

UNIVERSIDADE FEDERAL DE PELOTAS
Programa de Pós-Graduação em Agronomia



Tese

Genômica comparativa em gramíneas

Clauber Mateus Priebe Bervaldo

Pelotas, 2009.

CLAUBER MATEUS PRIEBE BERALD

Genômica comparativa em gramíneas

Tese apresentada ao Programa de Pós-Graduação em Agronomia, da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Ciências (área do conhecimento: Fitomelhoramento).

Orientador: Antônio Costa de Oliveira, PhD.

Co-orientador: Fernando Irajá Félix de Carvalho, PhD.

Pelotas, 2009.

Dados de catalogação na fonte:

(Marlene Cravo Castillo – CRB-10/744)

B552g Bervaldo, Clauber Mateus Priebe

Genômica comparativa em gramíneas/ Clauber Mateus Priebe
Bervaldo .. -Pelotas, 2009.

113f. : il.

Tese (Doutorado em Fitomelhoramento) –Programa de Pós-Graduação em Agronomia. Faculdade de Agronomia Eliseu Maciel. Universidade Federal de Pelotas. - Pelotas, 2009, Antônio Costa de Oliveira, Orientador; co-orientador Fernando Irajá Felix de Carvalho .

1. Microssatélites 2.Elementos cis-acting 3. Evolução de gramíneas 4.Bioinformática I Oliveira, Antônio Costa de (orientador) II .Título.

CDD 581.151

Banca examinadora:

Antonio Costa de Oliveira, PhD. – Dep. de Fitotecnia, FAEM/UFPeI

Fernando Irajá Félix de Carvalho, PhD. – Dep. de Fitotecnia, FAEM/UFPeI

Beatriz Helena Gomes Rocha, Dra. – Instituto de Biologia, UFPeI

Ariano Martins de Magalhães Júnior, Dr. – CPACT/EMBRAPA

José Fernandes Barbosa Neto, PhD. – Dep. de Plantas de Lavoura,
Agronomia/UFRGS

Dedico este trabalho a Deus, a meus pais, Guido e Brunilda e aos meus irmãos, Clóvis e Clenildo, e em especial à minha noiva Fabiana. Todos estarão sempre em meu coração.

Agradeço a Deus, pelas bênçãos e por todas as coisas conquistadas na minha vida até hoje e para todo o sempre.

Aos Professores Antonio Costa de Oliveira e Fernando Irajá Félix de Carvalho pela orientação e co-orientação, pelas conversas, conselhos, experiência e conhecimentos transmitidos, pela confiança depositada e principalmente, a amizade.

Aos colegas e amigos do Centro de Genômica e Fitomelhoramento, pela colaboração nos trabalhos, partilha de conhecimentos, convívio e amizade...muito obrigado pela contribuição com a minha formação.

Aos funcionários do Centro de Genômica e Fitomelhoramento pela amizade e constante auxílio prestado.

À Universidade Federal de Pelotas, a todos os professores, em especial aos do Programa de Pós-Graduação em Ciência e Tecnologia de Sementes e em Agronomia, pela oportunidade de formação profissional durante a Pós-Graduação.

Ao CNPq e CAPES pela concessão da bolsa de estudos durante a Graduação e Pós-Graduação.

A minha família pelo carinho e apoio recebido desde os tempos de graduação e durante todo o Curso de Pós-Graduação. Em especial, a minha noiva Fabiana pelo amor, carinho e apoio de todas as horas.

A todos que, direta ou indiretamente contribuíram de alguma forma para o desenvolvimento deste trabalho, meu sincero agradecimento.

MUITO OBRIGADO!!!

Resumo

BERVALD, Clauber Mateus Priebe. **Genômica comparativa em gramíneas**. 2009, 113f. Tese (Doutorado) – Programa de Pós-Graduação em Agronomia. Universidade Federal de Pelotas, Pelotas – RS.

O uso de seqüências de DNA tem sido à base da genômica comparativa para estudos evolucionários e a transferência de informações de espécies modelo para espécies agrícolas tem revolucionado a genética molecular e estratégias de melhoramento das culturas. A combinação de métodos clássicos de genética e melhoramento com tecnologias moleculares de análise genômica abre uma nova perspectiva para a ampliação do conhecimento das bases genéticas e aceleração de programas de melhoramento. A bioinformática está se tornando em uma ferramenta que será parte essencial na pesquisa científica de plantas e que poderá contribuir muito para este desafio, permitindo fazer inferências de função, estrutura e evolução de genes e genomas, busca de marcadores, desenho de *primers*, entre outras possibilidades. Neste sentido, três estudos de genômica comparativa em gramíneas foram realizados, utilizando a bioinformática como ferramenta. No primeiro trabalho, o objetivo do trabalho foi analisar a abundância de microssatélites no genoma de 13 espécies do gênero *Oryza* obtidas do GeneBank do National Center for Biotechnology Information – NCBI, com programa SSRLocator, descrevendo as taxas, freqüências e padrão de distribuição dos diferentes microssatélites. A freqüência de microssatélites varia inversamente proporcional ao tamanho do genoma da espécie no gênero *Oryza*. As espécies de genoma A apresentam as maiores freqüências para todos os tipos de microssatélites e totais no gênero *Oryza*. Os trímeros compostos por citosinas e guaninas apresentam o maior percentual de ocorrência entre as espécies de gênero *Oryza*, sem relação direta com o conteúdo GC da espécie. No segundo trabalho, a região promotora de 1000 pares de bases dos genes *OsNramp* de *Oryza sativa* subsp *japonica* cv. Nipponbare foi investigada quanto à abundância de elementos *cis-acting*. As seqüências foram analisadas utilizando o programa “Signal Scan Search” do portal “Plant Cis-acting Regulatory DNA Elements” (PLACE) para a identificação dos diferentes elementos *cis-acting* presentes em cada uma das regiões promotoras. Foram detectados 170 diferentes elementos *cis-acting* na região promotora dos genes membros da família *OsNramp*, sendo um total de 14 elementos comuns a região promotora dos oito membros da

família gênica *OsNramp* de *Oryza sativa* subsp. *japonica*. O elemento CACTFTPPCA1 foi o motivo mais freqüente na região promotora dos genes membros da família *OsNramp* de *Oryza sativa* subsp. *japonica*. O objetivo do último trabalho foi comparar as seqüências de ESTs de aveia com as espécies *Arabidopsis*, cevada, arroz, cana-de-açúcar, sorgo, trigo e milho, buscando inferir relações evolutivas dessas espécies com a aveia. Seqüências de oito espécies de plantas disponíveis atualmente, no banco de dados Unigene no National Center for Biotechnology Information - NCBI foram baixadas e as comparações com a aveia foram feitas utilizando as ferramentas de BLASTn e tBLASTx. Observou-se que o banco de ESTs de aveia UNIGENE-NCBI consiste de muitas seqüências pouco similares a outras espécies comparados a nucleotídeos (46,72%) e aminoácidos (74,97%). Uma minoria de 4,56% das seqüências presentes no banco de ESTs de aveia UNIGENE-NCBI apresentam alta similaridade com seqüências de espécies de mono e dicotiledôneas.

Palavras-chave: Microssatélites, elementos *cis-acting*, evolução de gramíneas, bioinformática.

Abstract

BERVALD, Clauber Mateus Priebe. **Comparative genomics in grasses**. 2009, 113f. Tese (Doutorado) – Programa de Pós-Graduação em Agronomia. Universidade Federal de Pelotas, Pelotas – RS.

The use of sequences of DNA has been the base of comparative genomics for evolutionary studies. The transfer of information from model species to agricultural species has been revolutionizing the molecular genetics and strategies of crop improvement. The combination of classic methods of genetics and breeding with molecular technologies of genomic analysis opens a new perspective for the increase of the understanding of the genetic basis and the acceleration of breeding programs. Bioinformatics is becoming a tool that will be an essential part of the plant scientific research contributing to this challenge, and allowing inferences of function, structure and evolution of genes and genomes, search of markers, primers design, among other possibilities. In this sense, three studies of comparative genomics in grasses were accomplished, using bioinformatics tools. In the first work, the objective was to analyze the microsatellite abundance in genome fractions of 13 species of the genus *Oryza*, describing the rates, frequencies and pattern of distribution of the different microsatellites. The microsatellite frequency varies inversely proportional to the size of the genome of the specie of the genus *Oryza*. The A genome species presented the highest occurrences and frequencies for all types and total microsatellites. The trimers composed by cytosines and guanines presented the percentile largest of occurrence among the species, without direct relationship with the GC content of the species. In the second work, the promoter region of 1,000 bases pairs of the genes *OsNramp* of *Oryza sativa* subsp *japonica* cv. Nipponbare was investigated as for the abundance of *cis-acting* elements. The sequences were analyzed using the software “Signal Scan Search” of the website “Plant Cis-acting Regulatory DNA Elements” (PLACE) to the identification of different *cis-acting* elements present in each one of the promoter regions. Were detected 170 different *cis-acting* elements in the upstream region of the genes members of the family *OsNramp* subsp *japonica* cv. Nipponbare. A total of 14 elements were common to the eight members of the family *OsNramp* subsp *japonica* cv. Nipponbare. The element CACTFTPPCA1 was the most frequent motif in the promoter region of the genes of the family *OsNramp* of *Oryza sativa* subsp *japonica* cv. Nipponbare. The

objective of the last work was to compare the sequences of oats with the species *Arabidopsis*, barley, rice, sugarcane, *Sorghum*, wheat and corn, looking for to infer evolutionary relationships of those species with the oats. Sequences of eight species of plants evaluable at the present moment, in the Unigene data bank in the National Center for Biotechnology Information - NCBI were downloaded and comparisons with oat made using the BLASTn and tBLASTx tools. It was observed that the bank of ESTs of oats UNIGENE-NCBI consists from many sequences a little similar to other species compared to nucleotides (46,72%) and aminoacids (74,97%). Only 4,56% of the sequences present in the UNIGENE-NCBI oat EST bank presents high similarity with sequences of mono and dicotyledonous species.

Keywords: Microsatellites, *cis-acting* elements, grasses evolution, bioinformatics.

Lista de Figuras

Figura 1.1. Frequências dos diferentes tipos de microssatélites, Classe I (≥ 20 pb) nos genomas de espécies do gênero <i>Oryza</i> .	42
Figura 1.2. Frequências dos diferentes tipos de microssatélites, Classe II (≥ 12 pb e < 20 pb), nos genomas de espécies do gênero <i>Oryza</i> .	43
Figura 1.3. Ocorrência percentual dos diferentes tipos de microssatélites, Classe I (≥ 20 pb) (a), Classe II (≥ 12 pb e < 20 pb) (b) e Classe I e II (≥ 12 pb) (c) nos genomas de espécies do gênero <i>Oryza</i> .	44
Figura 1.4. Ocorrência percentual média dos diferentes tipos de microssatélites, Classe I (≥ 20 pb) e Classe II (≥ 12 pb e < 20 pb), nos genomas de espécies do gênero <i>Oryza</i> .	45
Figura 1.5. Tamanho médio (pb) de locos dos diferentes tipos de microssatélites, Classe I (≥ 20 pb, em azul) e Classe II (≥ 12 pb e < 20 pb, em vermelho) nos genomas de espécies do gênero <i>Oryza</i> .	46
Figura 2.1. Número de elementos <i>cis-acting</i> encontrados nos genes <i>OsNramp</i> no genoma de <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare.	75
Figura 2.2. Ocorrência total dos elementos <i>cis-acting</i> em comum por gene da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare.	77
Figura 2.3. Ocorrência total de cada elemento <i>cis-acting</i> considerando todos os genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare.	79
Figura 2.4. Posicionamento do elemento <i>cis-acting</i> CACTFTPPCA1 mais comum encontrado nas regiões promotoras dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare (posicionamento da esquerda para direita)	80
Figura 2.5. Posicionamento do elemento <i>cis-acting</i> DOFCOREZM mais comum encontrado nas regiões promotoras dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare (posicionamento da esquerda para direita).	81
Figura 2.6. Posicionamento do elemento <i>cis-acting</i> CAATBOX1 mais comum encontrado nas regiões promotoras dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare (posicionamento da esquerda para direita).	82

Figura 2.7. Alinhamento do elemento <i>cis-acting</i> CACTFTPPCA1 (YACT) mais comum encontrado nas regiões promotoras dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare.	83
Figura 2.8. Dendograma da região promotora dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare, pelo método de agrupamento UPGMA.	84
Figura 2.9. Dendograma dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare, pelo método de agrupamento UPGMA.	85
Figura 3.1. Porcentagem de seqüências similares nas espécies para nucleotídeos (N) e aminoácidos (A), baseados na comparação realizada entre aveia e as demais espécies (7–número de espécies consideradas no cálculo de porcentagem de seqüências de aveia similares presentes, e assim consecutivamente, 0–seqüências de aveia sem similaridade).	105
Figura 3.2. Número de seqüências encontradas similares a nucleotídeos, aminoácidos, e ambos na comparação realizada entre aveia e as demais espécies.....	106
Figura 3.3. Número total de seqüências similares encontradas na comparação realizada entre aveia e as demais espécies e seqüências depositadas no banco UNIGENE.	107
Figura 3.4. Número de seqüências de aveia similares a <i>Oryza sativa</i>	108
Figura 3.5. Classificação funcional dos genes conservados entre as oito espécies.	109

Lista de Tabelas

Tabela 1.1.	Quantidade total de ocorrências e frequências (locos/mb) dos diferentes tipos de microssatélites, Classe I (>20pb) e Classe II (>12pb e <20pb), nos genomas de espécies do gênero <i>Oryza</i>	41
Tabela 1.2.	Distribuição da ocorrência de motivos monômeros, dímeros, trímeros e tetrameros, pentâmeros e hexâmeros e percentuais em espécies do gênero <i>Oryza</i>	47
Tabela 2.1.	Cromossomos de <i>Oryza sativa</i> subsp. <i>japonica</i> e os genes <i>OsNramp</i> envolvidos na homeostasia de ferro em plantas.	74
Tabela 2.2.	Genes <i>OsNramp</i> do genoma de <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare, com os respectivos cromossomos, quantidade de elementos <i>cis-acting</i> em cada loco, número de diferentes elementos <i>cis-acting</i> por loco, elementos <i>cis-acting</i> mais comuns encontrados em cada loco, com o respectivo número de ocorrências no loco.	76
Tabela 2.3.	Relação dos elementos <i>cis-acting</i> comuns encontrados na região promotora dos genes da família <i>OsNramp</i> em <i>Oryza sativa</i> subsp <i>japonica</i> cv. Nipponbare.	78
Tabela 3.1.	Espécies com o número de seqüências encontradas no banco UniGene do NCBI em 06/06/2009.	104

Sumário

Resumo.....	3
Abstract	5
Lista de Figuras.....	7
Lista de Tabelas	9
Introdução Geral.....	14
1. Microssatélites em espécies do gênero <i>Oryza</i>	18
INTRODUÇÃO.....	19
MATERIAL E MÉTODOS	22
RESULTADOS E DISCUSSÃO	23
CONCLUSÕES.....	34
REFERÊNCIAS	35
2. Elementos <i>cis-acting</i> em regiões promotoras de genes <i>OsNramp</i>.	49
INTRODUÇÃO.....	50
MATERIAL E MÉTODOS	53
RESULTADOS E DISCUSSÃO	54
CONCLUSÕES.....	65
REFERÊNCIAS	66
3. Análise de genômica comparativa de ESTs de aveia com espécies gramíneas e <i>Arabidopsis</i>.....	86
INTRODUÇÃO.....	87
MATERIAL E MÉTODOS	91
RESULTADOS E DISCUSSÃO	92
CONCLUSÕES.....	97
REFERÊNCIAS	98
Discussão geral.....	110
Referências Bibliográficas (Introdução e Discussão Geral).....	112
Vitae.....	114

Introdução geral

O melhoramento de plantas é a arte e a ciência de modificar geneticamente as plantas, em benefício do homem (ALLARD, 1971, FEHR, 1987). O melhoramento como arte depende da intuição e das experiências acumuladas, inerentes a cada indivíduo. Apesar de toda a evolução nas técnicas de melhoramento, a arte continua a ser um componente muito importante na carreira de um melhorista (FEDERIZZI, 1998). Como ciência, o melhoramento de plantas é multidisciplinar envolvendo o conhecimento da agronomia, botânica, genética e citogenética, genética molecular, fisiologia, patologia, entomologia, bioquímica, e estatística (SCHELGEL, 2003).

O sucesso da produção agrícola depende fortemente do uso de variedades com desempenho superior e adaptadas ao meio de cultivo, sendo este o objetivo contínuo dos programas de melhoramento genético de qualquer espécie cultivada (LORENCETTI et al., 2006). O grande desafio dos melhoristas de plantas consiste em disponibilizar, permanentemente, genótipos de qualidade elevada e identificar constituições genéticas que superem o rendimento de grãos e outros caracteres agrônômicos expressados pelos cultivares existentes no mercado (BROWN e FORSBERG, 1987, MARCHIORO et al. 2005). Para isto, o melhoramento genético de plantas requer três etapas fundamentais para a obtenção de genótipos superiores: presença da variabilidade genética, eficiência na seleção dos genótipos mais promissores e ajuste das melhores constituições genéticas ao ambiente de cultivo (CARVALHO et al., 2003).

Os melhoristas de plantas do mundo inteiro têm buscado constantemente aumento na eficiência da seleção, o melhor conhecimento e caracterização do

germoplasma e a maximização dos ganhos genéticos. Novas formas de alcançar estes objetivos têm sido constantemente perseguidas pelo melhoramento. (MILACH, 1998a). Com o advento das técnicas de biologia molecular, uma das alternativas tem sido a utilização de marcadores moleculares.

Os marcadores moleculares são definidos como qualquer fenótipo molecular oriundo de um gene expresso ou de um segmento específico de DNA (TANSKLEY, 1983; FERREIRA & GRATTAPAGLIA, 1998). MILACH (1998a) descreve que marcadores são características de DNA que diferenciam dois ou mais indivíduos e são herdados geneticamente. Os marcadores moleculares propiciam, aos pesquisadores, o acesso à variabilidade em nível de DNA em plantas, sem sofrerem a influência do ambiente em relação aos caracteres morfológicos, sendo uma poderosa ferramenta para programas de melhoramento genético (MILACH 1998b; ZIMMER et al., 2005).

Os marcadores de DNA são livres de efeitos pleiotrópicos, permitindo, dessa forma, que qualquer número de marcadores seja monitorado em uma única população. A análise de marcadores moleculares pode ser executada em qualquer estágio do ciclo de vida de um organismo e a partir de quase qualquer tipo de tecido, incluindo tecidos de herbários e mumificados. Muitos marcadores de DNA apresentam expressão co-dominante, permitindo que genótipos sejam determinados em qualquer esquema de melhoramento (KUMAR, 1999).

A área de maior impacto dos marcadores moleculares no melhoramento vegetal é através do emprego da seleção assistida por marcadores para identificação de genótipos superiores em populações segregantes (FEDERIZZI, 1998). A seleção assistida por marcadores moleculares pode, potencialmente, acelerar o progresso genético pelo aumento na eficiência de seleção, redução no intervalo de gerações e identificação de quebra de ligações indesejáveis (OLIVEIRA, 2007). Entretanto, como se acreditava inicialmente, essa tecnologia não substitui o melhoramento convencional na avaliação do resultado desses processos (VARSHNEY et al., 2005a).

A seleção assistida por marcadores baseia-se na concepção de que é possível inferir a presença de um gene por meio da presença de um marcador fortemente ligado a esse gene. Inicialmente, são identificados marcadores moleculares ligados ao gene de interesse e por meio desses marcadores, a herança e a introgressão do gene pode ser analisada nos programas de melhoramento.

Usando os marcadores é possível não só selecionar plantas que contêm os genes de interesse, mas também que carregam uma grande parte de genes do genoma do genitor doador (KUMAR, 1999). O uso de marcadores moleculares possibilita a seleção de vários caracteres ao mesmo tempo, permite a identificação rápida de plantas geneticamente superiores em cada geração e reduz o tamanho da população a ser desenvolvida em cada geração, minimizando esforços e recursos nas avaliações.

Outro fato importante nos últimos anos, é que a genômica e a bioinformática têm invadido as pesquisas em biologia, genética e melhoramento. Durante as últimas décadas, os principais avanços no campo da biologia molecular, junto com os avanços nas tecnologias genômicas, conduziram a um crescimento na informação biológica gerada pela comunidade científica. Principalmente nestas duas últimas décadas, o armazenamento de dados biológicos em bancos de dados públicos, vem se tornando cada vez mais comum e esses bancos de dados têm crescido exponencialmente. Essa revolução está ligada a importantes progressos nos métodos e tecnologias que permitem o seqüenciamento de DNA em grande escala (OLIVEIRA, 2007). A grande quantidade de informações geradas conduziu a uma exigência absoluta de bancos de dados computadorizados para armazenar, organizar e indexar os dados e por ferramentas especializadas para visualizar e analisar os dados (NCBI, 2009). A literatura biológica também está crescendo fortemente, dessa forma, é impossível até mesmo para o pesquisador mais cuidadoso, manter-se atualizado com as informações necessárias da área sem o auxílio de ferramentas computacionais (MAIA, 2007).

Conseqüentemente, surge uma nova área do conhecimento chamada de bioinformática, o campo da ciência em que a biologia, ciência da computação e a tecnologia da informação se fundem para formar uma única disciplina. Segundo GOODMAN (2002), a bioinformática é uma arte, uma engenharia e uma ciência, que esta cercada pelo desenvolvimento de um novo método computacional e a aplicação destes para resolver problemas biológicos e ser capaz de descobrir novos aspectos biológicos. A comparação de seqüências de DNA, RNA e seqüências protéicas fornece a base para muitas ferramentas de bioinformática e pode permitir a inferência da função, estrutura e evolução de genes e genomas (RHEE et al., 2006).

Como a quantidade de dados cresce exponencialmente, o campo da bioinformática evoluiu de tal forma que a tarefa mais urgente agora envolve a análise

e interpretação de vários tipos de dados, incluindo seqüências de nucleotídeos e aminoácidos, domínios de proteínas e estruturas de proteínas. Segundo SUMNER et al., (2003), ferramentas apropriadas de bioinformática e métodos de análise são a chave para a integração da metabolômica com outras ferramentas de genômica funcional, e a ciência de plantas em geral. Desta forma, a bioinformática está se tornando uma ferramenta que será parte essencial na pesquisa científica de plantas, e acredita-se que cada pesquisador irá incorporar cada vez mais ferramentas de bioinformática em seus projetos de pesquisa (RHEE et al., 2006).

A combinação de métodos clássicos de genética e melhoramento com tecnologias moleculares de análise genômica abre uma nova perspectiva para a ampliação do conhecimento genético e para a aceleração de programas de melhoramento (FERREIRA e GRATTAPAGLIA, 1998). Atualmente, a bioinformática também pode contribuir para este desafio. Neste sentido, o objetivo geral desta tese foi contribuir com a pesquisa científica com a realização de trabalhos aplicados as plantas, principalmente gramíneas, utilizando a bioinformática como ferramenta.

O trabalho teve os seguintes objetivos específicos: (1) analisar os genomas de 14 espécies do gênero *Oryza*, buscando elucidar as taxas, freqüências e padrão de distribuição de diferentes microssatélites; (2) verificar os elementos *cis-acting* em regiões promotoras de genes *OsNramp*; (3) analisar comparativamente seqüências de ESTs de aveia e outras gramíneas e *Arabidopsis* através da genômica comparativa.

1. Microsatélites em espécies do gênero *Oryza*

Genetics and Molecular Biology (ISSN 1415-4757)

RESUMO

O objetivo do trabalho foi analisar a abundância de microsatélites no genoma de 13 espécies do gênero *Oryza* obtidas do GeneBank do National Center for Biotechnology Information – NCBI, com o programa SSRLocator, descrevendo as taxas, frequências e padrão de distribuição dos diferentes microsatélites. A frequência de microsatélites varia inversamente proporcional ao tamanho do genoma da espécie no gênero *Oryza*. As espécies de genoma A apresentam as maiores frequências para todos os tipos de microsatélites e totais no gênero *Oryza*. Os trímeros compostos por citosinas e guaninas apresentam o maior percentual de ocorrência entre as espécies de gênero *Oryza*, sem relação direta com o conteúdo GC da espécie.

INTRODUÇÃO

Entre todas as técnicas, uma classe de marcadores moleculares muito promissores são os microssatélites ou repetições de seqüências simples (SSRs). A abundância genômica e o potencial inerente de variação, oriundo dessas mutações, tornam esta classe de marcadores uma poderosa ferramenta para várias aplicações na genética e melhoramento vegetal.

Microssatélites são seqüências de DNA formadas pela disposição em série de nucleotídeos repetidos em arranjos formados entre dois e seis pares de bases (Morgante e Olivieri, 1993). As regiões microssatélites estão mais propensas à ocorrência de laços (*loops*) ou estruturas conhecidas como *hairpins*, promovendo um aumento ou redução no número de seqüências das repetições devido à inserção ou exclusão de pares de bases, pois nestes trechos, durante o processo de replicação, a *DNA polimerase* pode sofrer um deslize (Iyer *et al.*, 2000). Ainda, os mesmos podem ser encontrados em qualquer genoma (eucarioto e procarioto) e em regiões-codificantes e não-codificantes (Katti, Ranjekar, e Gupta, 2001; Toth, Gaspari, e Jurka 2000), com SSRs sendo mais abundantes em regiões não-codificantes do que em exons (Hancock, 1995). O locus pode ser perfeito, imperfeito (interrompido por vários nucleotídeos não repetidos) ou por repetições compostas (Hearne *et al.*, 1992).

Estes marcadores são muito promissores por serem úteis em uma variedade de aplicações na genética e no melhoramento de plantas pela sua reprodutibilidade, natureza multialélica, alto grau de polimorfismo, herança co-dominante, relativa abundância e boa cobertura do genoma (Powell *et al.*, 1996). Os microssatélites tornaram-se uma fonte valiosa de marcadores genéticos (Temnykh, 2001; Varshney *et al.*, 2005b) por causa da sua abundância e potencial inerente para variação. Estes

marcadores são úteis para integração de mapas genéticos e físicos e tem provido os melhoristas e geneticistas de plantas com uma ferramenta eficiente para associar variação genética e fenotípica (Gupta e Varshney, 2000; Varshney *et al.*, 2005b).

Os microssatélites (revisado por Maia *et al.*, 2008) foram estudados em diferentes espécies, sendo reportadas ocorrências que correspondem a 0,85% do genoma de *Arabidopsis thaliana*, 0,37% do milho (*Zea mays* subsp. *mays*), 3,21% do peixe *Fugu rubripes*, 0,21% do *Caenorhabditis elegans*, 0,30% do *Sacharomyces cerevisiae* (Morgante *et al.*, 2002) e 3% no genoma humano (Subramanian *et al.*, 2003).

Os microssatélites são uma importante classe de marcadores moleculares utilizada para entender relações de espaço entre segmentos de cromossomos, podendo ajudar na análise das relações temporais e evolutivas entre espécies e gêneros (Kashi *et al.*, 1997; Maia, 2008). Muitas espécies selvagens têm sido usadas como fonte de variabilidade genética nos programas de melhoramento para espécies cultivadas. No caso do arroz, existe uma extensa variabilidade genética disponível no gênero que proporciona uma grande oportunidade para a melhoria da cultura no futuro (Chakravarthi e Naravaneni, 2006). O gênero *Oryza* compreende um extenso grupo formado atualmente por 23 espécies (Londo *et al.*, 2006), sendo as espécies *Oryza sativa* spp *japonica* e *O. sativa* spp *indica* as mais cultivadas e estudadas. Embora, as outras espécies não sejam de grande importância econômica e agrícola, podem ser muito úteis para o melhoramento genético como fonte de novas formas de alelos ou diferentes genes. Neste sentido, o desenvolvimento de marcadores alelo-específico para genes que controlam caracteres agrônômicos serão importantes para o avanço científico no melhoramento de plantas. Além disso, o genoma do arroz está completamente

seqüenciado e este fato poderá ser importante para a obtenção de marcadores, devido à proximidade com as outras espécies do gênero *Oryza*.

A obtenção de novos marcadores moleculares microssatélites é, usualmente, uma tarefa lenta e de altos custos, especialmente quando feita em laboratório a partir de protocolos convencionais. Considerando a vasta utilidade dos microssatélites, vários grupos de pesquisa tem se esforçado em caracterizar a sua abundância, distribuição e localização genômica utilizando métodos *in silico* (La Rota *et al.*, 2005; Garnica *et al.*, 2006, Grover e Sharma, 2007). As ferramentas *in silico* foram desenvolvidas para facilitar e reduzir os custos com marcadores moleculares. Neste contexto, o uso da bioinformática para análise de bancos de dados que disponibilizam trechos de regiões seqüenciadas do DNA de espécies de interesse, na busca de áreas de DNA repetitivo pode fornecer novos marcadores moleculares (Varshney *et al.*, 2005a).

