consideradas análogas quando E-value foi maior que $1e^{-10}$.

3 RESULTADOS E DISCUSSÃO

Foi encontrado 15.851 sequencias ESTs de Gérbera no banco de dados do NCBI, correspondente a 7.254.970 pb, com um tamanho médio dos transcritos de 457 pb. Com a eliminação da redundância, obteve-se 7.529 contigs, correspondentes a 3.941.688 pb e tamanho médio dos transcritos de 533 pb.

Através do software *SSRLocator*, foram encontradas 225 seqüências que continham microsatélites e estas foram utilizadas para alinhamento no banco de dados do NCBI. Destas 225 seqüências, somente 34 (15,11%) obtiveram E-Values inferior a $1e^{-10}$ e identidade superior a 80%, sendo consideradas seqüências homólogas as seqüências de Gérbera. A maioria das seqüências identificadas no presente estudo teve similaridade significativa com genes de outras espécies de plantas, como *Populus trichocarpa*, *Vitis vinifera* e *Ricinus comunis* (Figura 1). Moccia et al., 2009, trabalhando com *Silene latifolia*, encontrou resultados semelhantes ao nossos e, acredita que essa similaridade é conseqüência das informações já disponíveis destas espécies no banco de dados, visto que seus genomas foram sequenciados e não reflete a proximidade genética. Porém, estas espécies que apresentaram semelhanças com Gérbera são pertencentes às angiospermas, ao clado das eudicotiledôneas e subespécie Rosids, enquanto a Gerbera pertence à subespécie Asterids que se encontra na arvore filogenética ao lado de Rosids, no mesmo clado, por isso não podemos afirmar que essas espécies não apresentam semelhança genética. Assim, o correto seria testar a transferabilidade destes *primers* nas espécies semelhantes à Gérbera.

Com relação as possíveis funções das seqüências EST-SSR de Gérbera, os contigs g1333, g179, g231 e g219, obtiveram identidade superior a 95% e E-values inferiores a $1e^{-38}$, relacionados a espécies como, *Malus X domestica* (função: C3HL domais class transcripction factor), *Capsicum annum* (função: Rho), *Gerbera hybrida* (função: Aquaporina) e *Ricinus communis* (função: Serine-pyruvate aminotransferase), respectivamente. Os contigs g1176 e gi|58327929 não apresentaram função conhecida, porém foram semelhantes a *Populus trichocarpa* e *Glycine max*.

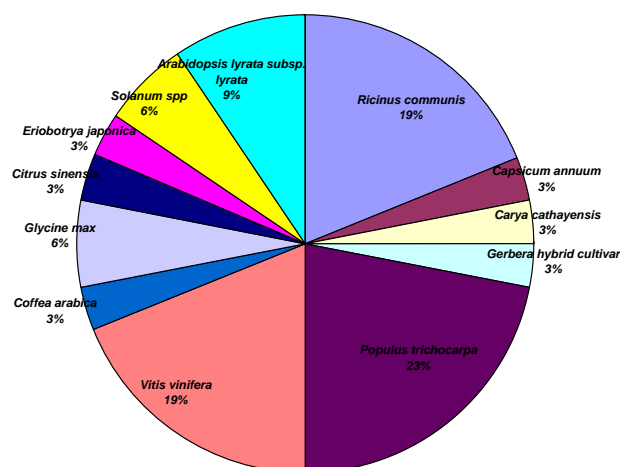


Figura 1- Espécies relacionadas as seqüências de Gérbera.

Tabela 1 – Lista das ESTs-SSR, E-Value, porcentagem de identidade, possíveis funções e espécies encontradas no alinhamento no banco de dados do NCBI, utilizando BLASTx.

