

ESTUDO ESTRUTURAL DA OCORRÊNCIA DE SSRs (*SIMPLE SEQUENCE REPEATS*) NO GENOMA DO ARROZ: CARACTERIZAÇÃO COMPLETA PÓS-SEQÜENCIAMENTO

Luciano Carlos da Maia⁽¹⁾, Darío Abel Palmieri⁽²⁾, Velci Queiroz de Souza⁽¹⁾, Ariano Martins de Magalhães Junior⁽³⁾, Fernando Irajá Félix de Carvalho⁽¹⁾, Antonio Costa de Oliveira⁽¹⁾.¹Centro de Genômica e Fitomelhoramento – FAEM/UFPEL, Campus Universitário, s/nº · Caixa Postal 354, 96010-900, Pelotas, RS; ²Laboratório de Estudos em Meio Ambiente - Universidade Católica do Salvador; ⁽³⁾Embrapa Clima Temperado - lmaia.faem@ufpel.edu.br

Microssatélites ou SSRs (*Simple Sequence Repeats*) são estruturas de DNA repetitivo que mostram os maiores índices de mutações no genoma. Segundo Morgante & Olivieri (1993) estas sequências de DNA são constituídas por nucleotídeos que se repetem lado a lado (em série) e são amplamente encontrados nos genomas de eucariotos. Para Li et al. (2004) várias evidências demonstram que a ocorrência dos microssatélites não é ao acaso o que pode prover vantagens para rápida adaptação a novos ambientes. Nos cereais (milho, trigo, cevada, sorgo e arroz) valores entre 1,5% e 7,5% das seqüências expressas (ESTs) contêm SSRs (Thiel et al. 2003). Estas regiões do genoma são de grande interesse para o desenvolvimento de marcadores moleculares em plantas, os quais são utilizados no mapeamento genético comparativo, em estudos evolutivos e, mais recentemente, no mapeamento associativo. Dessa forma, a utilização desta classe de marcadores é de grande valia para a agricultura, uma vez que tem sido descrita como a que tem maior capacidade de ligar as alterações genotípicas a diferenças do fenótipo, constituindo uma potente ferramenta auxiliar para os melhoristas.

Trabalhos previamente publicados estudaram a ocorrência dessas regiões no genoma do arroz. Entretanto, um estudo no pós-sequenciamento desse genoma pode prover informações mais conclusivas quando comparado com aqueles efetuados anteriormente. Segundo Morgante et al. (2002), a ocorrência de microssatélites corresponde a 0,85% do genoma de *Arabidopsis thaliana*, 0,37% do milho, 3,21% do peixe *Fugu rubripes*, 0,21% do *Caenorhabditis elegans* e 0,30% de *Sacharomyces cerevisiae*. Níveis de ocorrências de SSRs em plantas foram elucidados por Morgante et al. (2002) (Tabela 1).

No presente trabalho relatamos a ocorrência de todos os possíveis arranjos de SSRs compreendidos entre monômeros e decandros, a partir do banco de dados do genoma do arroz sequenciado pelo consórcio IRGSP (*International Rice Genome Sequence Project*).

Arquivos em formato *Fasta* (padrão internacional para análises de dados genômicos) contendo as seqüências completas referentes à versão 4 de cada um dos 12 cromossomos do arroz, foram obtidos a partir do *website* do IRGSP (<http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html>). A identificação dos SSRs foi realizada com auxílio de uma ferramenta computacional desenvolvida nas linguagens *Delphi* e *Perl* (Maia, 2007). O algoritmo utilizado identifica todos os SSRs perfeitos presentes nas seqüências analisadas que correspondam aos seguintes parâmetros previamente definidos: monômeros com 20 repetições, dímeros com 10, trímeros com sete, tetrâmeros com cinco, pentâmeros e hexâmeros com quatro, heptâmeros, octâmeros, e nonâmeros com três repetições, respectivamente. Foram analisadas 181.732 regiões (pseudo-molécula) do genoma do arroz, totalizando 370.733.456 pb.