Microssatélites tem sido freqüentemente usados para auxiliar na classificação e identificação de diferentes espécies (Lawson e Zhang, 2006), visto que SSRs polimórficos aumentam a possibilidade de detecção de diferenças alélicas entre espécies próximas, dentro de uma espécie, ou até mesmo entre indivíduos numa população (Yu *et al.*, 1999). Os microssatélites dividem-se em Classe I (≥ 20 pb) e Classe II (≥ 12 pb e < 20 pb) baseado no tamanho e seu potencial como marcadores genéticos informativos com aplicação para o melhoramento genético (Temnkyh *et al.*, 2001). Marcadores moleculares indexados no genoma de espécies *Oryza* podem facilitar estudos moleculares e a seleção assistida por marcadores moleculares, logo o conhecimento da ocorrência e freqüência dos diferentes tipos de SSRs nos genomas das espécies *Oryza* contribuirá para o entendimento da sua distribuição e

também no desenvolvimento de marcadores SSRs promissores para análises genéticas.

Neste sentido, o objetivo do trabalho foi analisar os genomas de 14 espécies do gênero *Oryza*, buscando elucidar as taxas, frequências e padrão de distribuição de diferentes microssatélites, assim como apontar os locos mais promissores para serem usados como marcadores moleculares, considerando que pouco ainda é conhecido sobre a abundância e variação de microssatélites nestas espécies.

MATERIAL E MÉTODOS

As seqüências dos genomas de 13 espécies do gênero *Oryza* foram obtidas do GeneBank do National Center for Biotechnology Information - NCBI (<http://www.ncbi.nlm.nih.gov/>). As frações genômicas das espécies *Oryza* baixadas e seus respectivos tamanhos entre parênteses foram *O. alta* (81 Mbp), *O. australiensis* (176 Mbp), *O. brachyantha* (46 Mbp), *O. coarctata* (136 Mbp), *O. glaberrima* (41 Mbp), *O. granulata* (98 Mbp), *O. minuta* (102 Mbp), *O. nivara* (72 Mbp), *O. officinalis* (76 Mbp), *O. punctata* (51 Mbp), *O. ridleyi* (137 Mbp), *O. rufipogon* (96 Mbp) e *O. sativa*, de subespécies *indica* (339 Mbp) e *japonica* (371 Mbp).

As seqüências foram depositadas em arquivos no padrão Fasta em computadores do Centro de Genômica e Fitomelhoramento (CGF/FAEM/UFPel) e, posteriormente, analisadas utilizando o programa computacional SSRLocator (MAIA *et al.*, 2008) (<http://www.ufpel.edu.br/~lmaia.faem>) para a localização dos microssatélites. O programa foi configurado para localizar microssatélites Classe I e Classe II, nos tipos compreendidos entre monômeros e hexâmeros. O número mínimo de repetições para monômeros foi 12, seis para dímeros, quatro para trímeros, três para tetrâmeros e pentâmeros e dois para hexâmeros.

Após a busca, estes resultados foram analisados quanto à quantidade total de ocorrências, distribuição, percentagem, frequências e tamanho dos diferentes motivos SSRs.

RESULTADOS E DISCUSSÃO

Ocorrência Geral de Microssatélites

Neste trabalho de busca da distribuição de microssatélites em 13 espécies de arroz não foi realizada discriminação entre as diferentes partes funcionais do genoma. O objetivo principal foi obter um padrão de distribuição dos microssatélites nos genomas.

Na Tabela 1.1 são apresentados os genomas das espécies *Oryza* e seus respectivos tamanhos seqüenciados e tamanhos estimados em milhões de pares de bases (Mpb) (Ammiraju *et al.*, 2006), seguidos pela quantidade total de ocorrências para microssatélites Classe I e Classe II (≥ 12 pb) por tipo de microssatélite e a frequência do número de locos para cada milhão de pares de bases (locos/Mbp) em cada genoma. A espécie *O. sativa* subsp. *indica* e *O. sativa* subsp. *japonica* é a única espécie completamente seqüenciada até o momento, por isso, mesmas foi tomada como referência no estudo.

Os resultados revelaram grande variação nas quantidades totais de microssatélites, sendo encontrados desde 41.056 locos em *O. glaberrima* a 493.099 locos em *O. sativa* sub. *japonica*. As espécies *Oryza* com genoma de tamanho aproximado a 400 Mb apresentaram as maiores frequências totais, comparado as espécies que possuem genomas próximos a 1.000 Mb. As espécies com as maiores frequências são diplóides e possuem genoma AA, exceto *O. punctata* (BB, BBCC) e

O. brachyantha (FF) com genomas diferentes (Vaughan, Morishima e Hadowaki, 2003). Em *O. sativa* subsp. *indica* e *O. sativa* subsp. *japonica*, a maior frequência para todos os tipos de microssatélites e totais era esperada devido as maiores frações genômicas analisadas para esta espécie.

As cinco espécies de genoma AA, *O. punctata* (genoma BB, BBCC), e *O. brachyantha* (genoma FF), possuem genomas pequenos comparado com as demais espécies do gênero *Oryza* de tamanhos estimados maiores (Ammiraju *et al.*, 2006), nas quais, os bancos de dados ainda são pequenos, compreendendo inclusive algumas das espécies com genoma AA.

Os tipos de microssatélites com as maiores ocorrências e frequências entre as espécies foram os hexâmeros, trímeros e tetrâmeros seguidos pelos demais microssatélites (Tabela 1.1). Resultados semelhantes para os tipos de microssatélites com maior ocorrência e frequência foram encontrados em seqüências genômicas de arroz, no genoma completo de *Arabidopsis* para os hexâmeros e trímeros e em clones genômicos de *Medicago truncatula* somente para hexâmeros (Mun *et al.*, 2006). Ainda pode ser observado um padrão de ocorrência de microssatélites que varia inversamente ao tamanho do genoma para os monômeros, trímeros e hexâmeros, sendo também observado em dímeros e trímeros, com exceção para alguns genomas.

Deve ser dada particular importância para os trímeros e hexâmeros, pois estes tipos de seqüências repetitivas podem formar códons, e assim conseqüentemente codificar para um ou mais aminoácidos (Nelson e Cox, 2004). Da mesma forma, a alteração no número de repetições nestes microssatélites não causa mudanças na seqüência de leitura (frames), sendo mais tolerados por mecanismos associados à pressão de seleção.

Considerando o tamanho atual dos bancos de dados e o tamanho estimado do genoma destas espécies, o número de microssatélites encontrado no estudo pode ser tomado como amostra e servir para estimar a distribuição completa dos microssatélites nestas espécies.

Freqüências de Microssatélites Classe I e II

Os dímeros e os hexâmeros foram os locos microssatélites mais freqüentes para as classes I e II (Figuras 1.1 e 1.2), respectivamente. As freqüências dos diferentes tipos de microssatélites nos genomas de espécies do gênero *Oryza*, para a Classe I e Classe II foram maiores nas duas espécies seqüenciadas, exceto para monômeros. No entanto, pode se observar uma maior freqüência de locos microssatélites para aquelas espécies com genomas menores, e conseqüentemente ocorrências menores para as espécies com genomas maiores, exceto em *O. coarctata* nos dímeros, pentâmeros e hexâmeros da Classe II. Estes dados corroboram com os resultados encontrados de ocorrência e freqüência geral de microssatélites (Tabela 1.1).

Ocorrência percentual e percentual média dos diferentes tipos de microssatélites

Os resultados percentuais médios de ocorrência (Figura 1.4) mostram que os dímeros apresentaram a maior ocorrência (49%) de microssatélites Classe I, enquanto que para a Classe II, os hexâmeros foram os mais freqüentes com 69,5%. Em ambas as classes, a segunda maior ocorrência foi encontrada para os microssatélites dímeros, com 20,2% e 11,5% para a classe I e II, respectivamente.

A ocorrência percentual de microssatélites para as Classes I e II (Figuras 1.3A e 1.3B) foi maior para dímeros e hexâmeros, respectivamente. As espécies *Oryza*

com tamanho de genoma próximo a 400 Mb, sendo a maioria com genoma AA, apresentaram ocorrências menores que 49% para os dímeros na Classe I, exceto para *O. brachyantha* (genoma FF) com 52,5%. Além destas, as espécies *O. officinalis* (632 Mb, genoma CC) e *O. ridleyi* (1283 Mb, genoma HHJJ) também apresentaram valores de ocorrência menor que a média. As espécies com genomas com aproximadamente 1000 Mb, revelaram ocorrências maiores que a média. Pode ser observada a mesma tendência para os hexâmeros na Classe II, em que os genomas menores apresentaram ocorrências menores que a média de 69,5%, com exceção da espécie *O. punctata* com ocorrência superior. Para as espécies com genomas maiores, as ocorrências ficaram todas acima da média encontrada para o tipo de microsatélites nesta classe. Estes resultados indicam que as ocorrências percentuais encontradas estão relacionadas com o tamanho do genoma estimado por Ammiraju *et al.* (2006).

Quando os resultados para as classes I e II são analisados conjuntamente (Figura 1.3C), os hexâmeros apresentaram a maior ocorrência percentual, seguidos dos dímeros, trímeros, pentâmeros e monômeros, respectivamente.

Tamanho dos Microsatélites

Na Figura 1.5 são apresentados separadamente os tamanhos médios dos diferentes tipos de microsatélites para as Classes I e II, em cada genoma das espécies do gênero *Oryza* analisadas no estudo. Em ambas as classes, o maior tamanho médio foi detectado nos dímeros, assim como também a maior variação média nos locos microsatélites para as duas classes. Além disso, apareceram valores um pouco discrepantes na Classe I para os monômeros em *O. alta* e para

hexâmeros em *O. australiensis*, resultados estes justificados pela presença de alguns locos com tamanhos muito superiores a média.

Os microssatélites são divididos nas classes I e II, baseado no tamanho e seu potencial como marcadores genéticos informativos com aplicação para o melhoramento genético (Temnkyh *et al.*, 2001). Contudo, os microssatélites Classe I são descritos como os melhores na utilização como marcadores moleculares, pois são mais polimórficos devido à maior probabilidade de mutações em locos maiores, logo os locos dos tipos dímeros desta classe, apontados neste estudo são os melhores candidatos a marcadores.

Distribuição da ocorrência e percentuais de motivos monômeros, dímeros, trímeros, tetrameros, pentâmeros e hexâmeros

Os resultados para motivos, valores percentuais da ocorrência e total para cada espécie (Tabela 1.2) revelaram diferenças no padrão de ocorrências dos tipos de motivos encontrados nas espécies de arroz.

Monômeros

Os microssatélites monômeros apresentaram diferenças no padrão de ocorrência dos motivos C/G e A/T entre as espécies de arroz. Nas espécies *O. alta*, *O. brachyantha*, *O. sativa* subsp. *japonica*, os motivos C/G foram mais freqüentes com 57, 72 e 71%, respectivamente, enquanto que nas espécies *O. glaberrima*, *O. granulata*, *O. minuta*, *O. nivara*, *O. officinalis*, *O. punctata*, *O. rufipogon*, *O. australiensis*, *O. coarctata*, *O. ridleyi* e *O. sativa* subsp. *indica* apresentaram os motivos A/T mais freqüentes, com variação entre 55 e 78%.

Dímeros

Para os motivos dímeros também foram observadas diferenças nos padrões de ocorrência dos motivos entre as espécies *Oryza*. Os motivos AT/AT e TA/TA foram os mais abundantes, apresentando os maiores percentuais de ocorrência nas espécies *O. alta* (42 e 23%), *O. brachyantha* (30 e 27%), *O. minuta* (30 e 24%), *O. officinalis* (28 e 25%), *O. punctata* (27 e 23%), *O. australiensis* (34 e 22%), *O. coarctata* (31 e 26%), *O. ridleyi* (34 e 26%). McCouch *et al.* (1997) também encontraram o motivo dímero AT como o mais abundante em plantas, enquanto que AC foi o mais abundante para o genoma humano. Ainda foram encontrados os motivos dímeros AG/CT e AT/AT em *O. glaberrima* (24 e 21%), *O. sativa* subsp. *indica* (22 e 21%) e *O. sativa* subsp. *japonica* (34 e 29%), e GA/TC e AG/CT (23 e 22%) em *O. nivara*, sendo que estes motivos estão incluídos entre os maiores percentuais de ocorrência de microssatélites dímeros.

Os resultados do trabalho estão de acordo com Kumptala e Mukhopadhyay (2005), que encontraram o motivo AT como o segundo dímero mais abundante em espécies dicotiledôneas. Em solanáceas, o motivo AT/AT foi o mais freqüente com 8,29%, respectivamente (Maia *et al.*, 2009). Morgante *et al.* (2002), observaram resultados diferentes, com as repetições AT ocorrendo em freqüência mais baixa do que os demais motivos dímeros. Morgante *et al.* (2002) e Varshney *et al.* (2002) também demonstraram que AG/CT foram os microssatélites dímeros mais freqüentes observados em ESTs de plantas.

As repetições AC/CA/TG/GT apresentaram baixa freqüência nas espécies estudadas, corroborando com observações feitas por Kumptala e Mukhopadhyay (2005). Em trabalho de Varshney *et al.* (2002) com ESTs de cereais, os mesmos observaram que a repetição AC (igual ao grupo AC/CA/TG/GT no estudo) foi o

segundo dímero mais freqüente. A escassez das repetições AG/CT também foi observada nas espécies de arroz, corroborando com resultados encontrados também em outras espécies de plantas (Morgante e Olivieri, 1993, Lagercrantz *et al.*, 1993, Kumptala e Mukhopadhyay, 2005). No entanto, Maia *et al.* (2009) estudando microssatélites em Brassicas, Solanáceas e Poáceas, observaram que os dímeros AG/CT (6,72%) e GA/TC (5,61%) foram os motivos mais freqüentes nos cereais.

Em todas as espécies analisadas, as menores ocorrências foram para aqueles motivos formados por guanina e citosina (CG/CG), sendo que na espécie *O. alta* onde não foram encontrados locos formados por esses arranjos. Resultados semelhantes foram encontrados em *Medicago truncatula* (Mun *et al.*, 2006) e em Poaceas e até mesmo ausentes em espécies Brassicas e Solanáceas (Maia *et al.*, 2009). Uma explicação pode ser que as repetições (GC)_n são selecionadas contra devido ao aumento na estabilidade das estruturas de laços (GC)_n.

Outros estudos mostraram que os motivos AG e GA foram os mais freqüentes em *Arabidopsis* (Cardle *et al.*, 2000; Morgante *et al.*, 2002; Lawson e Zhang, 2006; Parida *et al.*, 2006) e AT/AT em *B. rapa* (Hong *et al.*, 2007). Em estudos com ESTs das mesmas espécies foram encontradas taxas de ocorrência entre 20-25% para o dímero AG e 15-20% para AT (Kumptla e Mukhopadhyay, 2005). Em outros estudos foram detectadas taxas de ocorrência entre 38-50% para o motivo AG em *Z. mays*, *H. vulgare*, *O. sativa*, *S. bicolor* e *T. aestivum* (Varshney *et al.*, 2002; Morgante *et al.*, 2002; Kantety *et al.*, 2002; Thiel *et al.*, 2003; Yu *et al.*, 2004; La Rota *et al.*, 2005) e taxas de 50% de AC em *H. vulgare* (Varshney *et al.*, 2002). Entretanto, alguns estudos mostraram GA como o motivo mais abundante em algumas dessas

gramíneas (Temnykh *et al.*, 2001; Kantety *et al.*, 2002; Nicote *et al.*, 2004; Parida *et al.*, 2006).

Trímeros

Os motivos compostos por C e G (CGC/GCG, GCC/GGC e CCG/CGG) apresentaram os maiores percentuais de ocorrência (24 a 43%) entre os trímeros na maioria das espécies, exceto nas espécies *O. alta*, *O. australiensis* e *O. ridleyi*. Os motivos ACT/AGT (*O. nivara*, *O. officinalis*, *O. coarctata* e *O. ridleyi*), ACA/TGT (*O. punctata*, *O. rufipogon*, *O. sativa* subsp. *indica* e *O. sativa* subsp. *japônica*) e AAC/GTT (*O. brachyantha*, *O. glaberrima* e *O. minuta*) apresentaram as menores ocorrências revelando serem os mais raros na maioria das espécies (1%). Os resultados encontrados para os motivos trímeros compostos por C e G foram muito variáveis e não podem ser relacionados com o tamanho do genoma e conteúdo GC das espécies.

Em estudo com cromossomos de arroz, os microssatélites trímeros (CGG/GCC)_n foram os motivos mais freqüentes (Akagi *et al.*, 1996). Os trímeros compostos por C e G foram os motivos mais freqüentes em Poaceas (Maia *et al.*, 2009). Outros estudos mostraram predominância na ocorrência do motivo CCG em espécies gramíneas (*Z. mays*, *H. vulgare*, *O. sativa*, *S. bicolor*, *T. aestivum*, *S. cereale* e *S. officinalis*) (Cordeiro *et al.*, 2001; Varshney *et al.*, 2002; Morgante *et al.*, 2002; Kantety *et al.*, 2002; Thiel *et al.*, 2003; Yu *et al.*, 2004; Peng *et al.*, 2005; Nicot *et al.*, 2004; La Rota *et al.*, 2005). Em *Arabidopsis* e *Brassica rapa* (Morgante *et al.*, 2002; Hong *et al.*, 2007) foram encontrados resultados contrastantes que indicaram ocorrência predominante do motivo AAG. Os motivos AAG/AGA/GAA foram os mais

freqüentes em *Arabidopsis*, *B. napus*, *B. rapa*, *S. lycopersicum* e *S. tuberosum* (Kumptla e Mukhopadhyay, 2005) e em *Citrus* (Jiang et al., 2006).

Em regiões gênicas do genoma, estes códons compostos por C e G têm potencial para codificar para os aminoácidos prolina (CCG), arginina (CGG, CGC), alanina (GCG, GCC) e glicina (GGC) (Nelson e Cox, 2004), porém entre estes, a expansão do motivo direcionando para adições do aminoácido prolina pode ter efeitos fortes na estrutura da proteína, enquanto em alanina e glicina terão efeitos bem menores (La Rota et al., 2005).

Tetrâmeros, Pentâmeros e Hexâmeros

Para os locos microssatélites formados por arranjos superiores a três nucleotídeos (tetrâmeros, pentâmeros e hexâmeros), somente os três motivos mais freqüentes, assim como os totais de cada um dos tipos de microssatélites são apresentados na tabela. Os resultados indicam diferenças no padrão de ocorrências dos três tipos de motivos entre as espécies de arroz.

Os tetrâmeros apresentaram um padrão de ocorrência de motivos com diferenças entre as espécies. Os motivos ATTA/TAAT em *Oryza brachyantha* (37%), *O. glaberrima* (17%), *O. nivara* (15%), *O. officinalis* (14%), *O. australiensis* (18%), *O. coarctata* (17%), *O. sativa* subsp. *indica* (17%), AAAT/ATTT em *O. minuta* (15%), *O. punctata* (16%), *O. ridleyi* (17%), e GAAA/TTTC em *O. alta* (18%) e *O. rufipogon* (16%) foram os que apresentaram as maiores ocorrências médias. Esses motivos foram os mais freqüentes entre as espécies, sempre alternando as primeiras posições entre si com eventuais exceções. Em *Hordeum vulgare*, ACGT foi o motivo mais abundante (Thiel et al., 2003). Estudos com *Arabidopsis* indicaram que os

motivos AAAG e AAAT foram os mais freqüentes e AAAT em *B. rapa* (Cardle *et al.*, 2000; Hong *et al.*, 2007).

Para os pentâmeros, os motivos GAAAA/TTTTTC e AAAAT/ATTTT juntamente com AAAAG/CTTTT foram os três mais freqüentes na maioria das espécies. Entre estes, o motivo AAAAG/CTTTT foi o mais comum, sendo encontrado em *O. minuta* (18%), *O. nivara* (24%) e *O. officinalis* (20%), *O. punctata* (27%), *O. rufipogon* (30%), *O. coarctata* (14%), enquanto que em *O. sativa* subsp. *japonica*, o motivo de maior ocorrência foi ATATA/TATAT (19%). Também foram encontrados os motivos AATCT/AGATT (26%) em *O. alta*, CGAGC/GCTCG (16%) em *O. brachyantha*, AAGAA/TTCTT (20%) em *O. granulata*, ATTTA/TAAAT (15%). Maia *et al.*, (2009), observaram os motivos GAAAA/TTTTTC, AAAAT/ATTTT e AAAAC/GTTTT em Brassicas, AAAAT/ATTTT, AAAAG/CTTTT em solanáceas com os mais freqüentes entre os pentâmeros. Também foram observados AAAAT (Hong *et al.*, 2007), e AAAAT/GAAAA (Jiang *et al.*, 2006) como os motivos mais freqüentes encontrados em Rosaceas e Citrus, respectivamente.

Os motivos mais freqüentes encontrados para os hexâmeros foram AGAAAA/TTTTCT em *O. glaberrima* (14%), *O. nivara* (16%), *O. rufipogon* (15%) e *O. sativa* subsp. *indica* (13%). Também foram encontrados AAAAAG/CTTTTT em *O. alta* (17%) e AAAAAT/ATTTTT em *O. coarctata* (13%). Entre os resultados encontrados no trabalho, os motivos AAAAAT em *Arabidopsis* (Zhang *et al.*, 2004) AAAAAG em *Citrus* (Jiang *et al.*, 2006) estiveram entre as maiores ocorrências para hexâmeros.

Discussão Geral

O gênero *Oryza* apresenta um grande número de espécies que podem ser utilizadas como fonte de variabilidade genética de alelos para o melhoramento genético de plantas. O conhecimento dos microssatélites presentes nas frações genômicas atualmente disponíveis das espécies de arroz pode facilitar os estudos moleculares e o uso da seleção assistida ao melhoramento de plantas, pois podem fornecer marcadores detectáveis para genes agronomicamente importantes.

No presente estudo, podemos verificar que a frequência de microssatélites encontrada varia inversamente proporcional ao tamanho do genoma da espécie do gênero *Oryza*. Além disso, as frequências encontradas podem ser tomadas como amostra e servir para estimar a distribuição completa dos microssatélites, considerando o tamanho atual dos bancos de dados e o tamanho estimado dos genomas destas espécies. Os resultados obtidos referentes aos principais motivos de microssatélites como a riqueza de ocorrências de dímeros e trímeros, fornecem uma visão final das classes destes marcadores que devem ser exploradas com mais intensidade, pois estas seqüências podem ser usadas como um ponto de partida, ao menos para acessar a abundância e o benefício dos microssatélites como marcadores moleculares nas espécies estudadas.

O programa SSRLocator mostrou ser eficiente e eficaz para a realização de estudos *in silico*, facilitando a obtenção de marcadores microssatélites, e conseqüentemente reduzindo os custos no desenvolvimento de microssatélites no futuro e beneficiar a genética comparativa e mapeamento físico, estudos de diversidade genética, mapeamento gênico, seleção assistida por marcadores e eventualmente à clonagem posicional de genes úteis em cereais e outras espécies importantes das gramíneas.

CONCLUSÕES

A frequência de microssatélites varia inversamente proporcional ao tamanho do genoma da espécie no gênero *Oryza*.

As espécies de genoma A apresentam as maiores frequências para todos os tipos de microssatélites e totais no gênero *Oryza*.

Os trímeros compostos por citosinas e guaninas apresentam o maior percentual de ocorrência entre as espécies de gênero *Oryza*, sem relação direta com o conteúdo GC da espécie.

REFERÊNCIAS

- Ammiraju JSS, Luo M, Goicochea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert G, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, SanMiguel P, Soderlund C, Jackson S, Wing R A. (2006). The *Oryza* bacterial artificial chromosome library resource: Construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. **Genome Research**. 16: 140-147.
- Akagi H, Yokozeki Y, Inagaki A, Fujimura T. (1996). Microsatellite DNA markers for the rice chromosomes. **Theoretical and Applied Genetics**. 93:1071-1077.
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. **Genetics**. 156(2):847-54.
- Chakravarthi BK, Naravaneni R. (2006). SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L). **African Journal of Biotechnology**. 5(9):684-688.
- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and *Sorghum*. **Plant science**. 160(6):1115-1123.
- Garnica DP, Pinzoni AM, Quesada-Ocampo LM, Bernal AJ, Barreto E, Grunwald NJ, Restrepo S. (2006). Survey and analysis of microsatellites from transcript sequences in *Phytophthora* species: frequency, distribution and potential as markers for the genus. **BMC Genomics**. 7:245.

- Grover A, Sharma PC. (2007). Microsatellite motifs with moderate GC content are clustered around genes on *Arabidopsis thaliana* chromosome 2. **In Silico Biology**. 7:0021.
- Gupta PK, Varshney RK. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. **Euphytica** 113:163-185.
- Hancock JM. (1995). The contribution of slippage-like process to genome evolution. **Journal of Molecular Evolution**. 41:1038-1047.
- Hearne C, Ghosh S, Todd J. (1992). Microsatellites for linkage analysis of genetic traits. **Trends in Genetics**. 8:288-294.
- Hong CP, Piao ZY, Kang TW, Batley J, Yang TJ, Hur YK, Bhak J, Park BS, Edwards D, Lim YP. (2007). Genomic distribution of simple sequence repeats in *Brassica rapa*. **Molecules and cells**. 30;23(3):349-56.
- Iyer RR, Pluciennik A, Rosche WA, Sinden RR, Wells RD. (2000). DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in *Escherichia coli*. **Journal of Biological Chemistry**. 275(3):2174-2184.
- Jiang D, Zhong GY, Hong QB. (2006). Analysis of microsatellites in *Citrus* unigenes. **Acta Genetica Sinica**. 33(4):345-53.
- Kantety RV, La Rota M, Matthews DE, Sorrells ME. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, *sorghum* and wheat. **Plant molecular biology**. 48(5-6):501-10.
- Kashi Y, King D, Soller M. (1997). Simple sequence repeats as a source of quantitative genetic variation. **Trends in Genetics**. 13:74-78.

- Katti MV, Ranjekar PK, Gupta VS. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. **Molecular Biology and Evolution**. 18:1161–1167.
- Kumpatla SP, Mukhopadhyay S. (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. **Genome**. 48(6):985-98.
- La Rota M, Kantety RV, Yu JK, Sorrells ME. (2005). Nonrandom distribution and frequencies of genomic and ESTderived microsatellite markers in rice, wheat and barley. **BMC Genomics**. 6: 23.
- Lagercrantz U, Ellegren H, Andersson L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. **Nucleic Acids Research**. 21:1111–1115.
- Lawson MJ, e Zhang L. (2006). Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. **Genome Biology**. 2(7):R14.
- Londo JP, Chiang Y, Hung K, Chiang T, Schaal BA. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. **Proceedings of the National Academy of Sciences of the USA** 103:9578–9583.
- Maia LC, Palmieri DA, Souza VQ, Koop MM, Carvalho FIF, Costa de Oliveira A. (2008). SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. **International Journal of Plant Genomics**. vol. 2008, Article ID 412696, 9 pages, 2008. doi:10.1155/2008/412696
- Maia LC, Kopp MM, Carvalho FIF, Oliveira AC. (2009). Tandem repeat distribution in gene transcripts of three plant families. **Genetics and Molecular Biology**. No prelo.

Morgante M, Hanafey M, Powell W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**. 30(2):194-200.

Morgante M, Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. **The Plant Journal**. 3(1):175-182.

Mun JH, Kim DJ, Choi HK, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dubez AM, Kiss GB, Roe B, Young ND, Cook DR. (2006). Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. **Genetics**. 172(4):2541-55.

Nelson DL, Cox MM. (2004). **Lehninger Principles of Biochemistry**. 4ed. Nova Iorque: W. H. Freeman.

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P. (2004). Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). **Theoretical and Applied Genetics**. 109 (4):800-805.

Parida SK, Anand RKK, Dalal V, Singh NK, Mohapatra T. (2006). Unigene derived microsatellite markers for the cereal genomes. **Theoretical Applied Genetics**. 112(5):808-17.

Peng JH, Lapitan NL. (2005). Characterization of EST-derived microsatellites in the wheat genome and development of SSR markers. **Functional & integrative genomics**. 5(2):80-96.

Powell W, Machray GC, Provan J. (1996). Polymorphism revealed by simple sequence repeats. **Trends in Plant Science** 7:215–222.

Subramanian S, Mishra RK, Singh L. (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. **Genome Biology**. 4:R13.