Contig	Blast (E-Value)	% de identidade	Função	Espécie
g20	7 e ⁻⁹⁵	80	<i>ATP-dependent peptidase</i>	<i>Ricinus communis</i>
g179	2 e ⁻⁸⁸	96	<i>Rho</i>	<i>Capsicum annuum</i>
g219	2 e ⁻⁶⁸	95	<i>Serine-pyruvate aminotransferase</i>	<i>Ricinus communis</i>
g228	3 e ⁻¹⁴³	92	<i>Putative chloroplast chlorophyll a/b-binding protein</i>	<i>Carya cathayensis</i>
g231	2 e ⁻¹⁴²	96	<i>Aquaporina</i>	<i>Gerbera hybrid</i>
g320	2 e ⁻¹⁴³	85	<i>Precursor of transferase serine</i>	<i>Populus trichocarpa</i>
g438	2 e ⁻¹⁴	88	<i>Hypothetical protein</i>	<i>Vitis vinifera</i>
g1033	5 e ⁻¹⁸	83	<i>Hypothetical protein</i>	<i>Populus trichocarpa</i>
g1053	9 e ⁻¹³	86	<i>Hypothetical protein</i>	<i>Vitis vinifera</i>
g1092	4 e ⁻⁷⁷	90	<i>Metionina synthase</i>	<i>Coffea arabica</i>
g1101	2 e ⁻³⁶	88	<i>Mitochondrial pyruvate dehydrogenase kinase isoform 1</i>	<i>Glycine max</i>
g1176	3 e ⁻²⁹	86	<i>Unknown</i>	<i>Populus trichocarpa</i>
g1250	8 e ⁻⁸⁴	82	<i>PSI reaction center subunit II</i>	<i>Citrus sinensis</i>
g1257	9 e ⁻⁷⁷	84	<i>Hypothetical protein LOC100259220</i>	<i>Vitis vinifera</i>
g1293	8 e ⁻¹⁷⁷	83	<i>Unnamed protein product</i>	<i>Vitis vinifera</i>
g1333	2 e ⁻³⁸	100	<i>Histone H4</i>	<i>Eriobotrya japonica</i>
g1677	5 e ⁻⁷⁴	85	<i>Acyltransferase</i>	<i>Ricinus communis</i>
g2105	2 e ⁻⁷³	91	<i>EF hand family protein</i>	<i>Solanum demissum</i>
gi 62943581	8 e ⁻²³	80	<i>Hypothetical protein ARALYDRAFT_494297</i>	<i>Arabidopsis lyrata subsp. lyrata</i>
gi 62942576	5 e ⁻⁵⁷	87	<i>Predicted protein</i>	<i>Populus trichocarpa</i>
gi 62942099	8 e ⁻⁸⁸	88	<i>Predicted protein</i>	<i>Populus trichocarpa</i>
gi 62942045	3 e ⁻⁵⁰	92	<i>Hypothetical protein LOC100252098</i>	<i>Vitis vinifera</i>
gi 62941101	2 e ⁻⁵⁸	82	<i>Unnamed protein product</i>	<i>Vitis vinifera</i>
gi 62940000	5 e ⁻²²	86	<i>Hypothetical protein ARALYDRAFT_485347</i>	<i>Arabidopsis lyrata subsp. lyrata</i>
gi 58329162	1 e ⁻³⁸	83	<i>Predicted protein</i>	<i>Populus trichocarpa</i>
gi 58329027	5 e ⁻¹⁹	82	<i>Hypothetical protein ARALYDRAFT_911661</i>	<i>Arabidopsis lyrata subsp. lyrata</i>
gi 58328640	9 e ⁻¹³	83	<i>ATP citrate (pro-S)-lyase</i>	<i>Populus trichocarpa</i>
gi 58328112	1 e ⁻²⁹	84	<i>Regulator of ribonuclease activity A</i>	<i>Ricinus communis</i>
gi 58327929	1 e ⁻⁴¹	72	<i>Unknown</i>	<i>Glycine max</i>
gi 58325973	1 e ⁻¹⁸	83	<i>Lung seven transmembrane receptor</i>	<i>Medicago truncatula</i>
gi 58325931	3 e ⁻¹⁶	89	<i>Histone h2a</i>	<i>Ricinus communis</i>
gi 58325855	1 e ⁻⁴⁵	86	<i>Ubiquinol--cytochrome c reductase</i>	<i>Solanum tuberosum</i>
gi 58323140	1 e ⁻⁵⁶	92	<i>Derlin-2</i>	<i>Ricinus communis</i>
gi 58322153	1 e ⁻⁷⁰	90	<i>Protein kinase</i>	<i>Medicago truncatula</i>

4 CONCLUSÕES

Pode-se concluir que 32 sequencias de EST-SSR de *Gérbera* apresentam funções conhecidas, podendo assim ser utilizadas para desenho de *primers* em estudos futuros.

5 REFERÊNCIAS

ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. L. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research*, v. 25, n. 17, p. 3389-3402, 1997.

HUANG X, MADAN A. CAP3: a DNA sequence assembly program. *Genome Research* v.9, p.868–877,1999.

KORF I.; YANDELL M., BEDELL J. *An Essential Guide to the Basic Local Alignment Search Tool*; O' Reilly & Associates, Inc., Sebastopol, U.S.A. 2003.

MAIA L.C. DA; PALMIERI D.A, DE; SOUZA V.Q; KOPP M.M; DE CARVALHO F.I; COSTA DE O.A. SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genomics*, p.1-9, 2008.

MOCCIA M.D; OGER-DESFEUX C.; MARAIS G.A; WIDMER A. A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics*. N.10, p.243, 2009.

TICONA W. G. C. Aplicação de algoritmos genéticos múltiplo-objetivo para alinhamento de seqüências biológicas. 2003. Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação- ICMC-USP. USP-São Carlos, Fevereiro de 2003.