Os resultados obtidos mostraram que 25.411 delas apresentaram a ocorrência de microssatélites. A somatória dos nucleotídeos encontrados nesses trechos correspondeu a 801.416 pb, mostrando que a proporção total do genoma de arroz constituído por SSRs é de 0,2161%. Na Tabela 2, é discriminada a ocorrência total de microssatélites, distribuídos

por motivos e cromossomos no genoma do arroz. Os dados obtidos evidenciam que os SSRs constituídos por di e trinucleotídeos são os mais abundantes, totalizando cada um 43,83% e 20,22% das ocorrências, respectivamente. A maior ocorrência de SSRs foi evidenciada no cromossomo 1, onde 3.122 microssatélites foram encontrados, o que corresponde a 12,3% do total encontrado, seguido pelos cromossomos 2 e 3, com 10,4% e 10,5%, respectivamente. Foi obtida também a frequência de ocorrência de microssatélites nos 10 motivos pesquisados; esses dados representam o número médio de ocorrências em intervalos de um milhão de pares de bases (Mpb) (Tabela 2). Microssatélites dinucleotídeos representaram o maior número de ocorrências por milhão de nucleotídeos, com um valor médio de 30 ocorrências. Estes dados diferem dos obtidos por Morgante et al. (2002). Assim, pode-se concluir que resultados de maior riqueza de ocorrências de dímeros e trímeros fornecem uma visão final de que essas duas classes de marcadores devem ser exploradas com mais intensidade.

Tabela 1. Frequência de ocorrência de microssatélites (microssatélites/Mpb) em DNA genômico de cinco espécies vegetais (Morgante et al., 2002).

Motivo	<i>Arabidopsis</i>	Soja	Milho	Trigo	Arroz
Monômeros	112.2	51.5	10.6	20.8	89.9
Dímeros	77.7	45.7	27.5	43.4	55.0
Trímeros	142.0	50.5	83.1	76.5	90.5
Tetrâmeros	86.0	69.1	45.7	49.8	86.0
Pentâmeros	29.7	19.5	13.1	8.1	22.3

Fonte: Morgante et al. (2002).

Tabela 2. Ocorrência e frequência (Mpb) de microssatélites, distribuídos por motivos e cromossomos no genoma do arroz. CGF/FAEM/UFPel, Pelotas, 2007.

Cr.	Mono	Di	Tri	Tetra	Penta	Outros	Total	%
1	138	1.335	643	302	358	346	3.122	12,3
2	101	1.085	571	277	303	297	2.634	10,4
3	94	1.172	618	241	275	265	2.665	10,5
4	69	863	410	177	250	225	1.994	7,8
5	68	880	429	210	271	239	2.097	8,3
6	59	929	431	206	263	260	2.148	8,5
7	64	826	389	195	244	211	1.929	7,6
8	63	860	413	191	209	237	1.973	7,8
9	63	738	278	186	167	172	1.604	6,3
10	39	664	301	130	168	178	1.480	5,8
11	43	898	320	236	215	166	1.878	7,4
12	42	887	336	205	212	205	1.887	7,4
Σ	843	11.137	5.139	2.556	2.935	2801	25.411	
%	3,3	43,8	20,2	10,1	11,6	11		
Frequência	2,3	30,0	13,9	6,9	7,9	7,5		

REFERÊNCIAS BIBLIOGRÁFICAS:

LI, Y.C.; KOROL, A.B.; FAHIMA, T.; NEVO, E. Microsatellites within genes: structure, function, and evolution. **Molecular Biology and Evolution**, v.21, n.6, p. 991-1007, 2004.

MAIA L.C. da, Desenvolvimento de ferramenta para análise *in silico* da ocorrência de microssatélites (*single sequence repeats*) no genoma do arroz. **Dissertação de mestrado**, Universidade Federal de Pelotas, UFPel, Brasil, 2007.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, v.30, n.2, p.194-200, 2002.

MORGANTE, M. & OLIVIERI, A.M. PCR-amplified microsatellites as markers in plant genetics. **The Plant Journal**. v.3, n.1, p.175-182, 1993.

THIEL, T.; MICHALEK, W.; VARSHNEY, W.; GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, v.106, n.3, p.411-422, 2003.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, v.23, n.1, p.48-55, 2005.