- Temnykh S, Declerck G, Lukashova A, Lipovich L, Cartinhour S, Mccouch S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. **Genome Research**. 11(11):1441-1452.
- Thiel T, Michalek W, Varshney W, Graner A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**. 106(3):411-422.
- Thoth G, Gaspari Z, Jurka J. (2000). Microsatellites in different eukaryotic genomes. **Genome Research**. 10(7):967-981.
- Varshney RK, Graner A, Sorrells ME. (2005a). Genomics-assisted breeding for crop improvement. **Trends in Plant Science**. 10(12):621-630.
- Varshney RK, Graner A, Sorrells ME. (2005b). Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**. 23(1):48-55.
- Varshney RK, Thiel T, Stein N, Langridge P and Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cellular and Molecular Biology Letters**. 7:537-546.
- Vaughan DA, Morishima H, Kadowaki K. (2003). Diversity in the *Oryza* genus. **Current Opinion Plant Biology**. 6:139-146.
- Yu JK, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME. (2004). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. **Genome**. 47(5):805-18.
- Yu K, Park SJ, Poysa V, Gepts P. (2000). Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris*). **Journal of Heredity**. 91:429-434.

Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K. (2004). Preference of simple sequence repeats in coding and non coding regions of *Arabidopsis thaliana*. **Bioinformatics**. 20:1081–1086.

Tabela 1.1. Quantidade total de ocorrências e frequências (locos/Mb) dos diferentes tipos de microssatélites, Classe I ($\geq 20\text{pb}$) e Classe II ($\geq 12\text{pb}$ e $< 20\text{pb}$), nos genomas de espécies do gênero *Oryza*.

Espécies	Genoma ^{a,b}	Conteúdo GC (%) ^b	Tamanho do genoma		Mono-		Di-		Tri-		Tetra-		Penta-		Hexa-		Total	
			Sequenciado	Estimado ^b	Qtde ^c	Freq. ^c	Qtde	Freq.	Qtde	Freq.	Qtde	Freq.	Qtde	Freq.	Qtde	Freq.	Qtde	Freq.
<i>O. sativa</i> subsp. <i>indica</i>	AA		339	-	17.194	50,72	31.922	94,17	71.241	210,15	45.018	132,80	14.352	42,34	268.382	791,69	448.109	1.322
<i>O. sativa</i> subsp. <i>japonica</i>	AA	44,0	371	-	16.487	44,44	36.025	97,10	80.496	216,97	48.394	130,44	16.192	43,64	295.505	796,51	493.099	1.329
<i>O. rufipogon</i>	AA	42,1	96	439	4.314	44,94	5.835	60,78	12.501	130,22	8.910	92,81	2.643	27,53	61.863	644,41	96.066	1.001
<i>O. glaberrima</i>	AA	41,7	41	357	2.264	55,22	2.425	59,15	4.836	117,95	4.124	100,59	1.104	26,93	26.303	641,54	41.056	1.001
<i>O. nivara</i>	AA	42,8	72	448	3.151	43,76	3.983	55,32	9.467	131,49	6.601	91,68	1.945	27,01	46.146	640,92	71.293	990
<i>O. punctata</i>	BB, BBCC	41,8	51	425	1.561	30,61	2.356	46,20	4.496	88,16	4.456	87,37	970	19,02	30.475	597,55	44.314	869
<i>O. brachyantha</i>	FF	40,9	46	362	1.629	35,41	3.275	71,20	5.564	120,96	5.813	126,37	1.370	29,78	31.220	678,70	48.871	1.062
<i>O. officinalis</i>	CC	42,9	76	651	1.206	15,87	3.573	47,01	6.869	90,38	5.987	78,78	1.192	15,68	44.288	582,74	63.115	830
<i>O. minuta</i>	BBCC	42,5	102	1.124	2.063	20,23	4.818	47,24	8.915	87,40	8.007	78,50	1.671	16,38	59.352	581,88	84.826	832
<i>O. alta</i>	CCDD	43,0	81	1.008	666	8,22	3.648	45,04	6.322	78,05	6.105	75,37	1.414	17,46	44.508	549,48	62.663	774
<i>O. australiensis</i>	EE	44,4	176	965	1.451	8,24	6.509	36,98	14.230	80,85	10.422	59,22	2.734	15,53	97.612	554,61	132.958	755
<i>O. granulata</i>	GG	45,6	98	882	1.440	14,69	3.369	34,38	7.436	75,88	5.552	56,65	933	9,52	54.563	556,77	73.293	748
<i>O. ridleyi</i>	HHJJ	43,4	137	1.283	1.554	11,34	5.208	38,01	11.551	84,31	10.196	74,42	2.017	14,72	73.822	538,85	104.348	762
<i>O. coarctata</i>	HHKK	40,3	136	-	2.514	18,49	7.481	55,01	11.498	84,54	12.932	95,09	2.433	17,89	79.816	586,88	116.674	858

^aVaughan et al., 2003

^bAmmiraju et al., 2006

^cQuantidade e frequência (locos/Mb) de locos microssatélites, respectivamente.

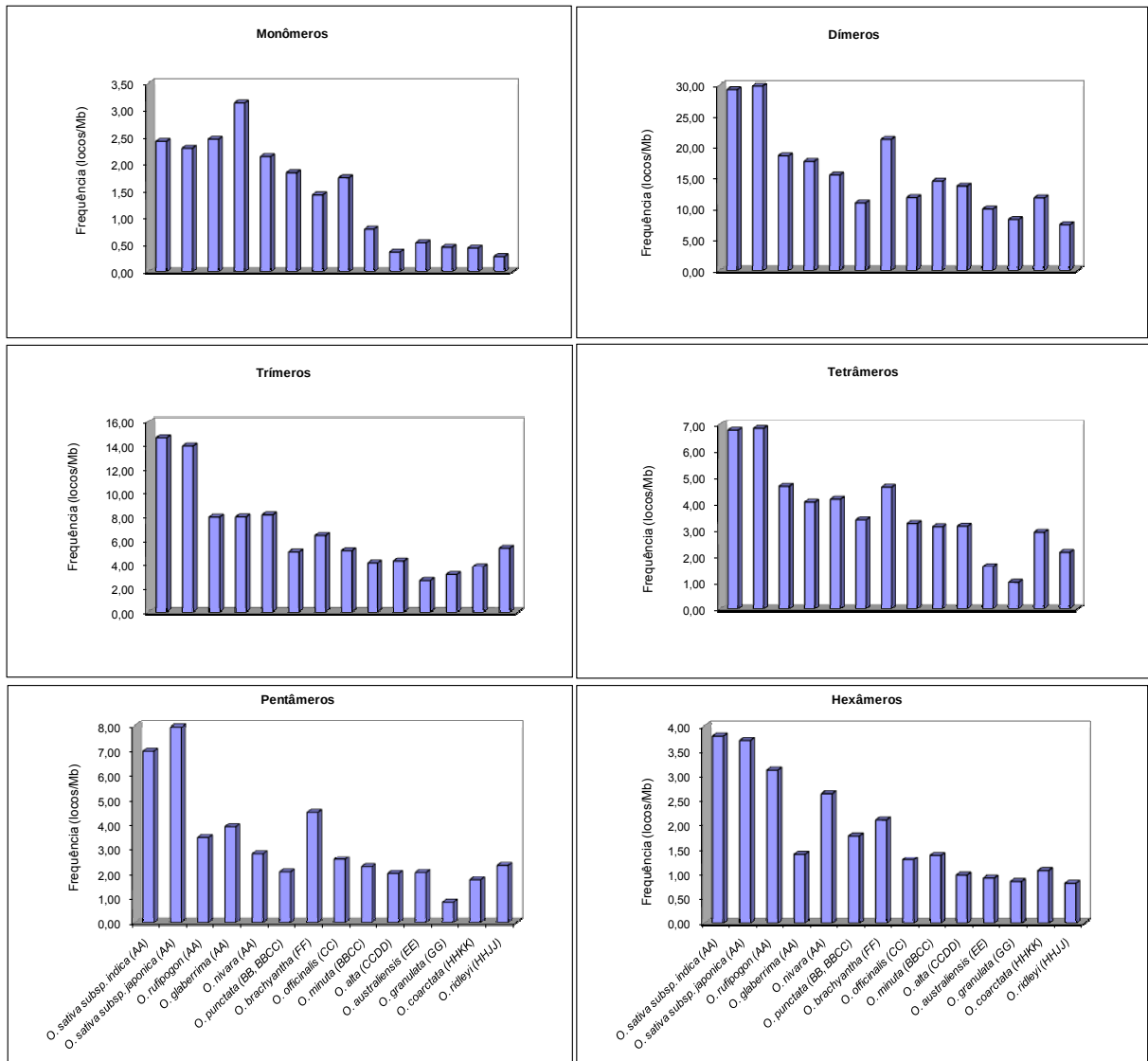


Figura 1.1. Frequências dos diferentes tipos de microssatélites, Classe I (≥ 20 pb) nos genomas de espécies do gênero *Oryza*.

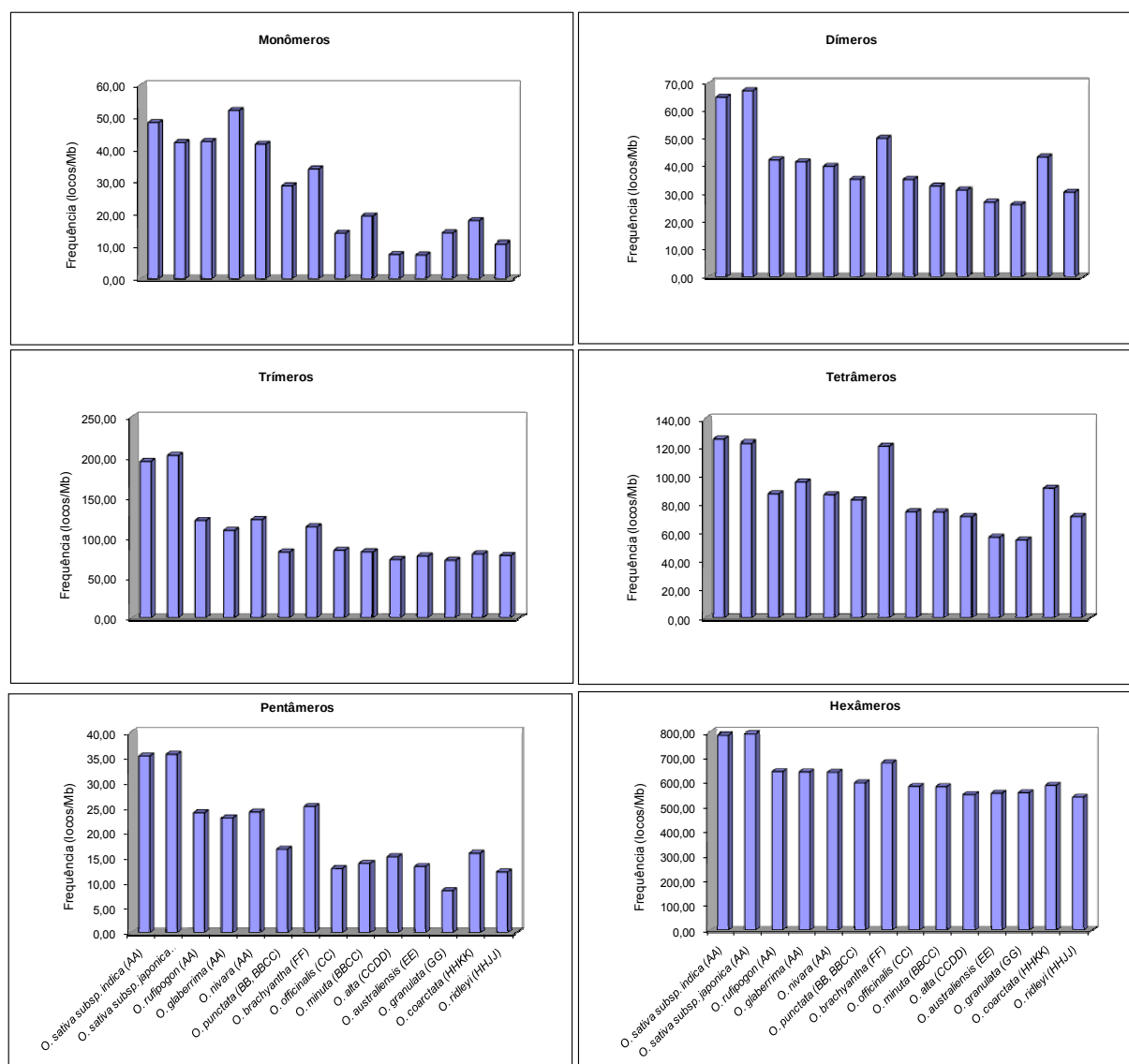


Figura 1.2. Frequências dos diferentes tipos de microssatélites, Classe II (≥ 12 pb e < 20 pb), nos genomas de espécies do gênero *Oryza*.

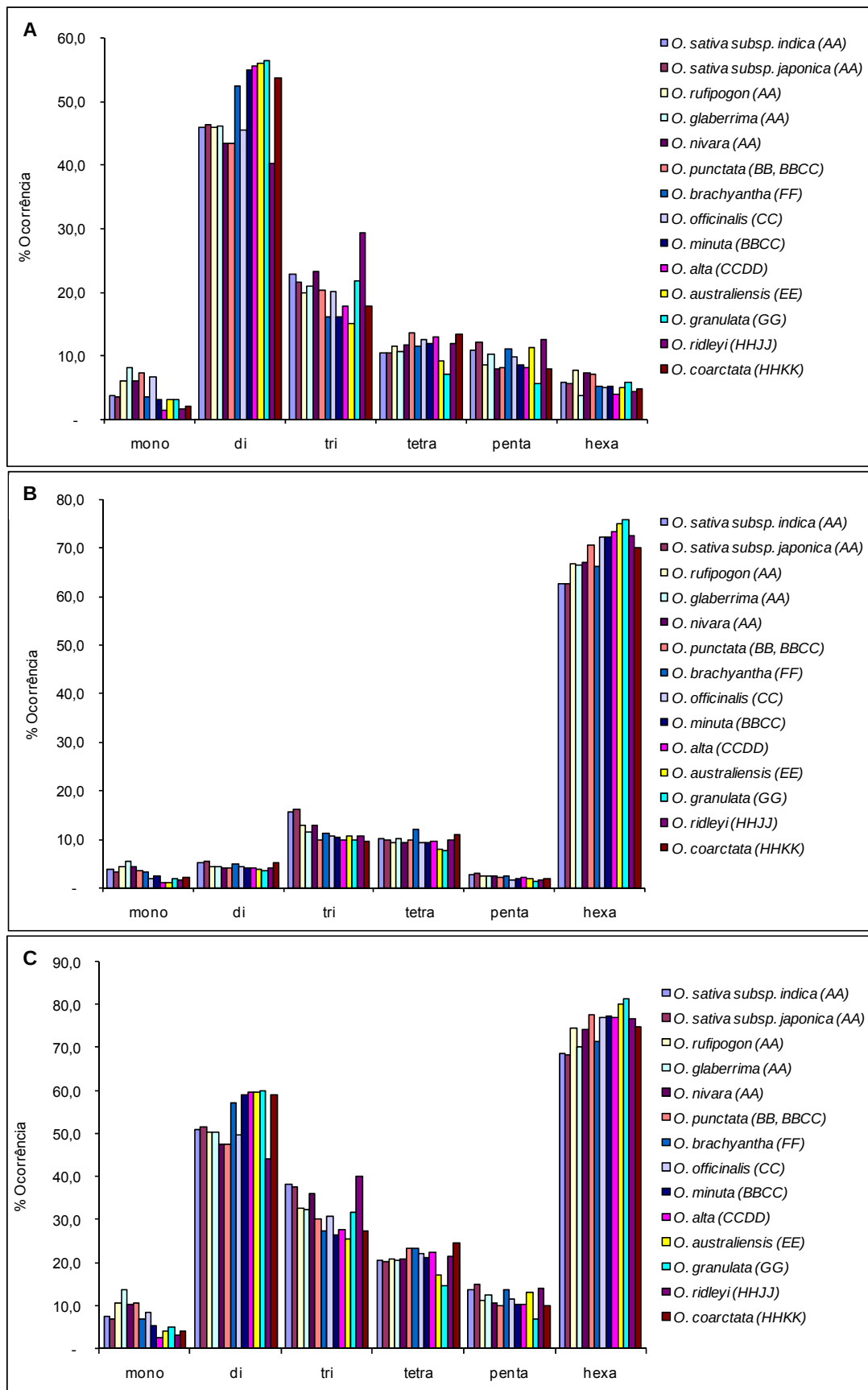


Figura 1.3. Ocorrência percentual dos diferentes tipos de microsatélites, Classe I ($\geq 20\text{pb}$) (A), Classe II ($\geq 12\text{pb}$ e $< 20\text{pb}$) (B) e Classe I e II ($\geq 12\text{pb}$) (C) nos genomas de espécies do gênero *Oryza*.

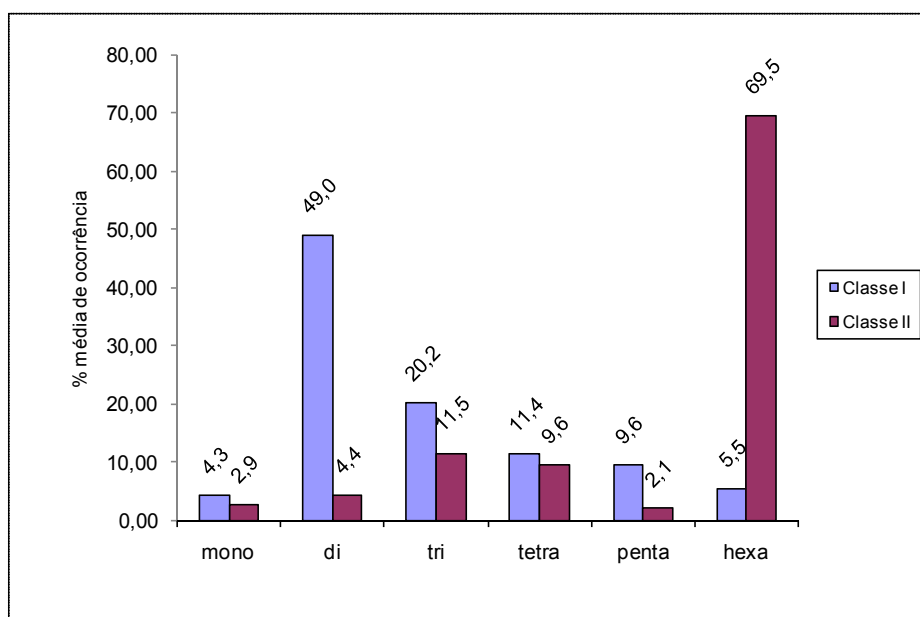


Figura 1.4. Ocorrência percentual média dos diferentes tipos de microsatélites, Classe I ($\geq 20\text{pb}$) e Classe II ($\geq 12\text{pb}$ e $< 20\text{pb}$), nos genomas de espécies do gênero *Oryza*.

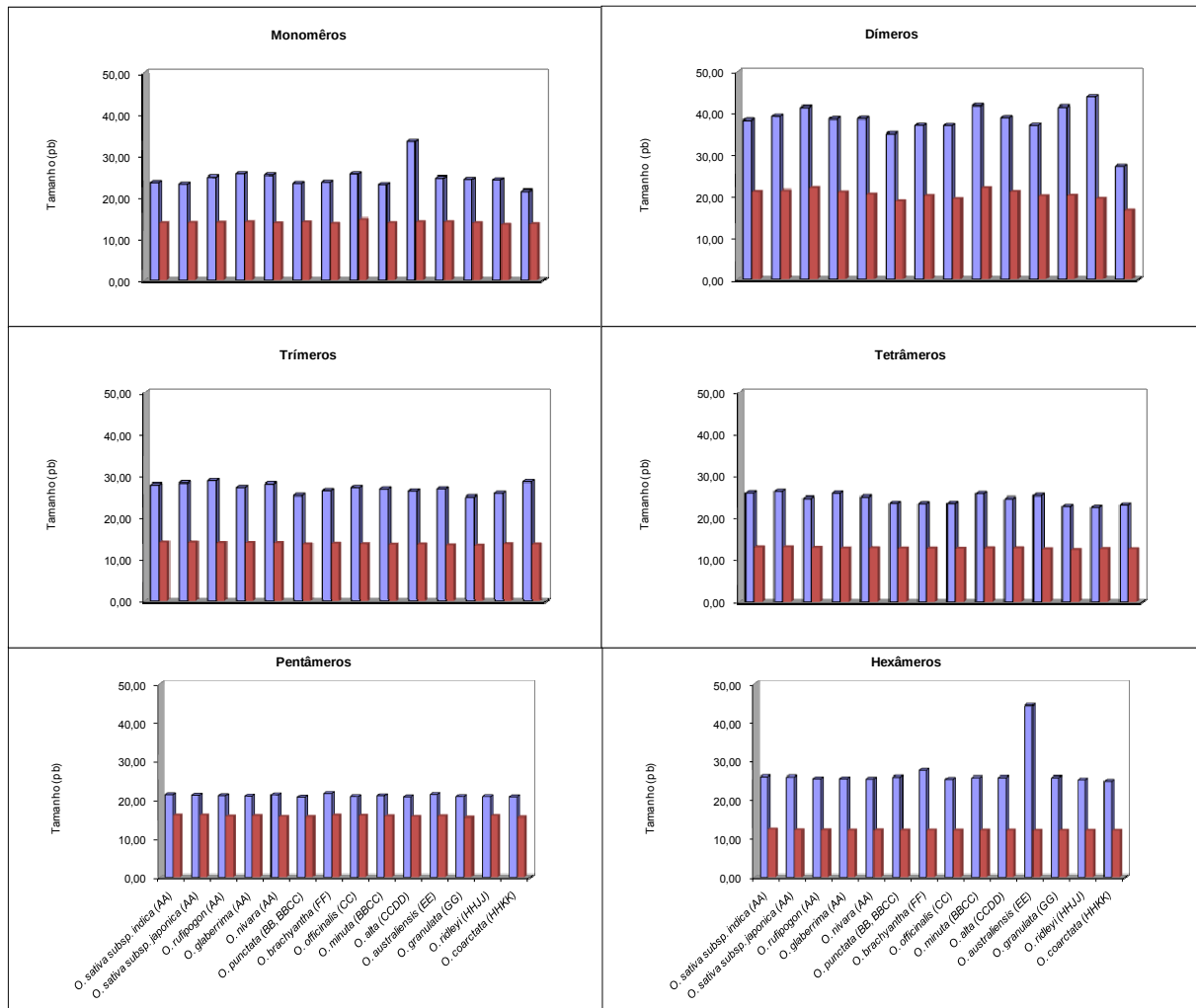


Figura 1.5. Tamanho médio (pb) de locos dos diferentes tipos de microssatélites, Classe I (≥ 20 pb, em azul) e Classe II (≥ 12 pb e < 20 pb, em vermelho) nos genomas de espécies do gênero *Oryza*.

Tabela 1.2. Distribuição da ocorrência de motivos monômeros, dímeros, trímeros e tetrâmeros, pentâmeros e hexâmeros e percentuais em espécies do gênero *Oryza*.

	<i>O. sativa</i> subsp. <i>indica</i> (AA)				<i>O. sativa</i> subsp. <i>japonica</i> (AA)				<i>O. rufipogon</i> (AA)				<i>O. glaberrima</i> (AA)				<i>O. nivara</i> (AA)				<i>O. punctata</i> (BB, BBCC)				<i>O. minuta</i> (BBCC)			
Mono	A/T	12256	0,71		C/G	602	0,71		A/T	3349	0,78		A/T	1771	0,78		A/T	2422	0,77		A/T	1189	0,76		A/T	1365	0,66	
	C/G	4937	0,29		A/T	248	0,29		C/G	964	0,22		C/G	492	0,22		C/G	728	0,23		C/G	371	0,24		C/G	697	0,34	
	Total	17193			850				4313				2263				3150				1560				2062			
Di	AG/CT	7075	0,22		AG/CT	3797	0,34		AT/AT	1292	0,22		AG/CT	582	0,24		GA/TC	911	0,23		AT/AT	629	0,27		AT/AT	1441	0,30	
	AT/AT	6748	0,21		AT/AT	3197	0,29		AG/CT	1234	0,21		AT/AT	505	0,21		AG/CT	890	0,22		TA/TA	545	0,23		TA/TA	1173	0,24	
	GA/TC	6711	0,21		GA/TC	1792	0,16		GA/TC	1214	0,21		GA/TC	498	0,21		AT/AT	722	0,18		AG/CT	415	0,18		AG/CT	750	0,16	
	TA/TA	5778	0,18		TA/TA	1515	0,14		TA/TA	1036	0,18		TA/TA	428	0,18		TA/TA	639	0,16		GA/TC	402	0,17		GA/TC	737	0,15	
	CA/TG	1935	0,06		CA/TG	398	0,04		CA/TG	438	0,08		CA/TG	177	0,07		CA/TG	284	0,07		CA/TG	155	0,07		CA/TG	333	0,07	
	AC/GT	1692	0,05		AC/GT	383	0,03		AC/GT	337	0,06		AC/GT	153	0,06		AC/GT	256	0,06		AC/GT	146	0,06		AC/GT	300	0,06	
	CG/CG	1199	0,04		CG/CG	10	0,00		CG/CG	174	0,03		CG/CG	54	0,02		CG/CG	175	0,04		CG/CG	32	0,01		CG/CG	47	0,01	
	GC/GC	784	0,02		GC/GC	9	0,00		GC/GC	110	0,02		GC/GC	28	0,01		GC/GC	106	0,03		GC/GC	32	0,01		GC/GC	37	0,01	
	Total	31922			11101				5835				2425				3983				2356				4818			
Tri	CGC/GCG	10966	0,15		CGC/GCG	734	0,14		CGC/GCG	1537	0,12		CGC/GCG	503	0,10		CGC/GCG	1235	0,13		CGC/GCG	412	0,09		CGC/GCG	748	0,08	
	CCG/CGG	10353	0,15		CCG/CGG	691	0,13		GCC/GGC	1436	0,11		CCG/CGG	401	0,08		CGC/GCG	1134	0,12		CGC/GCG	361	0,08		CCG/CGG	716	0,08	
	GCC/GGC	9090	0,13		GCC/GGC	450	0,09		CCG/CGG	1416	0,11		GCC/GGC	387	0,08		GCC/GGC	1129	0,12		GCC/GGC	337	0,07		GCC/GGC	671	0,08	
	CTC/GAG	3603	0,05		CTC/GAG	304	0,06		AGG/CCT	728	0,06		AGG/CCT	261	0,05		AGC/GCT	477	0,05		GAA/TTT	284	0,06		GAA/TTT	569	0,06	
	GGA/TCC	3115	0,04		AGG/CCT	239	0,05		AGC/GCT	720	0,06		AGC/GCT	254	0,05		AGG/CCT	463	0,05		CTC/GAG	244	0,05		GCA/TGC	503	0,06	
	AGG/CCT	2663	0,04		GGA/TCC	223	0,04		GGA/TCC	540	0,04		CTC/GAG	252	0,05		CTC/GAG	395	0,04		AGC/GCT	211	0,05		CTC/GAG	461	0,05	
	CGA/TCG	2629	0,04		AAT/ATT	220	0,04		CTC/GAG	510	0,04		GCA/TGC	207	0,04		GCA/TGC	332	0,04		GGA/TCC	183	0,04		AGG/CCT	407	0,05	
	AGC/GCT	2386	0,03		AAG/CTT	211	0,04		GAA/TTT	444	0,04		AAG/CTT	193	0,04		GGA/TCC	323	0,03		GCA/TGC	164	0,04		CAG/CTG	352	0,04	
	GCA/TGC	2043	0,03		GAC/GTC	188	0,04		GCA/TGC	402	0,03		GGA/TCC	191	0,04		GAA/TTT	301	0,03		AGA/TCT	159	0,04		AGC/GCT	343	0,04	
	GAC/GTC	1977	0,03		TAA/TTA	187	0,04		CAG/CTG	376	0,03		GAA/TTT	183	0,04		CAG/CTG	300	0,03		AAG/CTT	155	0,03		GGA/TCC	334	0,04	
	CAG/CTG	1867	0,03		GAA/TTT	186	0,04		CGA/TCG	357	0,03		CAG/CTG	173	0,04		CGA/TCG	277	0,03		AGG/CCT	152	0,03		AGA/TCT	332	0,04	
	GAA/TTT	1754	0,02		ATA/TAT	181	0,04		AAT/ATT	336	0,03		AAT/ATT	144	0,03		AAG/CTT	237	0,03		CAG/CTG	139	0,03		AAG/CTT	283	0,03	
	CCA/TGG	1714	0,02		AGA/TCT	170	0,03		AAG/CTT	334	0,03		CGA/TCG	138	0,03		AGA/TCT	236	0,02		AAT/ATT	131	0,03		CAA/TTG	270	0,03	
	ACG/CGT	1616	0,02		GCA/TGC	151	0,03		AGA/TCT	297	0,02		AGA/TCT	133	0,03		AAT/ATT	229	0,02		ATA/TAT	129	0,03		AAT/ATT	269	0,03	
	AAT/ATT	1496	0,02		CAG/CTG	128	0,02		CCA/TGG	288	0,02		CCA/TGG	129	0,03		CCA/TGG	213	0,02		CCA/TGG	128	0,03		TCA/TGA	252	0,03	
	AGA/TCT	1414	0,02		AGC/GCT	125	0,02		TCA/TGA	259	0,02		CAA/TTG	128	0,03		GAC/GTC	208	0,02		ACC/GGT	125	0,03		CGA/TCG	227	0,03	
	AAG/CTT	1384	0,02		CCA/TGG	111	0,02		ACC/GGT	236	0,02		GAC/GTC	108	0,02		TCA/TGA	207	0,02		TAA/TTA	116	0,03		ATA/TAT	220	0,02	
	TAA/TTA	1292	0,02		CGA/TCG	105	0,02		GAC/GTC	234	0,02		ATA/TAT	106	0,02		CAA/TTG	204	0,02		CGA/TCG	114	0,03		ACC/GGT	212	0,02	
	CAC/GTG	1244	0,02		ACC/GGT	93	0,02		ATA/TAT	223	0,02		TCA/TGA	103	0,02		ATA/TAT	183	0,02		TCA/TGA	108	0,02		ATC/GAT	206	0,02	
	ACC/GGT	1209	0,02		CAC/GTG	74	0,01		TAA/TTA	216	0,02		ACC/GGT	98	0,02		CAC/GTG	175	0,02		ATC/GAT	103	0,02		CAC/GTG	204	0,02	
	ATA/TAT	1090	0,02		ACG/CGT	65	0,01		CAA/TTG	214	0,02		ACG/CGT	98	0,02		ACG/CGT	172	0,02		ATG/CAT	98	0,02		GAC/GTC	175	0,02	
	TCA/TGA	914	0,01		ATC/GAT	58	0,01		ATG/CAT	202	0,02		TAA/TTA	98	0,02		ACC/GGT	166	0,02		CAC/GTG	97	0,02		ATG/CAT	169	0,02	
	CAA/TTG	895	0,01		ATG/CAT	57	0,01		CAC/GTG	198	0,02		CAC/GTG	94	0,02		ATC/GAT	154	0,02		GAC/GTC	93	0,02		CCA/TGG	168	0,02	
	GTA/TAC	875	0,01		CAA/TTG	52	0,01		ACG/CGT	196	0,02		ATG/CAT	86	0,02		ATG/CAT	138	0,01		GTA/TAC	87	0,02		TAA/TTA	160	0,02	
	ATC/GAT	771	0,01		TCA/TGA	43	0,01		GTA/TAC	182	0,01		ATC/GAT	82	0,02		TAA/TTA	138	0,01		CAA/TTG	80	0,02		ACG/CGT	131	0,01	
	CTA/TAG	713	0,01		AAC/GTT	38	0,01		ATC/GAT	156	0,01		CTA/TAG	75	0,02		AAC/GTT	107	0,01		ACT/AGT	65	0,01		ACA/TGT	128	0,01	
	ATG/CAT	698	0,01		ACT/AGT	27	0,01		AAC/GTT	143	0,01		GTA/TAC	74	0,02		GTA/TAC	100	0,01		CTA/TAG	63	0,01		GTA/TAC	119	0,01	
	ACT/AGT	590	0,01		CTA/TAG	20	0,00		CTA/TAG	132	0,01		ACT/AGT	49	0,01		CTA/TAG	84	0,01		ACG/CGT	60	0,01		CTA/TAG	106	0,01	
	AAC/GTT	444	0,01		GTA/TAC	20	0,00		ACT/AGT	107	0,01		ACA/TGT	48	0,01		ACA/TGT	77	0,01		AAC/GTT	57	0,01		ACT/AGT	96	0,01	
	ACA/TGT	436	0,01		ACA/TGT	19	0,00		ACA/TGT	82	0,01		AAC/GTT	40	0,01		ACT/AGT	73	0,01		ACA/TGT	41	0,01		AAC/GTT	84	0,01	
	Total	71241			5170				12501				4836				9467				4496				8915			
Tetra	ATTA/TAAT	2140	0,05		ATAG/CTAT	209	0,08		GAAA/TTTC	429	0,05		ATTA/TAAT	210	0,05		ATTA/TAAT	279	0,04		AAAT/ATTT	213	0,05		AAAT/ATTT	343	0,04	
	AAAT/ATTT	1731	0,04		AGAT/ATCT	186	0,07		ATTA/TAAT	382	0,04		AAAT/ATTT	169	0,04		GAAA/TTTC	255	0,04		GAAA/TTTC	207	0,05		ATTA/TAAT	308	0,04	
	ATGC/GCAT	1119	0,02		CATA/TATG	156	0,06		AAAT/ATTT	351	0,04		GAAA/TTTC	138	0,03		AAAT/ATTT	253	0,04		ATTA/TAAT	177	0,04		GAAA/TTTC	291	0,04	
	Total	45018			2533				8910				4124				6601				4456				8007			
Penta	AAAAT/ATTTT	828	0,06		ATATA/TATAT	173	0,06		AAAAG/CTTTT	266	0,10		AAAAT/ATTTT	81	0,07		AAAAG/CTTTT	153	0,08		AAAAG/CTTTT	71	0,07		AAAAG/CTTTT	77	0,05	
	AAAAG/CTTTT	528	0,04		AAAAG/CTTTT	142	0,05		AAAAT/ATTTT	116	0,04		AAAAG/CTTTT	50	0,05		AAAAT/ATTTT	109	0,06		AAAAT/ATTTT	46	0,05		AAAAT/ATTTT	60	0,04	
	CTCTC/GAGAG	388	0,03		AAAGA/TCITT	95	0,03		AAGAA/TTCTT	90	0,03		CTTCA/TGAAG	42	0,04		GAAAA/TTTTC	64	0,03		GAAAA/TTTTC	27	0,03		GAAAA/TTTTC	52	0,03	
	Total	14352			2924																							

Tabela 1.2. (continuação). Distribuição da ocorrência de motivos monômeros, dímeros, trímeros e tetrameros, pentâmeros e hexâmeros e percentuais em espécies do gênero *Oryza*.

	<i>O. officinalis</i> (CC)			<i>O. alta</i> (CCDD)			<i>O. australiensis</i> (EE)			<i>O. brachyantha</i> (FF)			<i>O. granulata</i> (GG)			<i>O. ridleyi</i> (HHJJ)			<i>O. coarctata</i> (HHKK)		
Mono	A/T	828	0,69	C/G	17	0,57	A/T	802	0,55	C/G	1174	0,72	A/T	823	0,57	A/T	1026	0,66	A/T	1510	0,60
	C/G	377	0,31	A/T	13	0,43	C/G	648	0,45	A/T	454	0,28	C/G	616	0,43	C/G	527	0,34	C/G	1003	0,40
		1205			30			1450			1628			1439			1553			2513	
Di	AT/AT	993	0,28	AT/AT	462	0,42	AT/AT	2234	0,34	AT/AT	942	0,29	AT/AT	856	0,25	AT/AT	1769	0,34	TA/TA	2290	0,31
	TA/TA	887	0,25	TA/TA	255	0,23	TA/TA	1450	0,22	TA/TA	848	0,26	AG/CT	751	0,22	TA/TA	1342	0,26	AT/AT	2033	0,27
	GA/TC	622	0,17	AG/CT	168	0,15	AG/CT	1006	0,15	AG/CT	543	0,17	TA/TA	707	0,21	AG/CT	801	0,15	AG/CT	1192	0,16
	AG/CT	535	0,15	GA/TC	151	0,14	GC/TC	984	0,15	GA/TC	414	0,13	GA/TC	695	0,21	CG/TC	658	0,13	GA/TC	969	0,13
	CA/TG	259	0,07	AC/GT	38	0,03	CA/TG	345	0,05	CA/TG	222	0,07	CA/TG	175	0,05	CA/TG	272	0,05	CA/TG	427	0,06
	AC/GT	219	0,06	CA/TG	33	0,03	AC/GT	334	0,05	AC/GT	220	0,07	AC/GT	145	0,04	AC/GT	231	0,04	AC/GT	377	0,05
	CG/CG	31	0,01				CG/CG	108	0,02	CG/CG	59	0,02	GC/GC	23	0,01	CG/CG	88	0,02	GC/GC	97	0,01
	GC/GC	27	0,01				GC/GC	48	0,01	GC/GC	27	0,01	CG/CG	17	0,01	GC/GC	47	0,01	CG/CG	96	0,01
		3573			1107			6509			3275			3369			5208			7481	
Tri	CGC/GCG	627	0,09	GAA/TTC	33	0,09	AGA/TCT	2767	0,19	CCG/CGG	509	0,09	CCG/CGG	644	0,09	CAA/TTG	735	0,06	CCG/CGG	905	0,08
	CCG/CGG	590	0,09	CCG/CGG	30	0,08	AAG/CTT	1230	0,09	CGC/GCG	474	0,09	GCC/GGC	575	0,08	AAG/CTT	656	0,06	CGC/GCG	898	0,08
	GCC/GGC	570	0,08	CGC/GCG	26	0,07	GAA/TTC	1098	0,08	GCC/GGC	417	0,07	GAC/GTC	554	0,07	ATA/TAT	610	0,05	GCC/GGC	807	0,07
	GAA/TTC	493	0,07	CTC/GAG	23	0,06	GCC/GGC	853	0,06	CTC/GAG	343	0,06	CGA/TCG	509	0,07	CCG/CGG	610	0,05	AAT/ATT	518	0,05
	AGG/CCT	394	0,06	GCC/GGC	23	0,06	CGC/GCG	792	0,06	CAG/CTG	276	0,05	CGC/GCG	431	0,06	CGC/GCG	567	0,05	AGC/GCT	517	0,04
	GCA/TGC	346	0,05	AAG/CTT	19	0,05	CCG/CGG	773	0,05	GCA/TGC	269	0,05	CAA/TTG	372	0,05	GGA/TCC	564	0,05	CTC/GAG	513	0,04
	CTC/GAG	334	0,05	AAT/ATT	19	0,05	GGA/TCC	500	0,04	GGA/TCC	253	0,05	ATC/GAT	307	0,04	TGA/TGA	564	0,05	GCA/TGC	460	0,04
	CAG/CTG	285	0,04	CAG/CTG	18	0,05	CAG/CTG	456	0,03	AGC/GCT	251	0,05	TAA/TTA	302	0,04	GCC/GGC	550	0,05	GAA/TTC	455	0,04
	CAA/TTG	247	0,04	GCA/TGC	17	0,05	CTC/GAG	400	0,03	AGG/CCT	236	0,04	AGC/GCT	264	0,04	AAT/ATT	540	0,05	GGA/TCC	426	0,04
	TCA/TGA	239	0,03	ATA/TAT	16	0,05	CAA/TTG	392	0,03	AAG/CTT	211	0,04	CGA/CTG	258	0,03	GAA/TTC	451	0,04	CGA/TCG	381	0,03
	AGC/GCT	232	0,03	TAA/TTA	16	0,05	AGC/GCT	386	0,03	GAA/TTC	191	0,03	GCA/TGC	258	0,03	CTC/GAG	448	0,04	CAG/CTG	372	0,03
	AGA/TCT	224	0,03	AGA/TCT	15	0,04	AGG/CCT	385	0,03	CGA/TCG	189	0,03	CTC/GAG	250	0,03	AGC/GCT	444	0,04	AAG/CTT	369	0,03
	GGA/TCC	217	0,03	GGA/TCC	13	0,04	AAC/GTT	368	0,03	AAT/ATT	188	0,03	CAG/CTG	238	0,03	GCA/TGC	444	0,04	AGC/CCT	366	0,03
	CGA/TCG	211	0,03	AGC/GCT	10	0,03	TCA/TGA	333	0,02	TCA/TGA	150	0,03	GAA/TTC	223	0,03	TAA/TTA	444	0,04	CCA/TGG	356	0,03
	AAG/CTT	194	0,03	AGG/CCT	8	0,02	GCA/TGC	324	0,02	AGA/TCT	141	0,03	AGG/CCT	220	0,03	AGA/TCT	440	0,04	AGA/TCT	349	0,03
	AAT/ATT	165	0,02	GAC/GTC	8	0,02	CGA/TCG	313	0,02	CCA/TGG	139	0,02	AGA/TCT	218	0,03	AGC/CCT	427	0,04	TAA/TTA	348	0,03
	GAC/GTC	156	0,02	ATC/GAT	7	0,02	ACC/GGT	312	0,02	TAA/TTA	139	0,02	GGA/TCC	217	0,03	CAG/CTG	326	0,03	ATC/GAT	324	0,03
	CAC/GTG	148	0,02	GTA/TAC	7	0,02	AAT/ATT	286	0,02	ACC/GGT	132	0,02	AAG/CTT	214	0,03	ATC/GAT	310	0,03	TCA/TGA	324	0,03
	CCA/TGG	144	0,02	TCA/TGA	7	0,02	ATA/TAT	277	0,02	ACG/CGT	129	0,02	ATA/TAT	214	0,03	GTA/TAC	296	0,03	ATA/TAT	319	0,03
	ATG/CAT	133	0,02	ACT/AGT	6	0,02	CAC/GTG	269	0,02	GAC/GTC	122	0,02	TCA/TGA	180	0,02	ATG/CAT	279	0,02	ATG/CAT	305	0,03
	ACC/GGT	122	0,02	ATG/CAT	6	0,02	GAC/GTC	241	0,02	CAC/GTG	106	0,02	ATG/CAT	152	0,02	CGA/TCG	261	0,02	CAA/TTG	301	0,03
	TAA/TTA	121	0,02	CGA/TCG	6	0,02	CCA/TGG	229	0,02	ATA/TAT	102	0,02	CCA/TGG	139	0,02	ACC/GGT	232	0,02	CAC/GTG	298	0,03
	ACG/CGT	114	0,02	AAC/GTT	5	0,01	ATC/GAT	220	0,02	ATG/CAT	99	0,02	AAT/ATT	129	0,02	GAC/GTC	231	0,02	ACC/GGT	269	0,02
	ATC/GAT	110	0,02	ACC/GGT	3	0,01	ATG/CAT	215	0,02	ATC/GAT	93	0,02	ACA/TGT	109	0,01	CCA/TGG	222	0,02	ACG/CGT	224	0,02
	ATA/TAT	103	0,01	CAA/TTG	3	0,01	ACG/CGT	172	0,01	GTA/TAC	91	0,02	CAC/GTG	92	0,01	CAC/GTG	181	0,02	GAC/GTC	221	0,02
	ACA/TGT	85	0,01	CAC/GTG	3	0,01	ACA/TGT	168	0,01	CAA/TTG	88	0,02	ACC/GGT	80	0,01	ACA/TGT	173	0,01	ACA/TGT	200	0,02
	CTA/TAG	84	0,01	TGT/ACA	3	0,01	TAA/TTA	166	0,01	ACA/TGT	65	0,01	CTA/TAG	78	0,01	ACG/CGT	160	0,01	CTA/TAG	199	0,02
	AAC/GTT	65	0,01	ACG/CGT	2	0,01	GTA/TAC	127	0,01	CTA/TAG	65	0,01	AAC/GTT	77	0,01	AAC/GTT	159	0,01	GTA/TAC	184	0,02
	GTA/TAC	61	0,01	CCA/TGG	2	0,01	ACT/AGT	90	0,01	ACT/AGT	64	0,01	ACT/AGT	76	0,01	CTA/TAG	131	0,01	AAC/GTT	167	0,01
	ACT/AGT	55	0,01	TAG/CTA	1	0,00	CTA/TAG	88	0,01	AAC/GTT	32	0,01	GTA/TAC	56	0,01	ACT/AGT	96	0,01	ACT/AGT	123	0,01
		6869			355			14230			5564			7436			11551			11498	
Tetra	ATTA/TAAT	240	0,04	GAAA/TTTC	22	0,09	ATCC/GGAT	657	0,06	ATTA/TAAT	797	0,14	ATGG/CCAT	417	0,08	AAAT/ATTT	603	0,06	ATTA/TAAT	732	0,06
	AAAT/ATTT	235	0,04	ATAC/GTAT	17	0,07	AAAT/ATTT	517	0,05	AAAT/ATTT	238	0,04	AAAT/ATTT	277	0,05	AATA/TATT	491	0,05	AAAT/ATTT	659	0,05
	GAAA/TTTC	217	0,04	AAAG/CTTT	16	0,06	GAAA/TTTC	483	0,05	ATAA/TTAT	192	0,03	ATTA/TAAT	244	0,04	ATAA/TTAT	407	0,04	AATA/TATT	503	0,04
		5987			257			10422			5813			5552			10196			12931	
Penta	AAAAG/CTTTT	61	0,05	AATCT/AGATT	16	0,10	ATTTA/TAAAT	136	0,05	CGAGC/GCTCG	45	0,03	AAGAA/TTCTT	56	0,06	AAAAT/ATTTT	141	0,07	AAAAG/CTTTT	82	0,03
	GAAAA/TTTTT	49	0,04	CTCTC/GAGAG	9	0,06	GAAAA/TTTTT	124	0,05	AAAAT/ATTTT	37	0,03	GAAAA/TTTTT	46	0,05	AAAAG/CTTTT	116	0,06	AAAAT/ATTTT	77	0,03
	AAAAT/ATTTT	41	0,03	GAAAA/TTTTT	9	0,06	AAGAA/TTCTT	112	0,04	GAAAA/TTTTT	32	0,02	AAAAG/CTTTT	41	0,04	AATAA/TTATT	70	0,03	GAAAA/TTTTT	77	0,03
		1192			162			2734			1370			933			2017			2433	
Hexa	GCGGAA/TTCCGC	740	0,02	AAAAAG/CTTTTT	4	0,05	TCAAAA/TTTTGA	1909	0,02	AAATTT/AAATTT	970	0,03	ACTCGG/CCGAGT	1653	0,03	TAAAAA/TTTTTA	767	0,01	AAAAAT/ATTTTT	764	0,01
	CGAAGA/TCTTCG	488	0,01	CTCTC/GGAGAG	4	0,05	GGCCTA/TAGGCC	1839	0,02	TAAAAA/TTTTTA	281	0,01	CTACTC/GAGTAG	1153	0,02	ACTCGG/CCGAGT	686	0,01	TAAAAA/TTTTTA	650	0,01
	AATATA/TATATT	444	0,01	GAAAAA/TTTTTC	3	0,04	CCTATC/GATAGG	1619	0,02	TATAAA/TTTATA	258	0,01	TCTACA/TGTAGA	888	0,02	GGCCTA/TAGGCC	630	0,01	AAAATA/TATTTT	645	0,01
		44289			80			97613			31221			54564			73283			79817	

2. Elementos *cis-acting* em regiões promotoras de genes *OsNramp*.

Genetics and Molecular Biology (ISSN 1415-4757)

RESUMO

Os genes *Nramp* constituem uma família altamente conservada de proteínas integrais de membrana que estão envolvidas no transporte de ferro em vários organismos, incluindo bactérias, fungos, plantas e animais. Estes genes são amplamente distribuídos em diferentes partes na planta e em todas as famílias de plantas, estando principalmente envolvidos no transporte de cátions divalentes. A região promotora de 1000 pares de bases dos genes *OsNramp* de *Oryza sativa* subsp. *japonica* cv. Nipponbare foi investigada quanto à abundância de elementos *cis-acting*. As seqüências foram analisadas utilizando o programa “Signal Scan Search” do portal “Plant Cis-acting Regulatory DNA Elements” (PLACE) (<http://www.dna.affrc.go.jp/PLACE/>) para a identificação dos diferentes elementos *cis-acting* presentes em cada uma das regiões promotoras. Um dendograma com os genes e as regiões promotoras dos mesmos foi construído pelo método de agrupamento UPGMA no programa CLCBio. Foram detectados 170 diferentes elementos *cis-acting* na região promotora dos genes membros da família *OsNramp*. Um total de 14 elementos são comuns a região promotora dos oito membros da família gênica *OsNramp* de *Oryza sativa* subsp. *japonica*. O elemento CACTFTPPCA1 foi o motivo mais freqüente na região promotora dos genes membros da família *OsNramp* de *Oryza sativa* subsp. *japonica*.

INTRODUÇÃO

A homeostasia das concentrações intracelulares de íons é fundamental na fisiologia de células vivas, fazendo-se necessária a regulação apropriada do fluxo de íons para as células para manter as concentrações baixas dos íons tóxicos e acumular os íons essenciais (Zhu, 2003).

O ferro é um destes íons, sendo elemento essencial para todos os organismos vivos, incluindo as plantas, devido a sua utilização em vários processos fundamentais. Este nutriente desempenha papel importante, uma vez que é parte integrante de diversas enzimas que participam de processos metabólicos como fotossíntese, respiração, fixação do nitrogênio, síntese de clorofila, entre outros (Taiz e Zeiger, 2004). O metabolismo do ferro é um mecanismo complexo sob um balanço homeostático, pois as plantas necessitam de mecanismos para absorver ferro do solo e atender a demanda para o crescimento e desenvolvimento. As plantas também devem evitar a toxicidade pelo excesso de ferro, já que ambas as situações são responsáveis por várias disfunções nutricionais que afetam a sua fisiologia (Ponnamperuma, 1972; Chaney *et al.*, 1972).

Existem várias famílias gênicas envolvidas na homeostase de ferro em plantas. Entre estes, os genes *Nramp* (*Natural Resistance-Associated Macrophage Protein*) constituem uma família altamente conservada de proteínas integrais de membrana que estão envolvidos no transporte de ferro em vários organismos, incluindo bactérias, fungos, plantas e animais (Cellier *et al.*, 2001; Thomine e Schroeder, 2004). Estes genes são amplamente distribuídos em diferentes partes na planta e em todas as famílias de plantas, estando envolvidos principalmente no transporte de cátions divalentes (Curie e Briat, 2003).

O entendimento dos mecanismos envolvidos na regulação da expressão dos genes é essencial para compreender a forma e a função dos sistemas vivos (Aceituno *et al.*, 2008). A expressão gênica é amplamente controlada em nível transcricional, onde as interações entre fatores de transcrição e elementos *cis-acting* em regiões promotoras de um gene desempenham um papel crucial (Brivanlou e Darnell, 2002).

Os elementos *cis-acting* são regiões do DNA que atuam como interruptores moleculares envolvidos na regulação da transcrição de uma rede gênica dinâmica (Strähle e Rastegar, 2007). Embora freqüentemente tenham somente 5 a 20 pb de tamanho, os elementos *cis-acting* são críticos para o entendimento da regulação gênica (Liu *et al.*, 2004). Estes controlam tanto o estado gene ligado e desligado em tipos particulares de células ou regulam os níveis de expressão do gene em vários processos biológicos, para adaptar-se a mudanças das condições fisiológicas e influências ambientais, incluindo respostas ao estresse abiótico, sinais hormonais e de processos do desenvolvimento.

A identificação dos elementos *cis-acting* é um dos maiores e principais desafios biológicos do genoma, pois os elementos reguladores normalmente são curtos, degenerados, escondidos em sucessões muito longas e não codificam proteínas (Wang e Stormo, 2005). Uma das maiores e principal tarefa na decifração de redes de regulação de transcrição é identificar todos os elementos *cis-acting* onde se ligam os fatores de transcrição que são codificados em um genoma, os quais eventualmente fornecerão a informação necessária para a construção de modelos de redes gênicas de regulação da transcrição (Li *et al.*, 2002).

Em regiões promotoras de vários genes, os elementos *cis-acting* têm sido definidos e investigados para isolar fatores de transcrição específicos, e buscar

elucidar como ocorre a regulação dos genes. O elemento *cis-acting* mais conhecido é o motivo TATAbox, que é reconhecido por proteínas que se ligam no sítio TATA. O TATAbox é uma sequência rica em T/A geralmente localizada a 25–35 pares de bases do sítio de iniciação da transcrição na região promotora dos genes (Burley e Roeder, 1996). Aproximadamente 85% dos genes de plantas seqüenciados contêm esta sequência, que tem papel crucial na transcrição, pois serve como um sítio de ligação do complexo de iniciação da transcrição. Também foram identificados vários outros elementos *cis-acting* envolvidos em uma série de processos metabólicos fisiológicos, como motivos específicos respondendo a hormônios, açúcares, condições ambientais e outros fatores (Lois *et al.*, 1989; Mundy *et al.*, 1990; Ballas *et al.*, 1993; Grierson *et al.*, 1994; Yamaguchi-Shinozaki e Shinozaki, 1994; Shinshi *et al.*, 1995; Savino *et al.*, 1997; Dunn *et al.*, 1998; Cercos *et al.*, 1999; Menke *et al.*, 1999; Shinozaki e Yamaguchi *et al.*, 2000; Trindade, 2003, Shinozaki *et al.*, 2003; Yamaguchi-Shinozaki e Shinozaki, 2005; Yamaguchi-Shinozaki e Shinozaki, 2006, Won *et al.*, 2009).

Em arroz, a rede gênica responsável pela resposta à toxidez por ferro envolve várias famílias de genes, contendo várias cópias de um mesmo gene (homólogo) para cada um dessas famílias. As regiões promotoras de diferentes genes podem conter diferenças estruturais quanto aos diferentes tipos de elementos *cis-acting* ativadores da transcrição gênica encontrados nestas regiões.

O conhecimento dos elementos *cis-acting* presentes na região promotora dos genes *OsNramp* poderá contribuir para a compreensão dos sistemas reguladores da expressão destes genes e dos demais envolvidos na homeostasia de ferro e de outros metais em arroz. Neste sentido, o objetivo do trabalho foi identificar os

elementos *cis-acting* presentes na região upstream dos genes *OsNramp* envolvidos na homeostasia de ferro em arroz.

MATERIAL E MÉTODOS

Seqüências de DNA dos genes da família *OsNramp*, com base em estudo *in silico* conduzido por Gross *et al.* (2005) foram baixadas dos bancos de dados referentes ao genoma do arroz (*Oryza sativa* subsp *japonica* cv. Nipponbare) do National Center for Biotechnology Information (NCBI). Baseados nas informações de localização dos genes no NCBI foram cortadas e baixadas seqüências de DNA de 1000 pb da região promotora destes genes.

A região promotora dos genes *OsNramp* foi investigada quanto à abundância de elementos *cis-acting*. As seqüências foram analisadas utilizando o programa “Signal Scan Search” do portal “Plant Cis-acting Regulatory DNA Elements” (PLACE) (<http://www.dna.affrc.go.jp/PLACE/>) para a identificação dos diferentes elementos *cis-acting* presentes em cada uma das regiões promotoras. O portal PLACE é um banco de dados de motivos de seqüências nucleotídicas encontrados em elementos *cis-acting* regulatórios de DNA (Higo *et al.*, 1999; Prestridge *et al.*, 1991). O banco de dados do PLACE contém uma breve definição e descrição de cada um dos motivos, e literatura relevante com números de ID do PubMed e números de acesso dos bancos de dados de seqüências nucleotídicas DDBJ/EMBL/GenBank também estão incluídas.

Após a submissão no PLACE, os resultados foram analisados e os elementos *cis-acting* em comum entre todos os locos da família gênica extraídos. Os elementos mais comuns foram agrupados em tabelas (quantidades) e em figuras, representando a posicionamento dos elementos *cis-acting* mais freqüentes na região

promotora dos genes da família *OsNramp*. Um dendograma com os genes e as regiões promotoras dos mesmos foi construído pelo método de agrupamento UPGMA (Unweighted Pair Group Method with Arithmetic mean). Com o Sequence Logo integrante do programa CLC Main Work Bench 5 (CLC Bio Inc.), foram feitos alinhamentos do elemento *cis-acting* mais comum em cada loco e quando todos estes estavam agrupados, verificando a frequência de conservação de cada nucleotídeo dentro do motivo.

RESULTADOS E DISCUSSÃO

Na Tabela 2.1 são apresentados os cromossomos, localização, tamanho e número de exons dos genes *OsNramp* (*Natural resistance-associated macrophage protein*) em *Oryza sativa* subsp. *japonica*. Os oito genes *OsNramp* estão presentes em apenas 6 dos 12 cromossomos que compõem o genoma do arroz. Destes oito locos, quatro genes *OsNramp* estão nos cromossomos 1 (*OsNramp6*), cromossomo 2 (*OsNramp4*) e cromossomo 3 (*OsNramp2* e *OsNramp8*), outros três genes localizados no cromossomo 6 (*OsNramp3*) e cromossomo 7 (*OsNramp1*, *OsNramp5*) e o gene *OsNramp7* no cromossomo 12. O gene *OsNramp6* apresentou o maior tamanho e 10 exons, enquanto que os genes *OsNramp4*, *OsNramp3*, *OsNramp5* e *OsNramp1* apresentaram entre 13 e 14 exons.

Os resultados obtidos para a família *OsNramp* indicaram a ocorrência de 170 diferentes elementos *cis-acting* descritos no portal PLACE (Figura 2.1), revelando a grande diversidade de elementos reguladores na região promotora, dados estes confirmados por Hatorangan *et al.* (2009) que analisaram regiões promotoras de genes transportadores de fosfato em arroz, e identificaram 153 elementos *cis-acting*.

No entanto, somente 14 destes elementos foram comuns a todos os oito membros da família gênica *OsNramp*.

Na Tabela 2.2 são apresentadas informações referentes aos elementos *cis-acting* mais comuns encontrados em cada loco. Todos os genes revelaram uma grande quantidade de elementos *cis-acting*, variando de 200 a 275. Os genes *OsNramp1* e *OsNramp7* foram os dois locos com as maiores ocorrências totais, apresentando 275 e 251 elementos, respectivamente, seguidos por *OsNramp5* (224), *OsNramp6* (220) e *OsNramp2* (220), *OsNramp3* (209), *OsNramp4* (206) e *OsNramp8* (200). O número de elementos *cis-acting* por loco juntamente com a ocorrência total dos diferentes elementos mostra que existe uma grande quantidade e diversidade de elementos *cis-acting* presentes nas regiões promotoras de uma família gênica *OsNramp*.

Os genes que apresentaram o maior número de diferentes elementos *cis-acting* em sua região promotora, não foram exatamente os mesmos que revelaram as maiores ocorrências totais, com exceção para os genes *OsNramp1* e *OsNramp7*. Logo, de forma geral, o número médio de repetições de cada elemento *cis-acting* foi maior naqueles locos que apresentaram o menor número de diferentes elementos, como no caso dos genes *OsNramp6*, *OsNramp8* e *OsNramp2*, os quais apresentaram o maior número médio de repetições.

Analisando mais detalhadamente os resultados por loco, em *OsNramp1* foram encontrados 96 diferentes elementos *cis-acting* e para os locos *OsNramp2*, *OsNramp3*, *OsNramp4*, *OsNramp5*, *OsNramp6*, *OsNramp7* e *OsNramp8*: 65, 77, 68, 71, 68, 83 e 65 elementos, respectivamente. Além destes dados, ainda são apresentados os nomes dos elementos *cis-acting* mais comuns encontrados em cada loco, com o número de repetições em parênteses. Outro fato interessante é a

grande quantidade de alguns elementos *cis-acting* nas regiões promotoras de alguns locos *OsNramp*, destacando o elemento ROOTMOTIFTABOX1 (ATATT), que apresentou 27 cópias na região promotora do gene *OsNramp2* e o elemento CGCGBOXAT (VCGCGB) com 20 cópias na região promotora do gene *OsNramp8*. O elemento ROOTMOTIFTAPOX1 foi previamente identificado em regiões promotoras do gene rolD de *Agrobacterium rhizogenes* (Elmayan e Tepfer, 1995) e do gene TTG2 de *Arabidopsis thaliana* (Johnson *et al.*, 2002) e parece estar associado à expressão em raiz. No entanto, as proteínas *NRAMP* parecem não estar envolvidas na absorção de ferro do solo, mas provavelmente são ativas em nível sub-celular (Belouchi *et al.*, 1997; Gross *et al.*, 2003; Thomine, 2000; Narayanan *et al.*, 2007).

A grande quantidade total, assim como a maior quantidade de alguns de elementos *cis-acting*, e a maior presença de alguns elementos específicos em alguns locos mostram a complexidade do controle genético desta família gênica. Existem alguns trabalhos relacionados a genes envolvidos no transporte intracelular de ferro, mostrando que as cópias de uma família gênica têm diferentes funções, assim, a expressão é temporal e espacial nas diferentes partes das plantas (Belouchi *et al.*, 1997; Curie *et al.*, 2000; Gross *et al.*, 2003; Narayanan *et al.*, 2007). Experimentos de expressão mostraram que, enquanto *OsNramp1* é expresso principalmente nas raízes e *OsNramp2* em folhas, *OsNramp3* é expresso em níveis semelhantes em raízes e folhas (Belouchi *et al.* 1997). Narayanan *et al.*, (2007), verificaram que os genes *OsNramp2*, *OsNramp4* e *OsNramp7* são altamente expressos em folhas de genótipos de arroz durante o estágio de enchimento do grão. As diferenças na expressão dos diferentes genes desta família podem ser

explicadas pela presença de elementos *cis-acting* específicos, ou também pela quantidade dos mesmos na região promotora.

A Figura 2.2 mostra a ocorrência total dos elementos *cis-acting*, considerando apenas os 14 elementos comuns os oito membros da família gênica *OsNramp*. Seis genes dos oito membros da família *OsNramp* apresentaram mais de 100 cópias de elementos *cis-acting* destas 14 categorias, exceto, *OsNramp3* e *OsNramp8* com valores totais menores. Os genes *OsNramp2* e *OsNramp7* apresentaram as maiores quantidades dos elementos comuns.

Os resultados encontrados podem sugerir uma maior importância responsiva destes elementos *cis-acting* aos fatores de transcrição específicos nos seus respectivos locos. Para Schwechheimer e Bevan (1998), a especificidade da expressão gênica depende dos elementos *cis-acting* presentes na região promotora, e as interações de fatores de transcrição específicos com os mesmos. Além disso, estas curtas seqüências de DNA determinam o tempo, a localização e o nível da expressão gênica (Pennacchio e Rubin, 2001), o que comprova a grande abundância dos motivos presentes dentro e entre os locos.

Na Tabela 2.3, são apresentados os 14 elementos *cis-acting* comuns a todos os locos da família gênica *OsNramp*, juntamente com informações referentes ao número total de cópias de cada elemento, porcentagem, seqüência, mecanismos em que estão envolvidos estes elementos *cis-acting* e as referências de trabalhos destes elementos comuns a todos os locos da região promotora da família gênica *OsNramp*. Na Figura 2.3 são apresentados graficamente os 14 elementos *cis-acting* comuns em todos os locos da família gênica *OsNramp*, mostrando as diferenças de ocorrência em cada loco.

O motivo CACTFTPPCA1 foi o mais comum entre os elementos *cis-acting* comuns, apresentando um total de 100 cópias e de 12,55%, considerando todos os diferentes locos dos genes *OsNramp*. A composição dos motivos mais comuns varia de quatro a seis nucleotídeos, sendo que a metade dos elementos *cis-acting* (sete) é composta por quatro nucleotídeos, enquanto que outros quatro motivos apresentam cinco, e mais quatro apresentam seis nucleotídeos, respectivamente. Dentre os 14 elementos *cis-acting* comuns, CACTFTPPCA1 (YACT), DOFCOREZM (AAAG), CAATBOX1 (CAAT), ARR1AT (AGGT) foram os que apresentaram as maiores ocorrências nos genes *OsNramp*, e estão entre os motivos compostos por apenas 4 nucleotídeos, enquanto que os elementos ACGTATERD1 (ACGT), CCAATBOX1 (CCAAT), GT1GMSCAM4 (GAAAAA) e IBOXCORE (GATAA), que revelaram as menores ocorrências são formados por cinco ou seis nucleotídeos, exceto ACGTATERD1 com apenas quatro. O elemento CACTFTPPCA1 é composto pelo motivo YACT, onde Y é uma pirimidina, podendo ser neste caso, citosina (C) ou timina (T).

A grande quantidade de informações referentes principalmente, ao número e os mecanismos em que estes diferentes elementos *cis-acting* estão envolvidos revelam e também comprovam a complexidade do controle da expressão gênica dos diferentes genes desta família. Estes elementos *cis-acting* estão envolvidos em uma ampla gama de processos metabólicos fisiológicos, relacionados na Tabela 2.3, mostrando a importância dos mesmos na sinalização celular. Para alguns autores, genes com expressão similar podem partilhar partes em comum nos seus mecanismos regulatórios (Vilo *et al.*, 2001), assim como quando a expressão dos mesmos difere, podem existir diferenças entre seus mecanismos regulatórios.

Ainda não existem trabalhos com os elementos mais comuns em relação ao ferro, exceto o estudo de Savino *et al.* (1997), que identificou uma sequência regulatória dependente de ferro (IDRS) na região promotora do gene ferritina de milho *ZmFer1*. Este elemento é responsável pela repressão da transcrição de *ZmFer1* sob baixas condições de suprimento de ferro. A literatura também traz mais alguns resultados interessantes que podem ser relacionados com este trabalho. O elemento *cis-acting* CACTFTPPCA1 foi também um dos motivos mais comuns encontrados nas regiões promotoras dos genes transportadores de fosfato (Hatorangan *et al.*, 2009). A grande quantidade deste elemento presente nas regiões promotoras dos genes envolvidos com o metabolismo de ferro poderia ser explicada pelo fato do elemento CACTFTPPCA1 estar relacionado com o processo fotossintético, pois este nutriente desempenha papel importante uma vez que é parte integrante de diversas enzimas que participam de vários processos metabólicos como fotossíntese, respiração, fixação do nitrogênio, síntese de clorofila, entre outros (Taiz e Zeiger, 2004).

Os elementos ARR1AT e CAATBOX1 estariam envolvidos em respostas hormonais (Hatorangan *et al.*, 2009). ARR1AT (5'-NGATT-3') é um elemento *cis-acting* presente em genes reguladores de resposta (ARRs) a estímulos externos (Sakai *et al.*, 2000), e está envolvido na resposta a citocinina ARR1, o qual é encontrado em *Arabidopsis* (Oka *et al.*, 2002; Sakai *et al.*, 2000), sendo também encontrado na região promotora do gene hemoglobina não simbiótica em arroz (NSHB) (Ross *et al.*, 2004). CAATBOX1 (5'-CAAT-3') é uma sequência promotora do gene legumina de ervilha responsável para a atividade específica de tecido (Shirsat *et al.*, 1989). Legumina é o principal componente que é acumulado nos vacúolos de sementes (Stoger *et al.*, 2001). Elementos MYCCONSensusAT

possuem o domínio MYC, e estão relacionados à resposta a estresse abiótico, sendo que foram encontrados em promotores de genes que respondem à seca e ao frio (Abe *et al.*, 1997; Chinnusamy *et al.*, 2003). Outro elemento *cis-acting* é WRKY71OS, que é reconhecido por WRKY71, um repressor transcricional da rota sinalizadora da giberelina (Zhang *et al.*, 2004).

Diversos estudos mostram que os fatores de transcrição ou proteínas geralmente atuam como fatores regulatórios, ligando-se a elementos *cis-acting* específicos no RNA ou DNA, localizados perto do sítio de começo da transcrição, promovendo ou reprimindo a expressão gênica pelo aumento ou diminuição da taxa de produção de mRNA via polimerização da RNA polimerase (Hentze e Kühn, 1996; Davidson *et al.*, 2002; Wray *et al.*, 2003).

A importância dos elementos *cis-acting* é bem conhecida na regulação da expressão dos genes. Vários trabalhos identificaram e analisaram elementos *cis-acting* envolvidos em uma série de processos metabólicos fisiológicos, como motivos específicos respondendo a hormônios, açúcares, condições ambientais e outros fatores (Trindade, 2003). São conhecidos vários motivos responsivos a hormônios como Auxin Responsive Elements – *AuxRE* (Ballas *et al.*, 1993), ácido abscísico – ABRE (Mundy *et al.*, 1990), etileno – ERE (Shinshi *et al.*, 1995), giberelinas – GARE (Cercos *et al.*, 1999) e ácido jasmônico – JERE (Menke *et al.*, 1999). O elemento SURE é uma sequência específica responsiva a sacarose (Grierson *et al.*, 1994). Entre os elementos *cis-acting* responsivos a sinais ambientais, vários elementos *cis-acting* reativos a luz – LRE, (Lois *et al.*, 1989), baixa temperatura – LTRE (Dunn *et al.*, 1998) e patógenos – *ELRE* (Lois *et al.*, 1989) também foram caracterizados. Os elementos ABRE e DRE/CRT são os dois principais elementos *cis-acting* com a expressão gênica induzida por estresse abiótico (Shinozaki e Yamaguchi-Shinozaki

et al., 2000; Shinozaki *et al.*, 2003; Yamaguchi-Shinozaki e Shinozaki, 2006). Análises das regiões promotoras do gene RESPONSIVE TO DEHYDRATION^{29A}/COLD-REGULATED⁷⁸/LOW-TEMPERATURE-INDUCED⁷⁸ (RD^{29A}/COR⁷⁸/LTI⁷⁸) induzido por seca, alta salinidade e frio, mostraram a presença destes dois elementos *cis-acting*, *ABRE* (dependente de ABA) e *DRE/CRT* (independente de ABA), os quais são elementos *cis-acting* que se expressam em resposta a estresse osmótico e frio, respectivamente (Yamaguchi-Shinozaki e Shinozaki, 1994). *DRE* também é encontrado em regiões promotoras de vários genes induzidos por seca e frio (Shinozaki e Yamaguchi-Shinozaki *et al.* 2000; Shinozaki *et al.* 2003). Além dos elementos *cis-acting* citados, outro trabalho de Yamaguchi-Shinozaki e Shinozaki, (2005), relaciona muitos outros envolvidos na expressão gênica responsiva a estresse osmótico e por frio.

Baseado no posicionamento na região promotora do elemento mais comum CACTFTPPCA1 (Figura 2.4), é observada uma maior abundância deste elemento na região promotora dos genes *OsNramp7* (17), *OsNramp2* e *OsNramp3* (16) e *OsNramp5* (15). No loco *OsNramp7*, o elemento CACTFTPPCA1 está mais distribuído entre as regiões proximal (-1pb a -400pb), intermediária (-400 pb a -600 pb) e distal (-600 pb a -1000 pb), enquanto ficaram mais localizados na porção proximal e distal da região promotora dos genes *OsNramp2*, *OsNramp3* e *OsNramp5*. Nos demais genes *OsNramp*, o elemento CACTFTPPCA1 apresentou um maior número de cópias na parte distal da região promotora, exceto em *OsNramp6*, onde os motivos ficaram localizados na região intermediária. Também foram estudados os posicionamentos dos elementos *cis-acting* DOFCOREZM e CAATBOX1 na região promotora dos genes *OsNramp* (Figura 2.5 e 2.6). As distribuições dos motivos dentro da região promotora dos locos *OsNramp* foram

muito variáveis entre os locos, sendo observadas concentrações dos motivos em regiões específicas da região upstream analisada para alguns locos.

Isto pode representar diferenças, como uma maior eficiência e especificidade na expressão de cada gene modulada por diferentes sinais externos recebidos pela espécie, em respostas a diferentes fatores de transcrição (proteínas) ativados por sinais de estresse abiótico e/ou biótico (Oliveira *et al.*, 2008; Bervald *et al.*, 2008). Além disso, apesar de os elementos mais comuns estarem presentes e conservados, estes não seguem um padrão de distribuição entre os locos estudados, alterando suas posições dentro da região upstream, fato que pode ser justificado pela presença de seqüências repetitivas, alterando a posição do elemento *cis-acting*. É importante também ressaltar que os elementos mais comuns não são descritos como sendo envolvidos na expressão de genes envolvidos no metabolismo de ferro, no entanto, isso pode ser justificado pela falta de trabalhos realizados com este enfoque.

Na Figura 2.7 é apresentado o alinhamento das ocorrências do elemento *cis-acting* CACTFTPPCA1 nas regiões promotoras dos genes da família *OsNramp*. Pode ser observado que as bases ACT são conservadas em todos os elementos *cis-acting* alinhados em um mesmo loco e também nos diferentes locos, exceto no loco *OsNramp3*, em que a base adenina não ocorre em 100% dos casos, mas ainda é mais conservada. Na posição da citosina ou timina, a maior ocorrência foi de timina nos locos *OsNramp1*, *OsNramp2*, *OsNramp3*, *OsNramp4*, *OsNramp5*, enquanto que nos locos *OsNramp6*, *OsNramp7* e *OsNramp8*, as citosinas foram mais conservadas. Estas diferenças podem ser devido a pequenas mutações ocorridas nos diferentes locos do elemento *cis-acting*.

Com base na análise filogenética da região promotora dos genes *OsNramp* através do método UPGMA conforme Figura 2.8, constata-se a formação de dois grupos, um formado por *OsNramp3*, *OsNramp2* e *OsNramp4* e o segundo por *OsNramp6*, *OsNramp1*, *OsNramp5*, *OsNramp7* e *OsNramp8*. Quando a análise é feita com base nos genes (Figura 2.9), também são formados dois grupos, porém com quatro genes cada, o primeiro grupo compostos pelos locos *OsNramp3*, *OsNramp4* que são mantidos neste grupo em comparação com o agrupamento baseado na região promotora dos genes, acrescidos dos locos *OsNramp1* e *OsNramp5*. Os locos *OsNramp7*, *OsNramp8* mantêm-se no segundo grupo, juntamente com os locos *OsNramp2* e *OsNramp6*. Considerando os dois dendogramas, os genes mais conservados quanto as suas seqüências na sua região estrutural e promotora são *OsNramp3* e *OsNramp4* para um grupo e *OsNramp6* e *OsNramp8* para o outro grupo. Os locos presentes no mesmo cromossomo, *OsNramp6* e *OsNramp8* (cromossomo 3) e *OsNramp1* e *OsNramp5* (cromossomo 7) tem comportamentos distintos. Enquanto *OsNramp2* e *OsNramp8* não permanecem no mesmo agrupamento quando comparadas as regiões codificantes e promotoras, sugerindo que a sua origem não seja uma duplicação recente ou estejam sob uma baixa pressão de seleção. Entretanto, os locos *OsNramp1* e *OsNramp5* estão localizados no mesmo cromossomo (1) e permanecem juntos no agrupamento tanto na comparação da região promotora como na comparação de regiões codificantes. Isto sugere que esta seja a duplicação mais recente dos genes *OsNramp* no genoma.

As diferenças na região promotora desta família gênica podem ser comprovadas pelo fato de que os genes *OsNramp* são amplamente distribuídos em diferentes partes na planta, estando envolvidos principalmente no transporte de

cátions divalentes (Curie e Briat, 2003). As proteínas NRAMP parecem não estar envolvidas na absorção de ferro do solo, mas provavelmente são ativas em nível sub-celular. Existem alguns trabalhos relacionados a genes envolvidos no transporte intracelular de ferro. A proteína AtNRAMP de *Arabidopsis* por exemplo, é expressa no tecido vascular de raízes, caules e folhas em condições de deficiência de ferro (Thomine, 2000, 2004). Esta proteína é um transportador de Fe^{2+} localizado no tonoplasto, estando possivelmente envolvido na remobilização do ferro que se encontra no vacúolo (Thomine *et al.*, 2003). Lanquar *et al.* (2005) mostraram que AtNRAMP3 e também AtNRAMP4 realizam a mobilização do estoque de ferro do vacúolo, o que, sob baixa concentração de ferro, passa a ser essencial para a germinação das sementes. A germinação de sementes em um mutante duplo *nramp3 nramp4* é atrasada sob baixa nutrição com ferro, e completamente normal com o suprimento de ferro. Belouchi *et al.* (1997) e Narayanan *et al.* (2007), em experimentos de expressão verificaram principalmente diferenças de localização, estádios de crescimento e o nível da expressão gênica dos genes *OsNramp* em plantas de arroz. As diferenças na expressão dos diferentes genes desta família podem ser explicadas pela presença de elementos *cis-acting* específicos, ou também pela quantidade dos mesmos na região promotora.

Os resultados indicam que existem diferenças entre o número e conteúdo de elementos *cis-acting* nos diferentes locos de uma família gênica *OsNramp*. No entanto, são necessários maiores estudos para verificar se estas diferenças estão associadas a padrões diferenciais na expressão dos genes *OsNramp* e/ou até mesmo para a identificação de elementos *cis-acting* responsivos ao ferro.

Os mecanismos de tolerância à toxidez por ferro são extremamente complexos, sendo que a compreensão dos mesmos, tanto em nível fisiológico

quanto molecular contribuirá, sobremaneira, com os programas de melhoramento de cultivares visando à obtenção de genótipos tolerantes. Conhecendo estes elementos *cis-acting* podem ser desenvolvidos trabalhos para aumentar ou diminuir a quantidade destes trechos nas regiões promotoras destes genes, aumentando ou diminuindo desta forma a expressão dos mesmos.

CONCLUSÕES

Um total de 170 diferentes elementos *cis-acting* ocorrem na região promotora dos genes membros da família *OsNramp* de *Oryza sativa* subsp. *japonica*.

Um total de 14 elementos são comuns a região promotora dos oito membros da família gênica *OsNramp* de *Oryza sativa* subsp. *japonica*.

O elemento CACTFTPPCA1 é o motivo mais freqüente na região promotora dos genes membros da família *OsNramp* de *Oryza sativa* subsp. *japonica*.

REFERÊNCIAS

- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, et al. (1997). Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. **Plant Cell**. 9:1859–1868.
- Aceituno FF, Moseyko N, Rhee SY and Gutiérrez RA. (2008). The rules of gene expression in plants: Organ identity and gene body methylation are key factors for regulation of gene expression in Arabidopsis thaliana. **BMC Genomics**. 9:438.
- Ballas N, Wong LM, Theologis A. (1993). Identification of the auxin-responsive element, AuxRE, in the primary indoleacetic acid-inducible gene of pea (*Pisum sativum*). **Journal of Molecular Biology**. 233:580-596.
- Belouchi A, Kwan T, Gros P. (1997). Cloning and characterization of the OsNramp family from *Oryza sativa*, a new family of membrane proteins possibly implicated in the transport of metal ions. **Plant Molecular Biology**. 33:1085–1092.
- Bervald C, Maia LC, Vital C, Bresolin A, Almeida AM, Carvalho, FIF, Oliveira AC. (2008). **Cis-acting elements in promoter regions of the Nramp genes family in rice genome**. In: 54º Congresso Brasileiro de Genética, 2008, Salvador-BA. Anais do 54º Congresso Brasileiro de Genética, 2008.
- Brivanlou A, Darnell J. (2002). Signal transduction and the control of gene expression. **Science**. 295:813–818.
- Burley SK, Roeder RG. (1996) Biochemistry and structural biology of transcription factor IID (TFIID). **Annual Review of Biochemical**. 65:769–799.
- Cellier M, Privé G, Belouchi A, Kwan T, Rodrigues V, Chia W and Gros P (1995). *Nramp* defines a family of membrane proteins. **Proceedings of National Academy of Science USA**. 92:10089-10093.

- Cellier MFM, Bergevin I, Boyer E, Richer E. (2001). Polyphyletic origins of bacterial *Nramp* gene transporters. *Trends Genetics*. 17:365–370.
- Cercos M, Gomes-Cadenas A, Ho THD. (1999). Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: cis-and trans-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. **Plant Journal**. 19:107-118.
- Chaney RL, Brown JC, Tiffin LO. (1972). Obligatory reduction of ferric chelates in iron uptake by soybean. *Plant Physiology*. 50:208-213.
- Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, et al. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. **Gene Development**. 17:1043–1054.
- Curie C, Alonso JM, Le Jean M et al. (2000). Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. **Biochemical Journal**. 347:749-755.
- Curie C, Briat JF. (2003). Iron transport and signaling in plants. *Annual Review of Plant Biology*. 54:183–206.
- Davidson EH, Rast JP, Oliveri P. et al. (2002). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. **Developmental Biology**. 246:162–190.
- Dunn MA, White AJ, Vural S, Hughes MA. (1998). Identification of promoter elements in a low-temperature responsive gene (bl4.9) from barley (*Hordeum vulgare* L.). **Plant Molecular Biology**. 38:551-564.
- Elmayan T, Tepfer M. (1995). Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter. **Transgenic Research**. 4:388–396.

- Grierson C, Du JS, Zabala MT, Beggs K, Smith C, Holdsworth M, Bevan M. (1994). Separate *cis* sequences and *trans* factors direct metabolic and developmental regulation of potato storage protein gene. **Plant Journal**. 5:815-826.
- Gross J; Stein RJ; Fett-Neto AG, Fett JP. (2003). Iron homeostasis related genes in rice. **Genetics and Molecular Biology**. 26:477-497.
- Hatorangan Mr, Sentausa E, Wijaya GY. (2009). In Silico Identification of Cis-Regulatory Elements of Phosphate Transporter Genes in Rice (*Oryza sativa* L.). **Journal of Crop Science and Biotechnology**. 12(1):25-30.
- Hentze MW, Kühn LC. (1996). Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. **Proceedings of the National Academy of Sciences**. 93:8175–8182.
- Higo K, Ugawa Y, Iwamoto M, KORENAGA T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database. **Nucleic Acids Research**. 27:297-300.
- Johnson CS, Kolevski B, Smyth DR (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. **Plant Cell**. 14:1359–1375.
- Lanquar V, Lelievre F, Bolte S, Hames C, Alcon C, Neumann D, Vansuyt G, Curie C, Schröder A, Kramer U, Barbie-Brygoo H, Thomine S. (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. **European Molecular Biology Organization**. 24:4041-4151.
- Li B, Perabekam S, Liu G, Yin M, Song S, Larson A. (2002). Experimental and bioinformatics comparison of gene expression between T cells from TIL of liver cancer and T cells from UniGene. **Journal of Gastroenterology**. 37(4):275-282.

- Liu Y, Liu XS, Wei L, Altman RB, Batzoglou S. (2004). **Eukaryotic regulatory element conservation** analysis and identification using comparative genomics. **Genome Research**. 14:451–458.
- Lois R, Dietrich A, Hahlbrok K, Schulz W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. **Embo Journal**. 8:1641-1648.
- Menke FL, Champion A, Kijne JW, Memelink J. (1999). A novel jasmonate- and elicitor-responsive element in the piriwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. **Embo Journal**. 18:4455-4463.
- Mundy J, Yamaguchi-Shinozaki K, Chua NH. (1990). Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. **Proceedings of the National Academy of Sciences**. 87:1406-1410.
- Narayanan NN, Vasconcelos WM, Grusak AM. (2007). Expression profiling of *Oryza sativa* metal homeostasis genes in different rice cultivars using a cDNA macroarray. **Plant Physiology and Biochemistry**. 45:277-286.
- Oka A, Sakai H, Iwakoshi S. (2002). His-Asp phosphorelay signal transduction in higher plants: receptors and response regulators for cytokinin signaling in *Arabidopsis thaliana*. **Genes & Genetic Systems**. 77:383–391.
- Oliveira AC; Bervald CMP, Vital CE, Bresolin APS, Santos LS, Maia LC, Almeida AM, Carvalho FIF. (2008). **Abundance of *cis-acting* elements in promoter regions of the *NRAMP* gene family**. In: The 5th International Crop Science Congress, 2008, Jeju - Korea. The 5th International Crop Science Congress, 2008.
- Pennacchio L, Rubin E. (2001). Genomic strategies to identify mammalian regulatory sequences. **Nature Review Genetics**. 2:100-109.

Ponnamperuma FN. (1972). The chemistry of submerged soils. *Advanced Agronomy*. 24:29-96.

Prestridge DS. (1991). SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. **Computer Applications In The Biosciences**. 2:203-206.

Ross EJH, Stone JM, Elowsky CG, Arredondo-Peter R, Klucas RV, Sarath G. (2004) Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor, ARR1. **Journal of Experimental Botany**. 55:1721-1731.

Sakai H, Aoyama T, Oka A. (2000). Arabidopsis ARR1 and ARR2 response regulators operate as transcriptional activators. **Plant Journal**. 24:703-711.

Savino G, Briat JF, Lobreaux S. (1997). Inhibition of the iron-induced *ZmFer1* maize ferritin gene expression by antioxidants and serine/threonine phosphatase inhibitors. **Journal of Biological Chemistry**. 272:33319–33326.

Savino G, Briat JF, Lobreaux S. (1997). Inhibition of the iron-induced *ZmFer1* maize ferritin gene expression by antioxidants and serine/threonine phosphatase inhibitors. **Journal of Biological Chemistry**. 272:33319–33326.

Schwechheimer C, Bevan M. (1998). The regulation of transcription factor activity in plants. **Trends in Plant Science**. 3:378–383.

Shinozaki K, Yamaguchi-Shinozaki K, Seki M. (2003). Regulatory network of gene expression in the drought and cold stress responses. **Current Opinion of Plant Biology**. 6:410–417.

Shinozaki K, Yamaguchi-Shinozaki K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. **Current Opinion of Plant Biology**. 3:217–23

- Shinshi H, Usume S, Ohme-Takagi M. (1995). Identification of an ethylene-responsive region in the promoter of a tobacco class I chitinase gene. **Plant Molecular Biology**. 27:923-932.
- Shirsat AH, Wilford N, Croy RRD, Boulter D. (1989). Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. **Molecular Genetics and Genomics**. 15:326–331.
- Stoger E, Parker M, Christou P, Casey R. (2001). Pea legumin overexpressed in wheat endosperm assembles into an ordered paracrystalline matrix. **Plant Physiology**. 125:1732-1742.
- Strähle U, Rastegar S. (2008). Conserved non-coding sequences and transcriptional regulation. *Brain Research Bulletin*. 75:225–230.
- Taiz L, Zeiger E. (2004). *Fisiologia vegetal*. Porto Alegre: Artmed, p.449-484.
- Thomine S, Lelièvre F, Debarbieux E et al. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. **Plant Journal**. 34:685-695.
- Thomine S, Schroeder JI. (2004). Plant Metal Transporters with Homology to Proteins of the NRAMP Family. *The NRAMP Family*. Eureka.com and Kluwer Academic / Plenum Publishers.
- Thomine S, Wang R, Ward JM. et al. (2000). Cadmium and iron transport by members of a plant transporter gene family in Arabidopsis with homology to NRAMP genes. **Proceedings of the National Academy of Sciences**. 97:4991–4996.
- Trindade LM. (2003). **Potato life-cycle: from genes to promoters to cis-acting elements**. PhD thesis. Wageningen University, The Netherlands. 135f.

- Vilo J, Kivinen K. (2001). Regulatory sequence analysis: application to the interpretation of gene expression. **European Neuropsychopharmacology**. 11:399-411.
- Wang T, Stormo GD (2005). Identifying the conserved network of cis-regulatory sites of a eukaryotic genome. **Proceedings of National Academy of Science USA**. 102(48):17400-17405.
- Won S., Yong-Ju L, Ha-Yeon L, Yoon-Kyung H, Misuk C, and Hyung-Taeg C. (2009). Cis-Element- and Transcriptome-Based Screening of Root Hair-Specific Genes and Their Functional Characterization in Arabidopsis. **Plant Physiology**, 150:1459–1473.
- Wray CJ, Mammen JM, Hershko DD, Hasselgren PO. (2003). Sepsis upregulates the gene expression of multiple ubiquitin ligases in skeletal muscle. **International Journal of Biochemistry & Cell Biology** 35:698–705.
- Yamaguchi-Shinozaki K, Shinozaki K. (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. **The Plant Cell**. 6:251–264.
- Yamaguchi-Shinozaki K, Shinozaki K. (2005). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. **Trends in Plant Science**. 10:88–94.
- Yamaguchi-Shinozaki K, Shinozaki K. (2006). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. **Trends in Plant Science**. 10:88–94
- Zhang JZ, Creelman RA, Zhu JK. (2004). From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops. **Plant Physiology**. 135:615–621.

Zhu JK. (2003) Regulation of ion homeostasis under salt stress. *Current Opinion in Plant Biology*. 6(5): 441-445.

Tabela 2.1. Cromossomos de *Oryza sativa* subsp. *japonica* e os genes *OsNramp* envolvidos na homeostasia de ferro em plantas.

Cromossomo	Gene	Localização (pb)	Tamanho (pb)	Nº exons
1	<i>OsNramp6</i>	c17461080-17450132	10948	10
2	<i>OsNramp4</i>	c1662644-1658576	4068	13
3	<i>OsNramp2</i>	c5626641-5622651	3990	4
3	<i>OsNramp8</i>	c22627467-22623629	3838	5
6	<i>OsNramp3</i>	c27681501-27675896	5605	14
7	<i>OsNramp5</i>	c8911062-8903800	7262	13
7	<i>OsNramp1</i>	8998211-9002982	4771	13
12	<i>OsNramp7</i>	c24157351-24153493	3858	4

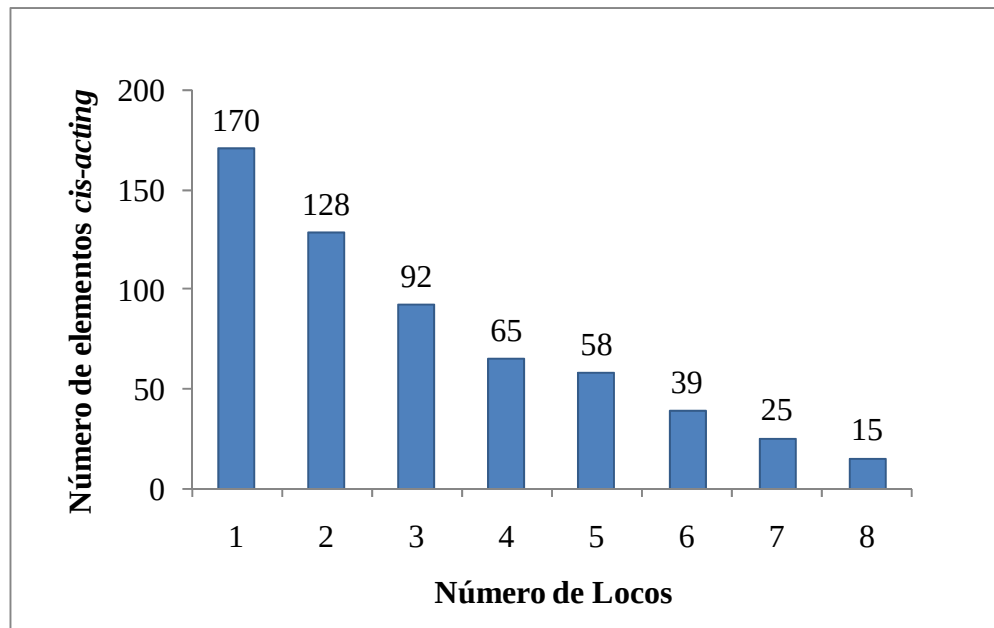


Figura 2.1. Número de elementos *cis-acting* encontrados nos genes *OsNramp* no genoma de *Oryza sativa* subsp *japonica* cv. Nipponbare.

Tabela 2.2. Genes *OsNramp* do genoma de *Oryza sativa* subsp *japonica* cv. Nipponbare, com os respectivos cromossomos, quantidade de elementos *cis-acting* em cada loco, número de diferentes elementos *cis-acting* por loco, elementos *cis-acting* mais comuns encontrados em cada loco, com o respectivo número de ocorrências no loco.

Loco	Cromossomo	Quantidade de elemento <i>cis-acting</i>	Diferentes elementos <i>cis-acting</i> por loco	Número médio de cada elementos <i>cis-acting</i> no loco	Elementos <i>cis-acting</i> mais comuns encontrados por loco (número de ocorrências no loco)
<i>OsNramp1</i>	7	275	96	2,86	EBOXBNNAPA (14), MYCCONSUSAT (14), CACTFTPPCA1 (10), GTGANTG10 (9), WRKY71OS (9), ACGTATERD1 (8), CAATBOX1 (7), DOFCOREZM (7), SORLIP2AT (6), GATABOX (5), GT1CONSUSAT (5)
<i>OsNramp2</i>	3	220	65	3,38	ROOTMOTIFTAPOX1 (27), CACTFTPPCA1 (16), POLLEN1LELAT52 (15), CAATBOX1 (12), TATABOX5 (11), CARGCW8GAT (8), DOFCOREZM (8), POLASIG2 (8), GT1CONSUSAT (7), BOXIINTPATPB (6)
<i>OsNramp3</i>	6	209	78	2,68	CACTFTPPCA1 (16), DOFCOREZM (12), GTGANTG10 (10), ARR1AT (8), CAREOSREP1 (8), CARGCW8GAT (8), ROOTMOTIFTAPOX1 (8), WRKY71OS (7), GATABOX (7), WBOXATNPR1 (5), WBOXHVIS01 (5), WBOXNTERF3 (5)
<i>OsNramp4</i>	2	206	68	3,03	ARR1AT (15), CAATBOX1 (14), GTGANTG10 (12), MYCCONSUSAT (12), CURECORECR (10), CACTFTPPCA1 (9), CGACGOSAMY3 (9), DOFCOREZM (6), CBFHV (5), POLLEN1LELAT52 (5), ROOTMOTIFTAPOX1 (5), WRKY71OS (5)
<i>OsNramp5</i>	7	224	71	3,15	CACTFTPPCA1 (15), DOFCOREZM (15), CURECORECR (12), ARR1AT (10), CAATBOX1 (10), EBOXBNNAPA (8), MYCCONSUSAT (8), WRKY71OS (8), POLLEN1LELAT52 (7), ROOTMOTIFTAPOX1 (7), GT1CONSUSAT (6), NODCON2GM (5), OSE2ROOTNODULE (5)
<i>OsNramp6</i>	1	220	65	3,38	POLLEN1LELAT52 (15), CACTFTPPCA1 (16), CAATBOX1 (12), TATABOX5 (11), CARGCW8GAT (8), DOFCOREZM (8), POLASIG2 (8), GT1CONSUSAT (7), SEF4MOTIFGM7S (7), BOXIINTPATPB (6), ROOTMOTIFTAPOX1 (5)
<i>OsNramp7</i>	12	251	82	3,06	CACTFTPPCA1 (17), DOFCOREZM (17), GT1CONSUSAT (11), GTGANTG10 (11), CAATBOX1 (10), SURECOREATSULTR11 (10), ARR1AT (9), WRKY71OS (9), CURECORECR (8), EBOXBNNAPA (6), GATABOX (6), MYCCONSUSAT (6), WBOXNTERF3 (6), CCAATBOX1 (5), MYBCORE (5), NODCON2GM (5), OSE2ROOTNODULE (5), POLLEN1LELAT52 (5)
<i>OsNramp8</i>	3	200	64	3,13	CGCGBOXAT (20), ARR1AT (12), GT1CONSUSAT (8), GTGANTG10 (8), ROOTMOTIFTAPOX1 (8), CGACGOSAMY3 (7), SURECOREATSULTR11 (7), CACTFTPPCA1 (6), EBOXBNNAPA (6), MYCCONSUSAT (6), SORLIP1AT (6)

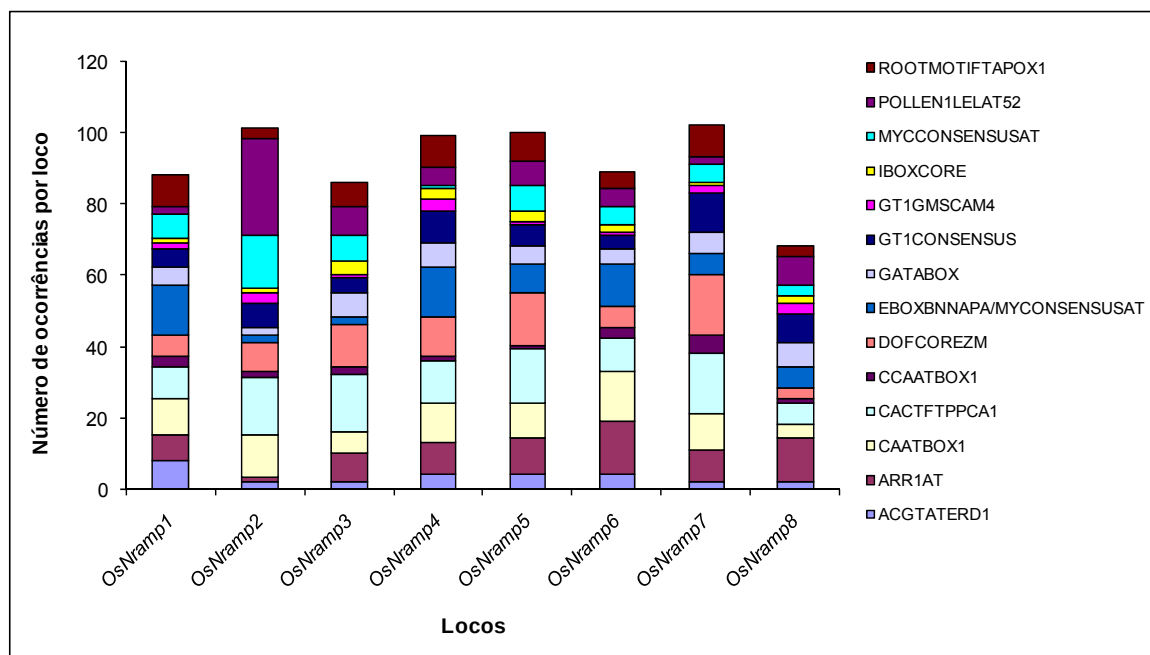


Figura 2.2. Ocorrência total dos elementos *cis-acting* em comum por gene da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare.

Tabela 2.3. Relação dos elementos *cis-acting* comuns encontrados na região promotora dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare.

Elementos <i>cis-acting</i>	Número total de cada elemento <i>cis-acting</i> considerando todos os locos	Porcentagem de cada elemento <i>cis-acting</i> (%)	Seqüência de elementos <i>cis-acting</i>	Mecanismos em que os elementos <i>cis-acting</i> estão envolvidos	Referências
ACGTATERD1	28	3,51	ACGT	ACGT; etiolation; erd	Simpson SD et al. (2003). Plant J. 33: 259-270. PubMed: 12535340
ARR1AT	71	8,91	NGATT	ARR1; Response regulator	Sakai H, Aoyama T, Oka A. (2000). Plant J. 24: 703-711 (2000). PubMed: 11135105 and PubMed: 15258171
CAATBOX1	77	9,66	CAAT	CAAT; legA; seed	Shirsat A. et al. (1989). Mol Gen Genet 215:326-331 (1989) PubMed: 2710102;
CACTFTPPCA1	100	12,55	YACT	mesohpyll; CACT	Gowik U. et al. (2004). Plant Cell, 16: 1077-1090. PubMed: 15100398.
CCAATBOX1	18	2,26	CCAAT	HSE (Heat shock element); CCAAT box	Rieping M., Schoffl F(1992). Mol Gen Genet. 231: 226-232. PubMed: 1736093, Haralampidis K (2002) Plant Physiol. 129: 1138-1149 (2002). PubMed: 12114568 and Wenkel S et al. (2006) Plant Cell. 18:2971-2984
DOFCOREZM	78	9,79	AAAG	Dof; C4PEPC; CyPPDK; PEPC; C4; leaf; shoot	Yanagisawa S, Schmidt RJ (1999) Plant J 17:209-214. PubMed: 10074718 and Yanagisawa S (2000) Plant J 21:281-288. PubMed: 10758479
EBOXBNNAPA/ MYCCONSUSUSAT	64	8,03	CANNTG	EBOXBNNAPA (napA; storage protein; ABRE; E-box; seed) MYCCONSUSUSAT (MYC; rd22BP1; ABA; leaf; seed; stress; CBF3; cold; CBF/DREB1; ICE1; RRE)	EBOXBNNAPA (Stalberg K et al. (1996) Planta 199:515-519. PubMed: 8818291 and Hartmann U et al. (2005) Plant Mol Biol. 57: 155-171. PubMed: 15821875). MYCCONSUSUSAT (Abe H. et al. Plant Cell 15: 63-78 (2003) PubMed: 12509522; Chinnusamy V. et al. Genes Dev. 17: 1043-1054 (2003) PubMed: 12672693; Chinnusamy V, Schumaker K, Zhu JK. J Exp Bot. 55: 225-236 (2004). PubMed: 14673035; Oh SJ. et al. Plant Physiology 138: 341)
GATABOX	43	5,39	GATA	ASF-2; GATA box; Cab; chlorophyll a/b binding protein	Lam E, Chua NH. Plant Cell 1:1147-1156 (1989). PubMed: 2535536; Gilmartin PM. et al. Plant Cell 2:369-378. (1990) PubMed: 2152164; Benfey PN, Chua NH. Plant Cell 2:369-378 (1990) PubMed: 2152164; Gidoni D. et al. Mol Gen Genet 215: 337-344 (1989). PubMed:
GT1CONSUSUS	54	6,77	GRWAAW	GT-1; light; TATA; TFIIA; TBP; HR; SAR; TMV; leaf; shoot	Terzaghi WB et al. Annu Rev Plant Physiol Plant Mol Biol 46:445-474(1995); Villain P, Mache R, Zhou DX. J Biol Chem 271:32593-32598 (1996) PubMed: 8955086; Le Gourrierec J, Li YF, Zhou DX. Plant J 18:663-668 (1999)PubMed: 10417717; Buchel AS et al. Plant
GT1GMSCAM4	16	2,01	GAAAAA	GT-1 box	Park HC et al. Plant Physiol. 135: 2150-2161 (2004). PubMed: 15310827
IBOXCORE	17	2,13	GATAA	I box; I-box; rbcS; light regulation	Terzaghi WB, Cashmore AR. Annu Rev Plant Physiol Plant Mol Biol 46:445-474 (1995)
POLLEN1LELAT52	50	6,27	AGAAA	pollen; lat52; endo-beta-mannnanase; MAN	Bate N, Twell D. Plant Mol Biol. 37:859-869 (1998) PubMed: 9678581; Filichkin AS. et al. Plant Physiol. 134 1080-1087 (2004) PubMed: 14976239
ROOTMOTIFTAPO11	64	8,03	ATATT	root; rold	Elmayan T, Tepfer M. Transgenic Res 4:388-396 (1995) PubMed: 7581519
WRKY71OS	53	6,65	TGAC	WRKY; GA; MYB; W box; TGAC; PR proteins	Zhang ZL. et al. Plant Physiol. 134:1500-1513(2004) PubMed: 15047897; Xie Z. et al. Plant Physiol. 137:176-189 (2005) PubMed: 15618416; Eulgem T. et al., EMBO J. 18:4689-4699 (1999) PubMed: 10469648; Eulgem T. et al. Trends Plant Sci. 5:199-206. (2000) PubMed: 10785665

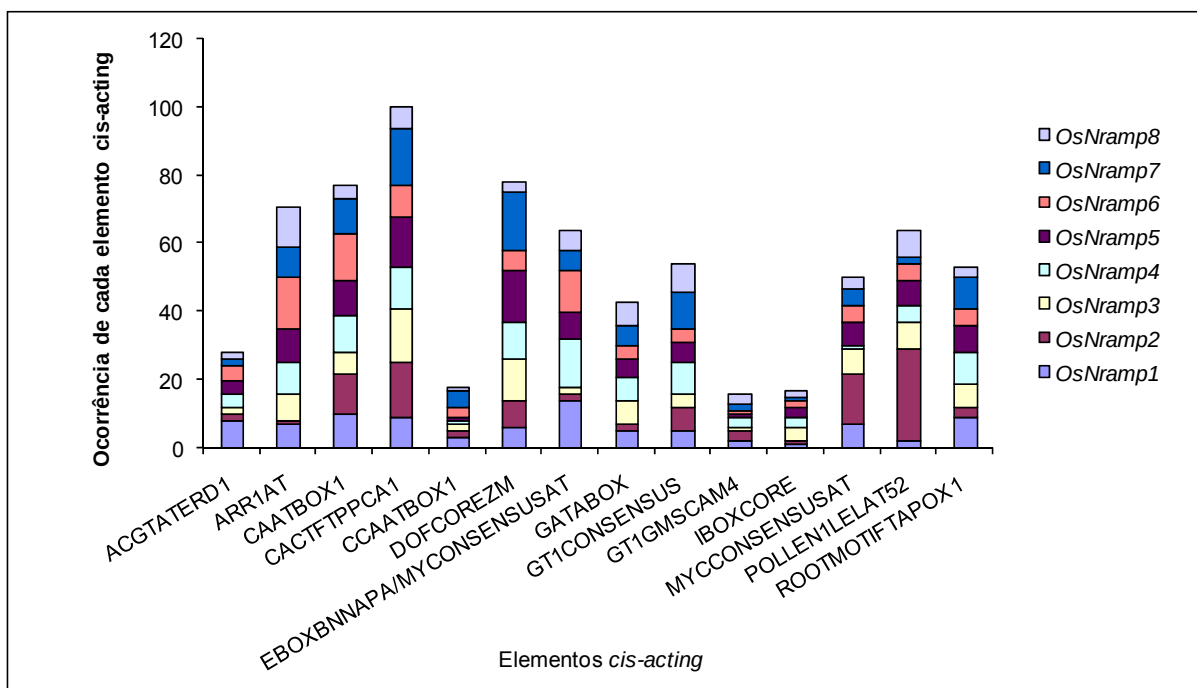


Figura 2.3. Ocorrência total de cada elemento *cis-acting* considerando todos os genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare.

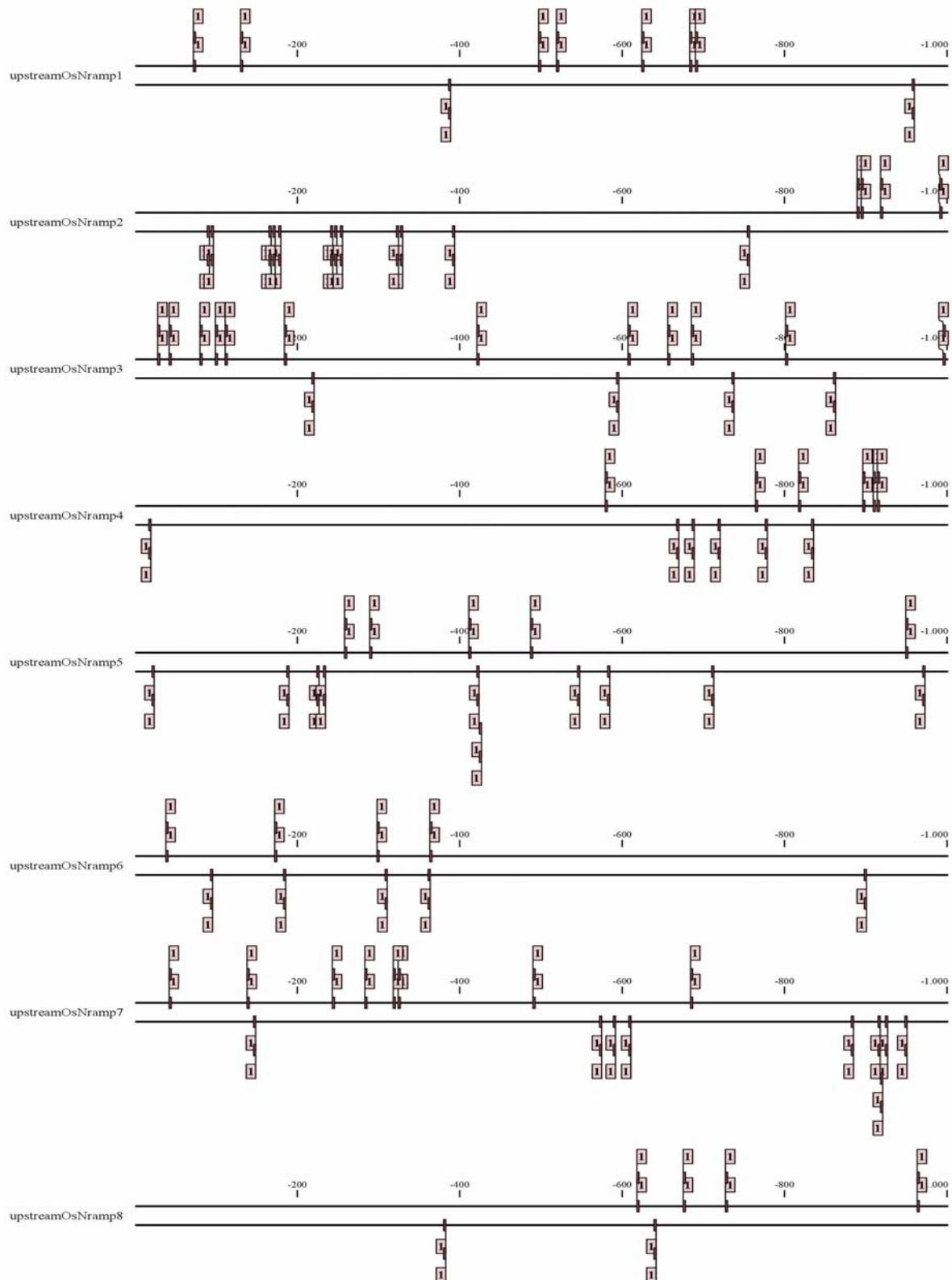


Figura 2.4. Posicionamento do elemento *cis-acting* CACTFTPPCA1 mais comum encontrado nas regiões promotoras dos genes da família *OsNrap* em *Oryza sativa* subsp *japonica* cv. Nipponbare (posicionamento da esquerda para direita)

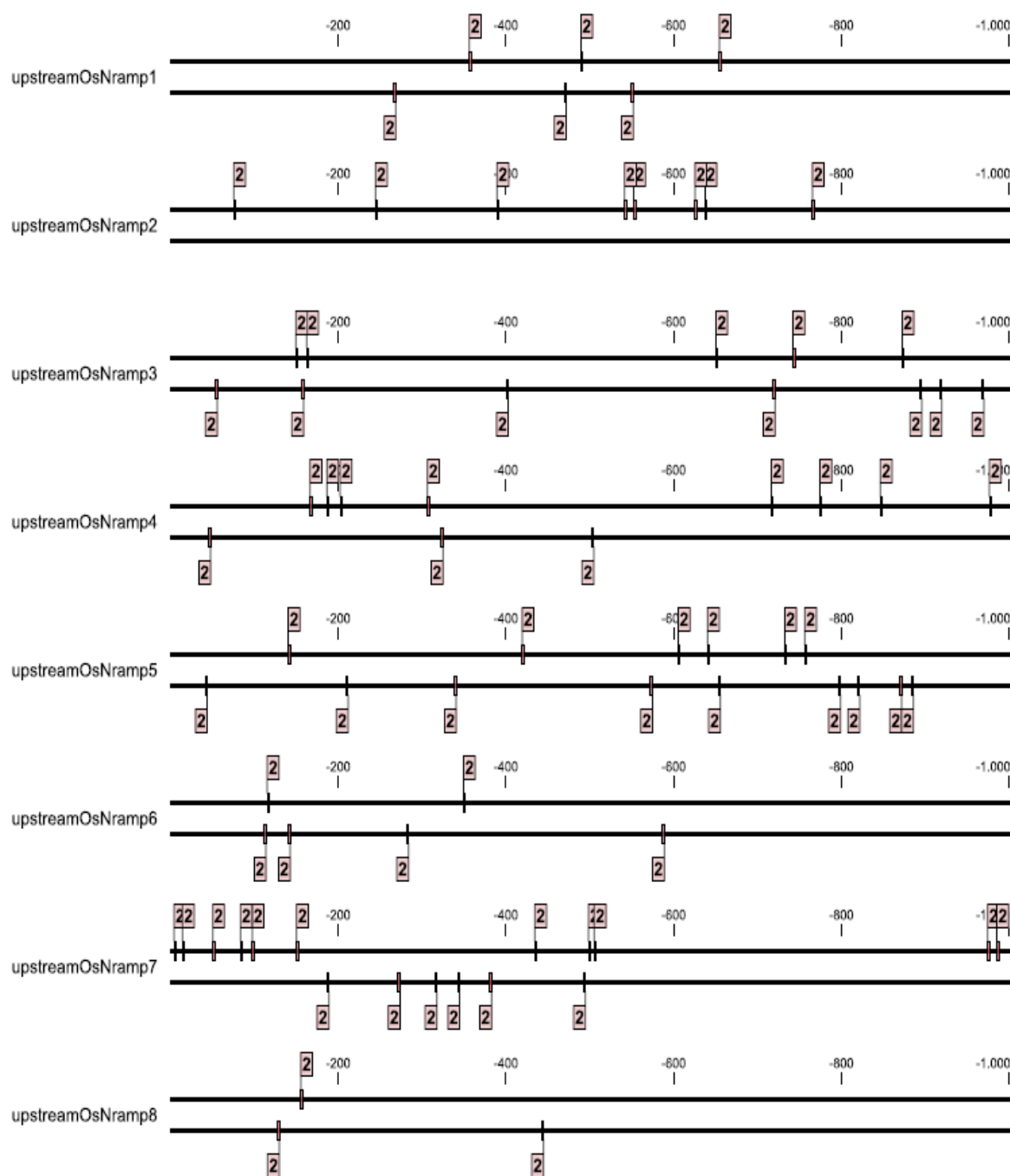


Figura 2.5. Posicionamento do elemento *cis-acting* DOFCOREZM mais comum encontrado nas regiões promotoras dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare (posicionamento da esquerda para direita).

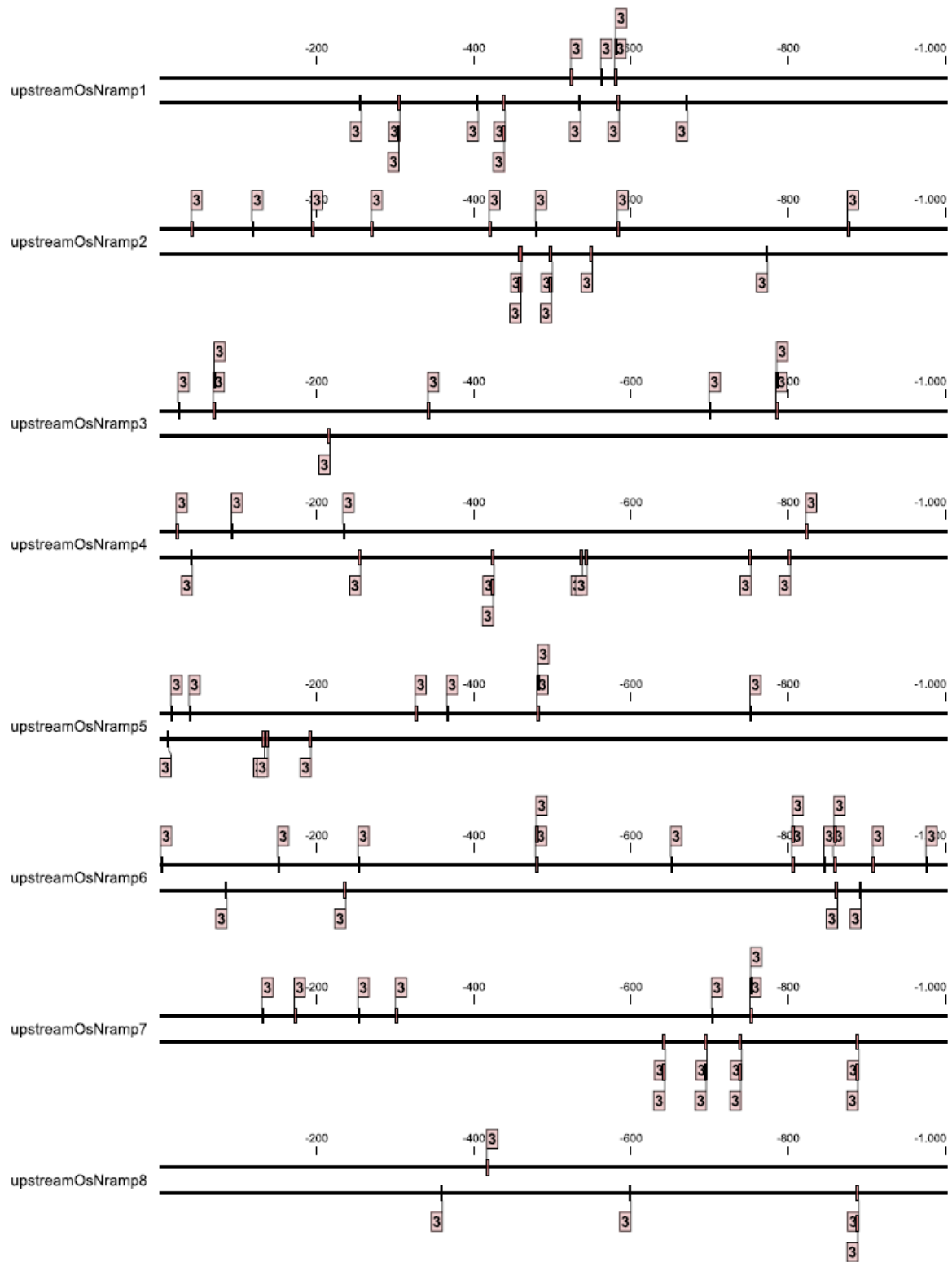


Figura 2.6. Posicionamento do elemento *cis-acting* CAATBOX1 mais comum encontrado nas regiões promotoras dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare (posicionamento da esquerda para direita).

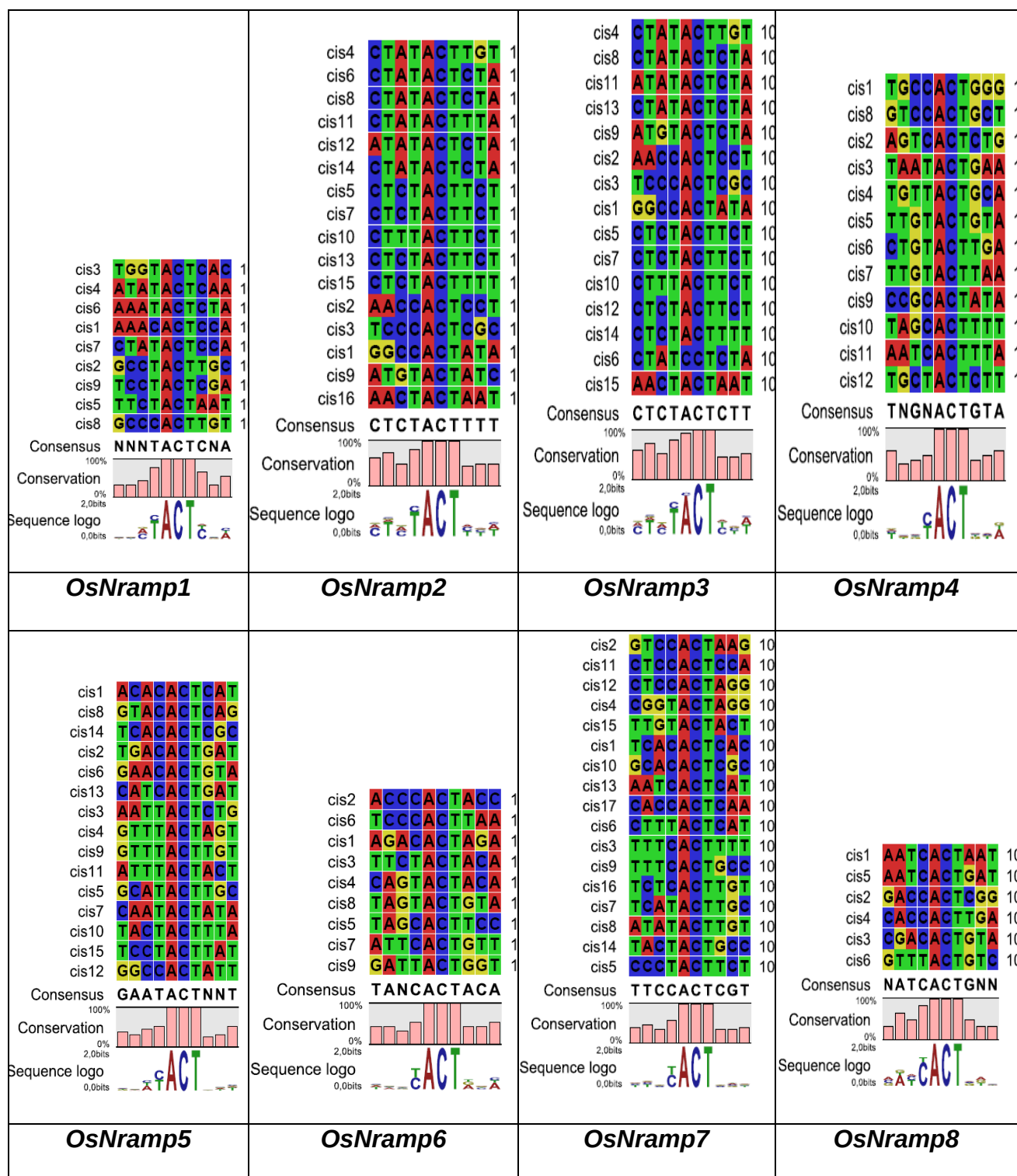


Figura 2.7. Alinhamento do elemento *cis-acting* CACTFTPPCA1 (YACT) mais comum encontrado nas regiões promotoras dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare.

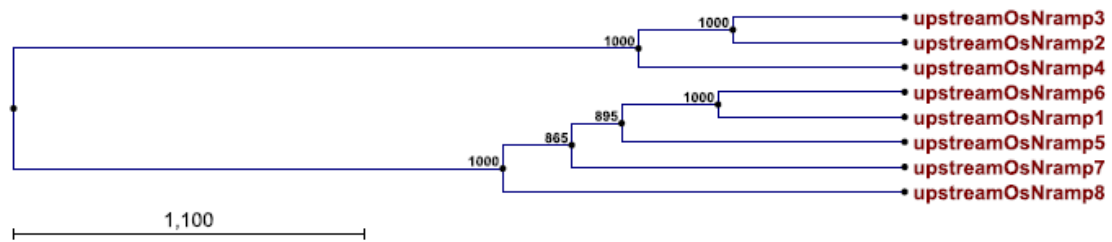


Figura 2.8. Dendrograma da região promotora dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare, pelo método de agrupamento UPGMA.

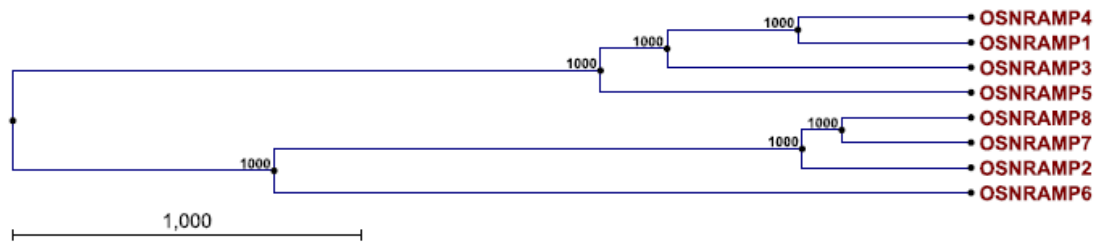


Figura 2.9. Dendrograma dos genes da família *OsNramp* em *Oryza sativa* subsp *japonica* cv. Nipponbare, pelo método de agrupamento UPGMA.

3. Análise de genômica comparativa de ESTs de aveia com espécies gramíneas e *Arabidopsis*.

Genetics and Molecular Biology (ISSN 1415-4757)

RESUMO

A genômica comparativa possibilita a comparação da organização de genomas relacionados, buscando inferir os processos básicos da evolução dos genomas, transferir a informação de espécies modelo para aquelas relacionadas, e integrar a informação da localização do gene e a expressão das espécies. Neste sentido, o objetivo do trabalho foi comparar as seqüências de ESTs de aveia com as espécies *Arabidopsis*, cevada, arroz, cana-de-açúcar, sorgo, trigo e milho, buscando inferir relações evolutivas dessas espécies com a aveia. As comparações foram feitas utilizando as ferramentas de BLASTn e tBLASTx. O banco de ESTs de aveia UNIGENE-NCBI consiste de muitas seqüências pouco similares a outras espécies comparados a nucleotídeos (46,72%) e aminoácidos (74,97%). Uma minoria de 4,56% das seqüências presentes no banco de ESTs de aveia UNIGENE-NCBI apresenta alta similaridade com seqüências de espécies de mono e dicotiledôneas.

INTRODUÇÃO

A família das Poaceas (gramíneas) é composta por mais de 10.000 espécies (Kellogg, 2000), incluindo as espécies de maior importância econômica e principais culturas de grãos, como o arroz (*Oryza sativa* L.), milho (*Zea mays* L.), trigo (*Triticum aestivum* L.), sorgo (*Sorghum vulgare* L.), cevada (*Hordeum vulgare* L.), centeio (*Secale cereale* L.), aveia (*Avena sativa* L.) e outras espécies, como a cana-de-açúcar (*Saccharum officinarum* L.). Estas espécies apresentam uma grande diversidade no número básico de cromossomos (arroz, 12; milho, 10; sorgo, 10; cevada, trigo e aveia com 7, e cana-de-açúcar com 8), nível de ploidia e o tamanho do genoma (arroz, 398 Mb; milho, 2700 Mb; sorgo, 770 Mb; cevada, 5000 Mb; trigo, 16000 Mb; aveia, 11300 Mb) (Devos e Gale, 1998; Keller e Feuillet, 2000).

As gramíneas divergiram de um ancestral comum a aproximadamente 50 milhões de anos atrás (Kellogg, 2001), e apresentam grande variação no tamanho de genoma, nível de ploidia e número de cromossomos, mesmo assim estudos comparativos indicam que existe um alto nível de conservação na ordem e conteúdo dos genes em nível macromolecular (Gale e Devos 1998), e que a maior parte da variação no tamanho do genoma é causada pela poliploidia e amplificação de elementos transponíveis (SanMiguel *et al.*, 1998).

Estudos comparativos das espécies sempre apresentaram grande importância científica para tentar elucidar aspectos evolutivos. Similaridades e diferenças entre genomas de espécies distintas podem ser estudadas pela genômica comparativa, um área da genética que permite a compreensão da função de tais genomas e os processos evolucionários atuantes (Jardim, 2007). A genômica comparativa tem se mostrado uma ferramenta poderosa para decifrar genes e na evolução dos genomas, além de melhorar a anotação do genoma (Lu *et al.*, 2009).

As análises de genômica comparativa demonstraram que a ordem dos genes entre espécies de plantas relacionadas permanecem largamente conservadas ao longo dos milhões de anos de evolução (Devos e Gale, 2000), por isso, as gramíneas têm servido como uma família modelo para a genética e genômica comparativa de plantas desde a década passada (Bennetzen e Freeling, 1993).

O nível de conservação dos genes é um importante critério para determinar a extensão com que o conhecimento comparativo possa ser aplicado entre as espécies (Devos, 2005). A disponibilidade da informação da seqüência de um fragmento de DNA genômico oferece uma oportunidade única para este tipo de análise (Schmidt, 2002). Desta forma, os principais objetivos da pesquisa em genômica comparativa são a comparar a organização de genomas relacionados e inferir os processos básicos da evolução dos genomas, transferir a informação de espécies modelo para aquelas relacionadas, e integrar a informação da localização do gene e a expressão das espécies. A genômica comparativa tem fornecido a base e o estímulo para integrar o conhecimento para aplicação em todas as culturas de cereais (Devos e Gale, 2000).

Lu *et al.* (2009), relacionam vários aspectos que podem ser beneficiados pelas comparações múltiplas entre as espécies, desde revelar novos fatos na evolução do genoma (Bennetzen, 2007), duplicação do genoma (Ilick *et al.*, 2003), e na origem do gene, além de também possibilitar a identificação prévia dos componentes do genoma desconhecidos ou poucos caracterizados, tais como novos elementos transponíveis (Lai *et al.*, 2005) e novos elementos funcionais (Stark *et al.*, 2007). Estudos comparativos podem fornecer informações chave da estrutura genômica e dos mecanismos responsáveis pelas diferenças no tamanho e na evolução do genoma das gramíneas (Feuillet e Keller, 2002).

Um bom conhecimento da organização do genoma é necessário para definir as melhores estratégias e ferramentas para isolar genes de importância agronômica dos amplos e complexos genomas dos cereais. O uso de seqüências de DNA tem sido à base da genômica comparativa para estudos evolucionários e a transferência de informações de espécies modelo para espécies agrícolas tem revolucionado a genética molecular e estratégias de melhoramento das culturas.

Entre as gramíneas, o arroz apresenta um genoma pequeno, mapas genéticos e físicos, e seu seqüenciamento completo (IRGSP, 2005), e devido a estes fatos tem sido considerado um modelo para o melhoramento genético de gramíneas (Izawa e Shimamoto, 1996; Gale e Devos, 1998; Malone, 2005). A maioria das aplicações comparativas utiliza as seqüências genômicas do arroz como fonte para marcadores de genes de interesse em outras espécies gramíneas (Brunner et al., 2003; Sutton et al., 2003; Gottwald et al., 2004; Devos, 2005). A alta conservação na ordem e conteúdo dos genes dentro dos cereais indica que a pesquisa com arroz pode proporcionar benefícios aos programas de pesquisa de outras gramíneas (Bennetzen e Ma, 2003). A espécie pode prover valiosas informações para investigação de características semelhantes no genoma de outras gramíneas (Bennetzen, 2002).

Os programas de melhoramento das culturas podem usar a genética comparativa para transferir informação sobre genes de espécies modelo para outras espécies de interesse, para auxiliar na identificação de genes que controlam características de interesse, e para acessar a diversidade alélica intra-espécies de forma que os melhores alelos possam ser identificados e agrupados em variedades superiores (Sorrels et al., 2008).

De acordo com Varshney *et al.* (2005), a sintenia e colinearidade entre diferentes genomas de gramíneas, provem novas oportunidades para uma eficiente transferência de informações genéticas das espécies modelos para grupos de plantas melhoradas e para grupos das espécies denominadas “órfãs”, ou seja, com poucos investimentos em pesquisa genômica, resultando em menores custos financeiros para o melhoramento genético das espécies.

Os avanços na genômica de espécies agrícolas têm resultado em um entendimento unificado da biologia das plantas, assim como tem gerado um grupo poderoso de ferramentas e métodos moleculares e de bioinformática. Estes avanços proporcionam uma oportunidade para a transferência de informação de espécies modelo e culturas principais para as culturas órfãs (Naylor *et al.*, 2004).

Entre os cereais de maior importância agronômica, a aveia é considerada uma cultura órfã devido ao menor rendimento econômico gerado. A aveia é cultivada com a finalidade de produção de grãos, forragem verde, feno, silagem, adubação verde no sistema de semeadura direta e na rotação de culturas (Carvalho *et al.*, 1987). Os grãos de aveia são destinados basicamente ao arração animal e em torno de 20% da produção mundial é processada para elaboração de alimentos humanos. No entanto, este paradigma está mudando, e o interesse pela aveia vem aumentando rapidamente por causa de suas características únicas. O grão de aveia tem excelente valor nutricional, elevada qualidade protéica e porcentagem de lipídios com predominância de ácidos graxos insaturados, aliado ao adequado conteúdo de carboidratos, que apresenta alta proporção de polissacarídeos não amiláceos, principais constituintes das fibras alimentares, das quais se destacam as chamadas β -glucanas, de grande responsabilidade pelas propriedades hipocolesterolêmicas da aveia (De Sá *et al.*, 2000).

Este trabalho proverá uma visão geral das relações evolutivas existentes entre a aveia e outras espécies gramíneas. Neste sentido, o objetivo do trabalho foi comparar as seqüências de aveia com as espécies *Arabidopsis*, cevada, arroz, cana-de-açúcar, sorgo, trigo e milho, buscando observar as relações evolutivas dessas espécies com a aveia.

MATERIAL E MÉTODOS

Obtenção das seqüências das espécies

Seqüências de oito espécies de plantas disponíveis atualmente, no banco de dados Unigene no National Center for Biotechnology Information - NCBI (<http://www.ncbi.nlm.nih.gov/>) foram baixadas (tabela1).

Comparação das seqüências

Para a eliminação da redundância, as seqüências de nucleotídeos (667), ESTs (25.357) e GSS (2.671) de aveia foram reunidas em arquivo de formato fasta, e submetidas ao programa CAP3, gerando 11.678 seqüências não redundantes de aveia. Também foi construído um “script” de algoritmo e banco de dados local para a comparação e armazenamento dos resultados das análises realizadas, respectivamente.

As seqüências de aveia foram comparadas com seqüências de *Arabidopsis* e outras seis espécies oriundas do Banco Unigene, utilizando as ferramentas de BLASTn e tBLASTx. Para o arroz também foi feita a comparação com o genoma completo da espécie através de BLASTn, com o objetivo de conhecer alinhamentos em regiões intergênicas do genoma do arroz. Seqüências com e-value maior do que $1e^{-15}$ e tamanhos maiores que 150 bases foram consideradas como similares.

Para aquelas seqüências conservadas entre a aveia e as demais espécies analisadas foi realizada uma classificação funcional dos genes de acordo com os bancos Gene Ontology (GO) e o InterPRO. O GO serve para a descrição do produto gênico e o InterPRO é um banco de dados de famílias de proteínas e características relacionadas, sendo que estes dois bancos são geralmente utilizados na classificação gênica.

RESULTADOS E DISCUSSÃO

Neste estudo, um total de 11.678 ESTs de *Avena sativa* foram analisados comparativamente com seqüências de *Arabidopsis*, *Hordeum vulgare*, *Oryza sativa*, *Saccharum officinarum*, *Sorghum bicolor*, *Triticum aestivum* e *Zea mays*. Somente 4,56% e 4,57% destes ESTs apresentaram seqüências similares para nucleotídeos e aminoácidos, respectivamente (Figura 3.1), para todas as espécies analisadas. Do total de ESTs de *Avena sativa*, 46,72% não apresentaram seqüências similares a nucleotídeos em nenhuma espécie. Embora a probabilidade de encontrar seqüências similares para aminoácidos seja maior, devido ao códon do aminoácido ser degenerado, o valor de ESTs sem similaridade aumentou para 74,97%.

A porcentagem de seqüências encontradas similares a nucleotídeos, aminoácidos, e ambos na comparação realizada entre *Avena sativa* e as demais espécies é apresentada na Figura 3.2. As espécies *Triticum aestivum*, *Hordeum vulgare* e *Zea mays* apresentaram os maiores números de seqüências similares com *Avena sativa* para nucleotídeos e aminoácidos, seguidos por *Saccharum officinarum*, *Sorghum bicolor* e *Oryza sativa*, respectivamente.

A espécie *Arabidopsis thaliana* apresentou menor número de seqüências similares para nucleotídeos, porém, quando aminoácidos foram comparados, a

conservação de seqüências com similaridade foi maior, inclusive maior do que nas espécies *Saccharum officinarum*, *Sorghum bicolor* e *Oryza sativa*. Muitas funções gênicas em diversos eucariotos, incluindo as angiospermas, são dirigidas por genes que exibem muito mais similaridade em nível de aminoácidos e mesmo em nível de nucleotídeos (Rubin *et al.*, 2000). Além disso, *Arabidopsis thaliana*, por ser uma espécie dicotiledônea, pertencente à família Brassicaceae ou Cruciferae (Meyerowitz e Somerville, 1994; Meinke *et al.*, 1998), está mais distante evolutivamente das gramíneas. Monocotiledôneas e dicotiledôneas divergiram há aproximadamente 200 milhões de anos (Wolfe *et al.*, 1989). No entanto, apesar desta distância evolutiva, em uma análise recíproca, 90% das proteínas de *Arabidopsis thaliana* têm um homólogo putativo no genoma de *Oryza sativa* (IRGSP, 2005). As funções protéicas são mantidas mesmo com as mutações silenciosas nos códons, devido à capacidade dos códons degenerados produzirem os mesmo aminoácidos (Nelson e Cox, 2004).

Comparações de seqüências de DNA em nível de nucleotídeos são muito mais sensíveis a mudanças, mesmo regiões de espécies distantemente relacionadas podem ser confiavelmente analisadas (Schmidt, 2002). Os genes tendem a ser altamente conservados em nível de seqüência de DNA (Devos e Gale, 2000). Segundo Chen *et al.* (1997), as seqüências evolutivamente conservadas nos segmentos de cromossomos correspondem a genes, com seqüências de introns evoluindo muito mais rapidamente do que os exons. As regiões codificantes são geralmente bem conservadas entre genes homólogos de gramíneas, mas a distância entre os genes parece estar correlacionada com o tamanho dos genomas (Bennetzen, 2000), que está conseqüentemente relacionado também com o nível de ploidia das espécies. Com base em análises comparativas, as seqüências de genes

e DNAs repetitivos evoluem a diferentes taxas, sendo os genes mais altamente conservados (Chen *et al.*, 1998; SanMiguel *et al.*, 1998). As seqüências derivadas de diferentes espécies quando analisadas tendem a apresentar similaridades menores quando o tamanho do genoma aumenta (Bennetzen, 2002), e a perda da colinearidade aumenta quanto maior a distância das espécies (Laurie e Devos, 2002).

Todas as espécies analisadas pertencem à família Poaceae, exceto *Arabidopsis thaliana*. Acredita-se que há aproximadamente 50 milhões de anos, a família das gramíneas divergiu em dois clados conhecidos como BEP-Bambusoideae, Ehrhartoideae e Pooideae (incluem arroz, trigo, cevada e aveia) e **PACC** (milho, sorgo, cana de açúcar e milheto) (Doebley *et al.*, 1990; Kellogg, 2001; Gaut, 2002; Paterson *et al.*, 2004).

As espécies *Triticum aestivum*, *Hordeum vulgare* e a *Avena sativa* pertencem a sub-família Pooideae (Kellogg *et al.*, 2001; Gaut, 2002) e estima-se que as duas primeiras espécies tenham divergido da aveia a cerca de 25 milhões de anos (Gaut, 2002), fatos que confirmam o maior número de seqüências similares encontradas entre aveia e as duas espécies.

A grande quantidade de seqüências similares encontradas para a espécie *Zea mays* pode ser explicada pelo número de seqüências depositadas no banco UniGene (82.630), duas vezes maior que o banco de *Triticum aestivum* (40.439), e em torno de três vezes maior do que o de *Hordeum vulgare* (23.595) (Figura 3.3). Esta similaridade não seria esperada, uma vez que a espécie *Zea mays* é da sub-família Panicoideae, e evolutivamente mais distante de *Avena sativa*, já que divergiram a aproximadamente 50 milhões de anos (Kellogg *et al.*, 2001; Gaut, 2002).

O menor número de seqüências similares encontradas entre *Avena sativa* e as espécies *Saccharum officinarum* e *Sorghum bicolor* ocorreu conforme o esperado, uma vez que estas espécies fazem parte da mesma sub-família do milho (Gaut, 2002), e o número de seqüências depositadas no banco UniGene destas duas espécies, é em torno de seis vezes menor que a quantidade de seqüências depositadas no banco do milho.

Oryza sativa pertence à sub-família Ehrhartoidae, que divergiu a 46 milhões de anos (Gaut, 2002), da sub-família Pooideae (aveia, cevada e trigo) (Kellog, 2001; Gaut, 2002), e a aproximadamente 50 milhões de anos da sub-família Panicoideae (Gaut, 2002). As espécies *Oryza sativa*, *Saccharum officinarum* e *Sorghum bicolor* apresentaram valores aproximados de seqüências similares à aveia, o que é consistente com as distâncias evolutivas propostas acima (Figura 3.3 e Tabela 3.1). Neste caso, a maior representação do banco de arroz, que chega a 2,3 a 3 vezes o tamanho dos outros dois bancos, não foi suficiente para causar um aumento relativo das seqüências similares do arroz. Isto sugere que o arroz esteja mais distante da sub-família Pooideae do que os membros da Panicoideae, como detectado para o milho. Esta inferência poderá ser confirmada quando um maior número de seqüências estiver disponível para o sorgo e a cana-de-açúcar.

Na Figura 3.4 é apresentado o número de seqüências encontradas em aveia similares ao arroz, considerando também alinhamentos relativos à região intergênica. Apresentaram similaridade acima do valor estabelecido como ponto de corte 3.112, 86 e 347 seqüências quando comparados nucleotídeos, aminoácidos e regiões intergênicas do arroz, respectivamente.

A similaridade dos ESTs com seqüências de regiões intergênicas, pode provavelmente ser oriunda do ganho ou perda de função na aveia, após a

divergência que ocorreu a aproximadamente 46 milhões de anos (Kellog, 2001; Gaut, 2002). Estas seqüências também podem ser retrotransposons, os quais são ubíquos em plantas e desempenham um papel principal em genes de plantas e na evolução do genoma, e são os principais responsáveis pelo tamanho dos genomas (Kumar e Bennetzen, 1999; Feschotte *et al.*, 2002). Estes transposons estão concentrados em regiões intergênicas e arranjados em grandes trechos de elementos inseridos (San Miguel *et al.*, 1996, 2002). Em arroz, os retrotransposons estão também na maioria dos casos inseridos em regiões intergênicas (Goff *et al.*, 2002; Yu *et al.*, 2002).

As seqüências conservadas entre a aveia e as demais espécies analisadas foram classificadas em 16 categorias funcionais acordo com os bancos GO e InterPRO (Figura 3.5). A categoria metabolismo de RNA apresentou 23%, seguido do metabolismo de proteínas e atividade catalítica com 12%, e metabolismo energético com 11%. Segundo Laurie e Devos (2002), vários aspectos do metabolismo celular são altamente conservados nos eucariotos, refletindo a sua origem precoce e importância para a viabilidade celular, confirmando os resultados encontrados no estudo.

De forma geral, os resultados encontrados no trabalho são confirmados pelos estudos de genômica comparativa e de evolução realizados. Apesar da divergência das gramíneas de um ancestral comum a aproximadamente 50 milhões de anos (Kellogg, 2001), e a presença de grande variação no tamanho de genoma, nível de ploidia e número de cromossomos, estudos comparativos indicam que existe um alto nível de conservação na ordem e conteúdo dos genes em nível macromolecular (Gale e Devos 1998), com pequenos rearranjos ocasionais (Lu *et al.*, 2009). A maior

parte da variação no tamanho do genoma das gramíneas é causada pela poliploidia e amplificação de elementos transponíveis (SanMiguel *et al.*, 1998).

Desta forma, as gramíneas podem ser consideradas como um sistema genético único (Bennetzen e Freeling, 1993). De acordo com Varshney *et al.* (2005), a sintenia (Moore *et al.*, 1995) e colinearidade (Chen *et al.*, 1997; Chen *et al.*, 1998) entre diferentes genomas de gramíneas, podem possibilitar uma eficiente transferência de informações genéticas das espécies modelos para grupos de plantas melhoradas e para grupos das espécies denominadas “órfãs”, como é o caso da aveia.

As estruturas dos genomas dos cereais e os seus respectivos genes são a chave para o entendimento das relações evolutivas dentro desta família, auxiliando na elucidação da biologia evolutiva dos cereais e ajudando os melhoristas de plantas nos seus objetivos de desenvolver melhores e mais estáveis linhagens. As seqüências em comum encontradas para as espécies podem ser usadas em estudos gênicos e de organização dos genes nos genomas destas espécies estudadas. Os avanços na genômica das gramíneas, principalmente em arroz, e inclusive em *Arabidopsis* podem beneficiar a cultura da aveia, pois apesar da grande disparidade no tamanho do genoma dos cereais e *Arabidopsis*, o tamanho, número e distribuição dos introns e exons são similares nos genes seqüenciados dos diferentes cereais e de *Arabidopsis* (Ware e Stein, 2003).

CONCLUSÕES

O banco de ESTs de aveia UNIGENE-NCBI consiste de muitas seqüências pouco similares a outras espécies comparados a nucleotídeos (46,72%) e aminoácidos (74,97%).

Uma minoria de 4,56 % das seqüências presentes no banco de ESTs de aveia UNIGENE-NCBI apresenta alta similaridade com seqüências de espécies de mono e dicotiledôneas.

REFERÊNCIAS

- Bennetzen JL (2007) Patterns in grass genome evolution. **Current Opinion in Plant Biology**. 10:176–181.
- Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: Genome composition, collinearity and compatibility. **Trends in Genetics**. 9:259–261.
- Bennetzen JL, Ma J. (2003). The genetic colinearity of rice and other cereals based on genomic sequence analysis. **Current Opinion in Plant Biology**. 6:128-133.
- Bennetzen, J. (2002). The rice genome: Opening the door to comparative plant biology. **Science**, 296(5565):60-63.
- Brunner S, Keller B, Feuillet C. (2003). A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the Rph7 locus. **Genetics**. 164:673-683.
- Carvalho FIF, Barbosa JF, Floss et al. (1987). Potencial genético da aveia como produtora de grãos no Sul do Brasil. **Pesquisa Agropecuária Brasileira**. 22(1):71-82.
- Chen M, SanMiguel P, Bennetzen JL. (1998). Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. **Genetics**. 148:435-443.
- Chen M, SanMiguel P, Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen, JL. (1997). Microcollinearity in sh2-homologous regions of the maize, rice, and sorghum

genomes. *Proceedings of the National Academy of Sciences of the United States of America*. 94:3431-3435.

De Sá RM, De Francisco A, Ogliari PJ, Bertoldi FC. (2000). Variação no conteúdo de betaglucanas em cultivares brasileiros de aveia. **Ciência e Tecnologia de Alimentos**. 20(1):99-102.

Devos KM, Gale MD (2000). Genome relationships: the grass model in current research. **The Plant cell**. 12(5):637-646.

Devos KM. (2005). Updating the 'crop circle'. **Current Opinion in Plant Biology**. 8:155-162.

Doebley JM, Durbin EM, Golenberg EM, Clegg MT, Ma DP. (1990). Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae). **Evolution**. 44:1097–1108.

Feschotte C, Jiang N, Wessler, SR. (2002). Plant transposable elements: where genetics meets genomics. **Nature Review Genetics**. 3:329–341.

Feuillet C, Keller B. (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Annals of Botany*. 89:3-10.

Gale MD, Devos, KM (1998). Comparative genetics in the grass. **Proceedings of the National Academy of Sciences of the United States of America**. 95:1971-1974.

Gibas C, Jambeck P. (2001). **Desenvolvendo Bioinformática: ferramentas de softwares para aplicações em biologia**. Rio de Janeiro. Campus.

Goff SA, Ricke D, Lan TH, et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). **Science**. 296:92-100.

Gottwald S, Stein N, Börner A, Sasaki T, Graner A. (2004) The gibberellic-acid insensitive dwarfing gene *sdw3* of barley is located on chromosome 2HS in a region

that shows high colinearity with rice chromosome 7L. **Molecular Genetics and Genomics**. 271:426–436.

Ilic K, SanMiguel PJ, Bennetzen JL. (2003). A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. **Proceedings of the National Academy of Sciences of the United States of America**. 100(21):12265-12270.

Izawa T, Shimamoto K. (1996). Becoming a model plant: The importance of rice to plant science. **Trends in plant science**, 1(3):95-99.

Jardim Belicuas, SN. (2007). Comparative genomics of grasses tolerant to aluminum. **Genetics and Molecular Research**. 6:1151-1155.

Keller B, Feuillet C. (2000). Colinearity and gene density in grass genomes. **Trends in Plant Science**. 5(6):246-251.

Kellogg EA. (2001). Evolutionary history of the grasses. **Plant Physiology**. 125:1198–1205.

Kumar A, Benntzen JL. (1999). Plant retrotransposons. **Annual Review Genetics**. 33:479-532.

Lai J, Li Y, Messing J, Dooner HK. Gene movement by Helitron transposons contributes to the haplotype variability of maize. (2005). **Proceedings of the National Academy of Sciences of the United States of America**. 102:9068–9073.

Laurie DA, Devos KM. (2002). Trends in comparative genetics and their potential impacts on wheat and barley research. **Plant Molecular Biology**. 48:729-740.

Lu F, Ammirajub JSS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Song X, Cheng Z, Oliveira AC de, Bennetzen JL. (2009). Comparative sequence analysis of *MONOCULM1*-orthologous regions in 14 *Oryza* genomes. **Proceedings of the**

National Academy of Sciences of the United States of America. 106(6):2071-2076.

Malone E. **Caracterização morfologia e molecular em mutantes de arroz (*Oryza sativa* L.) quanto à tolerância à toxidez por alumínio.** 2005. 67f. Dissertação (Mestrado em Biotecnologia Agrícola – Programa de Pós Graduação em Biotecnologia Agrícola) - Universidade Federal de Pelotas, Pelotas.

Meinke DW, et al. (1998) - *Arabidopsis thaliana*: A Model Plant for Genome Analysis. Science. 5389, 282, 662-682.

Meyerowitz EM, Somerville MR (1994). ***Arabidopsis***. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Naylor RL, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H, Nelson RJ. (2004). Biotechnology in the developing world: a case for increased investments in orphan crops. **Food Policy**. 29:15-44.

Nelson DL, Cox MM. **Lehninger Principles of Biochemistry**. 4ed. Nova Iorque: W. H. Freeman, 2004.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics.

Proceedings of the National Academy of Sciences of the United States of America. 101:9903-9908.

Paterson AH, Tanksley SD, Sorrells ME. (1991). DNA markers in plant improvement. **Advances in Agronomy**. 46:39-89.

Prosdocimi, Francisco et al. (2002). **Bioinformática: Manual do Usuário.** Biotecnologia Ciência & Desenvolvimento. 29:12-25.

Rubin GM, Yandell MD, Wortman JR, Miklos GLG, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W et al. (2000). Comparative genomics of the eukaryotes. **Science.**, 287:2204-2215.

San Miguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D. *et al.*, (1996). Nested retrotransposons in the intergenic regions of the maize genome. **Science.** 274:765-768.

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5Am. **Functional Integrative Genomics.** 2: DOI 10.1007/s10142-002-0056-4.

SanMiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima, and J. L. Bennetzen. (1998). The paleontology of intergenic retrotransposons of maize. **Nature Genetics.** 20:43-45.

Schmidt R. (2002). Plant genome evolution: lessons from comparative genomics at the DNA level. **Plant Molecular Biology.** 48:21-37.

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL et al.: (2003) Comparative DNA sequence analysis of wheat and rice genomes. **Genome Research.** 13:1818-1827.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N. et al. (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. **Nature.** 450:219-232.

- Sutton T, Whitford R, Baumann U, Dong CM, Able JA, Langridge P. (2003). The Ph2 pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. **Plant Journal**. 36:443-456.
- Varshney RK, Graner A, Sorrells ME. (2005) Genomics-assisted breeding for crop improvement. **Trends in Plant Science**. 10(12):621-630.
- Ware D, Stein L. (2003). Comparison of genes among cereals. **Current Opinion Plant Biology**. 6:121-127.
- Wolfe, K. H., M. Gouy, Y.-W. Yang, P. M. Sharp, and W.-H. Li. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. **Proceedings of the National Academy of Sciences** of the United States of America. 86:6201–6205.
- Yu J, Hu SN, Wang, et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). **Science**. 296:79-92.

Tabela 3.1. Espécies com o número de seqüências encontradas no Banco Unigene do NCBI em 06/06/2009.

Espécies	Número de seqüências UniGene
<i>Avena sativa</i> (aveia)	11.679
<i>Arabidopsis thaliana</i> (Arabidopsis)	30.576
<i>Hordeum vulgare</i> (cevada)	23.595
<i>Oryza sativa</i> (arroz)	40.972
<i>Saccharum officinarum</i> (cana-de-açúcar)	15.594
<i>Sorghum bicolor</i> (sorgo)	13.899
<i>Triticum aestivum</i> (trigo)	40.349
<i>Zea mays</i> (milho)	82.630

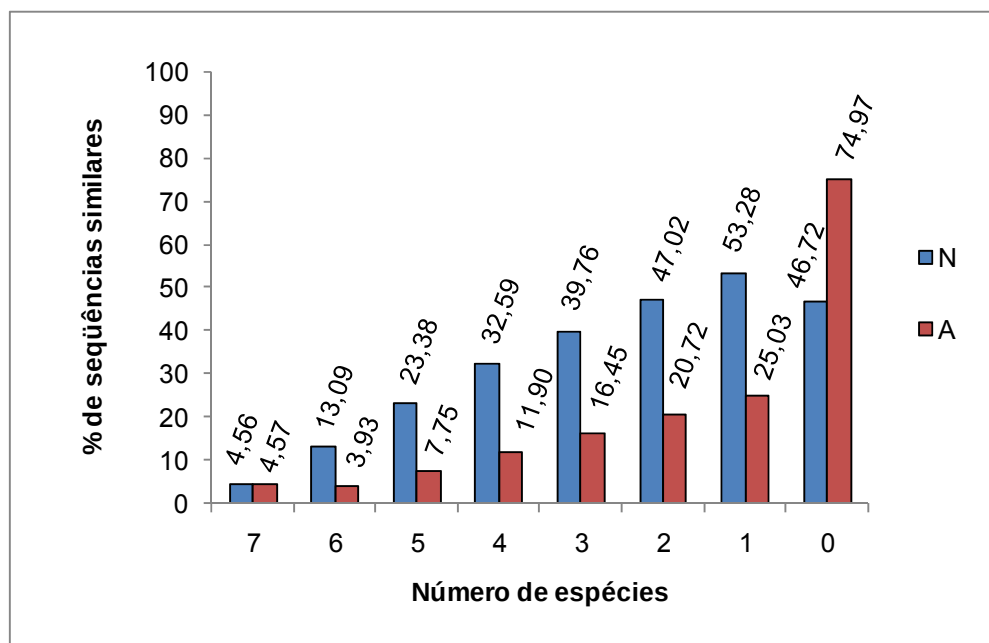


Figura 3.1. Porcentagem de seqüências similares nas espécies para nucleotídeos (N) e aminoácidos (A), baseados na comparação realizada entre aveia e as demais espécies (7–número de espécies consideradas no cálculo de porcentagem de seqüências de aveia similares presentes, e assim consecutivamente, 0–seqüências de aveia sem similaridade).

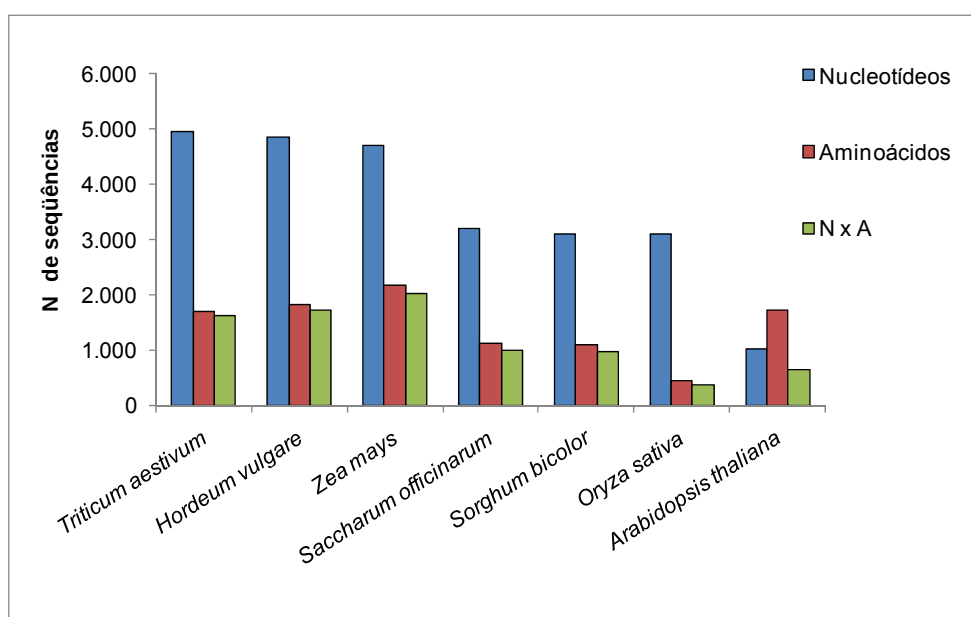


Figura 3.2. Número de seqüências encontradas similares a nucleotídeos, aminoácidos, e ambos na comparação realizada entre aveia e as demais espécies.

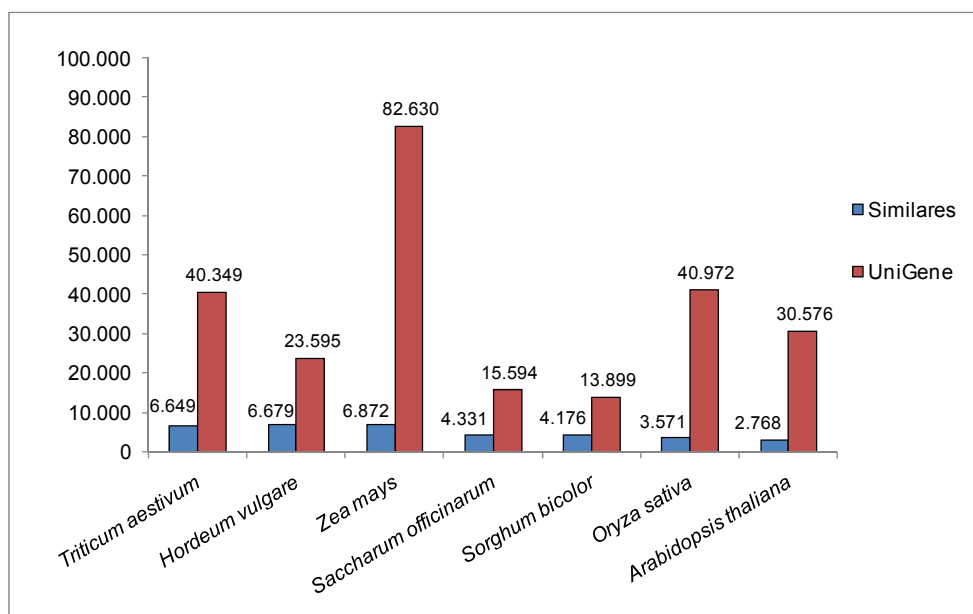


Figura 3.3. Número total de seqüências similares encontradas na comparação realizada entre aveia e as demais espécies e seqüências depositadas no banco UNIGENE.

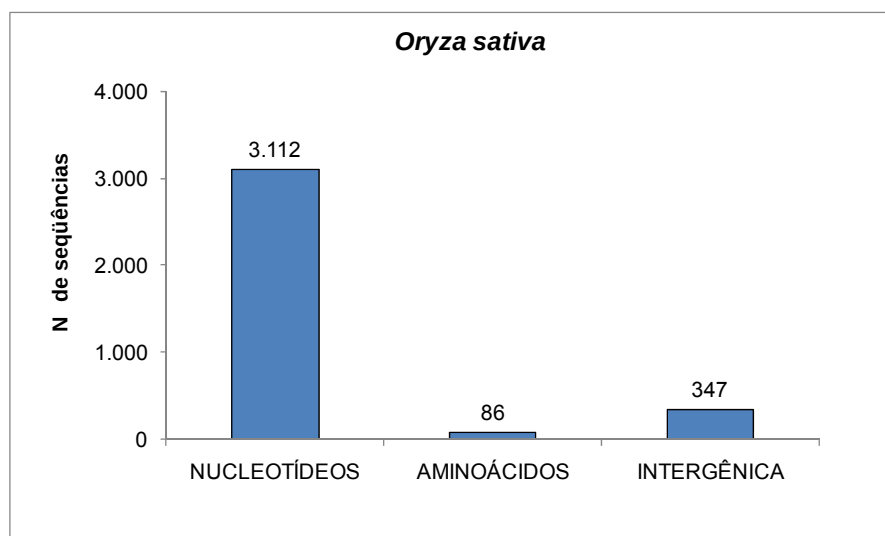


Figura 3.4. Número de seqüências de aveia similares a *Oryza sativa*.

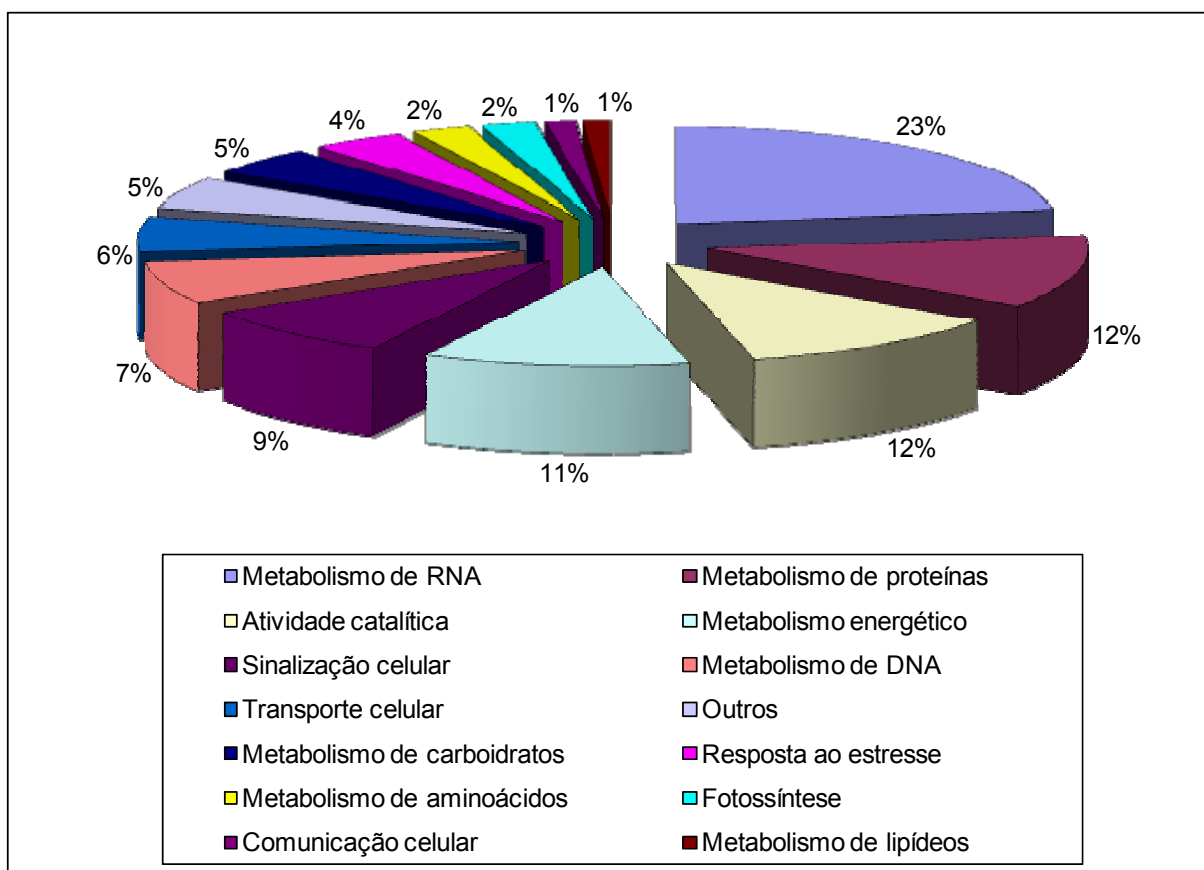


Figura 3.5. Classificação funcional dos genes conservados entre as oito espécies.

Discussão geral

Neste sentido, três estudos de genômica comparativa em gramíneas foram realizados, utilizando a bioinformática como ferramenta. No primeiro trabalho, o objetivo foi analisar a abundância de microssatélites no genoma de 13 espécies do gênero *Oryza*, descrevendo as taxas, frequências e padrão de distribuição dos diferentes microssatélites. O número de microssatélites encontrado no estudo pode ser tomado como amostra e servir para estimar a distribuição completa dos microssatélites, considerando o tamanho atual dos bancos de dados e o tamanho estimado dos genomas destas espécies. O gênero *Oryza* apresenta um grande número de espécies que podem ser utilizadas como fonte de variabilidade genética de alelos para o melhoramento genético de plantas. O conhecimento dos microssatélites presentes nas frações genômicas atualmente disponíveis das espécies de arroz pode facilitar os estudos moleculares e o uso da seleção assistida ao melhoramento de plantas, pois podem fornecer marcadores detectáveis para genes agronomicamente importantes. Os *primers* obtidos com base nos microssatélites devem ser validados em laboratório para uso posterior em trabalhos de variabilidade e diversidade genética, transferabilidade entre espécies, mapeamento genético e seleção assistida ao melhoramento nas espécies de arroz e em outras gramíneas. O programa SSRLocator mostrou ser eficiente e eficaz para a realização de estudos *in silico*, facilitando a obtenção de marcadores microssatélites, e conseqüentemente reduzir os custos no desenvolvimento de microssatélites no futuro e beneficiar a genética comparativa e mapeamento físico, estudos de diversidade genética, mapeamento gênico, seleção assistida por marcadores e eventualmente à clonagem posicional de genes úteis em cereais e outras espécies importantes das gramíneas.

No segundo trabalho, a região upstream dos genes *OsNramp* foi investigada quanto à abundância de elementos *cis-acting*. O conhecimento dos elementos *cis-acting* presentes na região promotora dos genes *OsNramp* possibilitará a escolha de candidatos para estudos posteriores de regulação da expressão destes genes, assim como contribuirá para a elucidação dos mecanismos regulatórios envolvidos no metabolismo do ferro em arroz, quem sabe futuramente no metabolismo de outros nutrientes na planta.

O objetivo do último trabalho foi comparar as seqüências de aveia com as espécies *Arabidopsis*, cevada, arroz, cana-de-açúcar, sorgo, trigo e milho, buscando observar as relações evolutivas dessas espécies com a aveia. Os resultados confirmaram a proximidade evolutiva da aveia com o trigo e a cevada. Como as gramíneas podem ser consideradas como um sistema genético único (BENNETZEN e FREELING, 1993), os avanços nos estudos de espécies modelo e culturas principais podem proporcionar uma oportunidade para a transferência de informação para as culturas órfãs (NAYLOR et al., 2004), e a aveia ser beneficiada pelos resultados científicos.

Referências Bibliográficas (Introdução e Discussão Geral)

- ALLARD, R.W. Principles of Plant Breeding. **2nd Ed. John Wiley and Sons. Inc.** New York, 1999.
- BROWN, C.M.; FORSBERG, R.A. Oat. In: FEHR, W.R. **Principles of cultivar development.** New York : Crop Species, 1987. 760p.
- CARVALHO, F.I.F.; LORENCETTI, C.; MARCHIORO, V.S.; SILVA, S.A. condução de populações no melhoramento genético de plantas. **Editora e gráfica Universitária - UFPel**, 2003.
- FEDERIZZI, L.C. Estrutura de um programa de melhoramento de plantas e possíveis aplicações de marcadores moleculares: visão do melhorista. In: MILACH, S. (Ed.). **Marcadores moleculares em plantas.** Porto Alegre: S. C. K. Milach, 1998. p.3-15.
- FEHR, W. R. Principles of cultivar development. New York: Mcmillian, 1987. 761p.
- FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.
- GOODMAN, N. Biological data becomes computer literate: new advances in bioinformatics. **Current Opinion in Biotechnology**, v.13, n.1, p. 68-71, 2002.
- KUMAR, L.S. DNA markers in plant improvement: an overview. **Biotechnology Advances**, Amsterdam, v.17, p.143-182, 1999.
- LORENCETTI, C.; CARVALHO, F.I.F.; OLIVEIRA, A.C.; VALÉRIO, I.P.; HARTWIG, I.; MARCHIORO, V.S.; VIEIRA, E.A. 2006. Retrocruzamento como uma estratégia de identificar genótipos e desenvolver populações segregantes promissoras em aveia. **Ciência Rural**, Santa Maria, v.36, n.4, p.1118-1125, 2006.

- MAIA, L.C. **Desenvolvimento de ferramenta e análise *in silico* da ocorrência de microssatélites (Simple Sequence Repeat) no genoma do arroz**. 2007. 120f. Dissertação (Programa de Pós-Graduação em Agronomia) - Universidade Federal de Pelotas. Pelotas.
- MILACH, S.C.K. Marcadores de DNA. **BIOTECNOLOGIA: Ciência e Desenvolvimento**, v.5, p.14-17, 1998a.
- MILACH, S.C.K. **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998b.
- NATIONAL CENTER OF BIOTECHNOLOGY INFORMATION - **The NCBI Homepage**. Disponível em: <http://www.ncbi.com> Acesso em: 2 jul. 2009.
- OLIVEIRA, A.C.B.; CAIXETA, E.T.; ZAMBOLIN, E.M.; ZAMBOLIN, L.; SAKIYAMA, N.S. Aplicação técnica de marcadores moleculares no melhoramento de plantas. **Documentos IAC**, Campinas: Instituto Agrônômico, n.81, 17p., 2007.
- RHEE, S.Y.; DICKERSON, J.; XU, D. Bioinformatics and Its Applications in Plant Biology. **Annual Review of Plant Biology**, v. 57: 335-360, 2006.
- SCHLEGEL, R.H.J. 2003. **Dictionary of plant breeding**. Food Products Press/The Haworth Reference Press, New York.
- SUMNER LW, MENDES P, DIXON RA. Plant metabolomics: largescale phytochemistry in the functional genomics era. **Phytochemistry**. 2003; 62:817–36.
- TANKSLEY SD. Molecular markers in plant breeding. **Plant Molecular Biology Reporter**. 1983.1:3–8.
- VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genomics-assisted breeding for crop improvement. **Trends in Plant Science**, v.10, n.12, p.621-630, 2005a.
- ZIMMER, PAULO DEJALMA ; OLIVEIRA, ANTONIO COSTA DE ; MALONE, GASPAR. Ferramentas biotecnológicas no melhoramento genético vegetal. Pelotas, RS, **Editora e Gráfica Universitaria**, UFPel, 2005.

Vitae

Clauber Mateus Priebe Bervald nascido em 25 de setembro de 1981, filho de Guido Bervald e Brunilda Priebe Bervald, natural de Pelotas – RS. Formado como Técnico em Agropecuária em 1998, pelo Conjunto Agrotécnico Visconde em Agropecuária, pertencente à Universidade Federal de Pelotas (UFPEL). Em 1999, ingressou na Faculdade de Agronomia Eliseu Maciel (FAEM) da UFPEL, obtendo o título de Engenheiro Agrônomo no ano de 2003. Durante este período, realizou estágios no Laboratório de Grãos do Departamento de Ciência e Tecnologia de Alimentos (DCTA), em administração e economia no Departamento de Ciências Sociais Agrárias (DCSA), foi bolsista do Programa Especial de Treinamento (PET), e Bolsista de Iniciação Científica PIBIC/UFPEL em Fisiologia Vegetal. Em março de 2004, iniciou no Mestrado em Ciência e Tecnologia de Sementes, sob orientação dos Professores Dario Munt de Moraes e Antônio Carlos Souza Albuquerque Barros, trabalhando com fisiologia de sementes. No ano de 2007 ingressou no Doutorado em Agronomia, área de concentração Fitomelhoramento, sob orientação dos Professores Antônio Costa de Oliveira e Fernando Irajá Félix de Carvalho, trabalhando com bioinformática. Além disso, ao longo deste período, desenvolveu trabalhos em fisiologia vegetal, metabolismo de carboidratos, sementes, melhoramento vegetal, biotecnologia e bioinformática.