

UNIVERSIDADE FEDERAL DE PELOTAS
Faculdade de Agronomia Eliseu Maciel
Programa de Pós-Graduação em Agronomia



Dissertação

Análise do transcriptoma de genótipos de arroz sob estresse por frio

Leomar Guilherme Woyann

Pelotas, 2014

Leomar Guilherme Woyann

Análise do transcriptoma de genótipos de arroz sob estresse por frio

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia da Faculdade de Agronomia Eliseu Maciel da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciências.

Orientador: Eng^o. Agr^o. Ph. D. Antonio Costa de Oliveira
Co-orientador: Eng^o. Agr^o. Dr. Luciano Carlos da Maia

Pelotas, 2014

Dados de catalogação na fonte:

(Gabriela Machado Lopes - CRB-10/1842)

W938a Woyann, Leomar Guilherme

Análise do transcriptoma de genótipos de arroz sob estresse por frio / Leomar Guilherme Woyann; orientador Antonio Costa de Oliveira; co-orientador Luciano Carlos da Maia – Pelotas, 2014.

109 f. : il.

Dissertação (Mestrado) Programa de Pós- Graduação em Agronomia. Faculdade de Agronomia Eliseu Maciel Universidade Federal de Pelotas. Pelotas, 2014.

1. Plântula 2. Transcriptoma 3. RNA-Seq 4. Genes diferencialmente expressos 5. Fatores de transcrição responsivos ao frio I. Oliveira, Antonio Costa de (orientador) II. Maia, Luciano Carlos da (co-orientador) III. Título.

Banca examinadora:

Antonio Costa de Oliveira, Ph. D. Professor – FAEM/UFPeI (presidente)

Luciano Carlos de Maia, Dr. Professor – FAEM/UFPeI

Daniel da Rosa Farias, Dr. – Pós-Doutorando – FAEM/UFPeI

Guilherme Loss de Moraes, Dr. – Pesquisador – Lab. Nac. Computação Científica

Agradecimentos

Agradeço primeiramente a minha mãe pela compreensão durante a minha ausência e pelo incentivo constante para que eu continuasse meus estudos.

Aos professores Antônio Costa de Oliveira e Luciano Carlos da Maia pela orientação, confiança e amizade durante todo o período de mestrado.

À Dra. Renata Ahlert pela inestimável colaboração na condução do experimento e na valiosa discussão a respeito das metodologias a serem empregadas neste trabalho.

Ao Dr. Daniel Farias, ao doutorando Luis Willian Pacheco Arge e ao graduando em Ciências da Computação Gustavo Neves pela grande ajuda na análise e discussão dos resultados obtidos.

À Embrapa, em nome do pesquisador Ariano Magalhães e do funcionário Alcides e Ailton pela ajuda sempre presente durante a realização deste projeto. Estendo este agradecimento a todos que de uma forma ou outra colaboraram na realização deste projeto.

Agradeço também a todos os professores que contribuíram na minha formação acadêmica, tanto em sala de aula quanto fora dela.

A todos os 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.

Agradeço às instituições de pesquisa CAPES, CNPq, FAPERGS, Fundação Pró-Sementes de Apoio à Pesquisa e Fundação Beachell-Bourlaug pela viabilização financeira para implantação deste projeto e financiamento da bolsa de pesquisa.

Ao Programa de Pós-Graduação em Agronomia, da Faculdade de Agronomia “Eliseu Maciel”, da Universidade Federal de Pelotas, por oportunizar o aprimoramento de minha formação profissional.

MUITO OBRIGADO!!!

Resumo

WOYANN, Leomar Guilherme. **Análise do transcriptoma de genótipos de arroz sob estresse por frio.** 2014. 107f.

Dissertação (Mestrado em Agronomia – Área de Concentração: Fitomelhoramento) - Programa de Pós-Graduação em Agronomia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Pelotas, 2014.

O estresse por frio pode trazer prejuízos econômicos significativos em diversas fases do ciclo da cultura do arroz. Em arroz (*Oryza sativa* L.), há grande variabilidade genética para tolerância ao frio sendo a subespécie *japonica* considerada tolerante e a subespécie *indica* considerada sensível. Foi realizado sequenciamento de RNA (RNA-Seq) visando identificar genes diferencialmente expressos entre os genótipos BRS Atalanta (*indica*) (A), Nipponbare (*japonica*) (N) e *bulks* formados por RILs (Linhas endogâmicas recombinantes) provenientes do cruzamento entre elas. *Bulks* (B) com as 15 RILs mais próximas a cada genitor (BA e BN) foram formados com base em dados de genotipagem. As plântulas, tanto da condição tratamento quanto da condição controle, permaneceram por 7d a 28°C. Para análise do transcriptoma, as plântulas da condição tratamento (T) foram expostas ao frio numa temperatura de 13°C por 24h e as plantas da condição controle (C) permaneceram mais 24h a 28°C. Após este período foi realizada a coleta das plântulas, extração de RNA total, síntese de cDNA e preparo das bibliotecas para sequenciamento (RNA-Seq). Este foi realizado em sequenciador HiSeq 2000, sendo as *reads* de 100b *single-end*. Diversas combinações foram analisadas para obter o número de genes diferencialmente expressos envolvendo as condições BRS Atalanta (controle e tratamento), Nipponbare (controle e tratamento), *bulk* de RILs mais próximos a cultivar BRS Atalanta (controle e tratamento) e *bulk* de RILs mais próximos ao genótipo Nipponbare (controle e tratamento). Um número significativo de genes mostra-se diferencialmente expressos em cada condição, sendo de modo geral mais genes estão subexpressos do que superexpressos, exceto para as condições AC x BAC, AT x BAT e NT x BNT. Um dendrograma usando as distâncias de Jensen-Shannon mostra a formação de dois ramos sendo que num deles está a cultivar BRS Atalanta e os *bulks* formados por sua semelhança genotípica com esta cultivar. No outro ramo estão o genótipo Nipponbare e os *bulks* formados pela semelhança com o genitor. Genes pertencentes a diversas famílias de fatores de transcrição estão diferencialmente expressos nas diferentes combinações, sendo que para grande parte destas famílias já foi demonstrada sua participação na resposta ao estresse por frio. As conclusões deste trabalho indicam que o número de genes diferencialmente expressos ($\log_2\text{-fold-change} \geq |1.5|$) varia grandemente entre as combinações, apresentando como limite superior 1481 genes subexpressos na combinação BNC x BAC e 1017 genes superexpressos em NT x AT, e seis genes subexpressos na combinação NT x BNT e cinco genes

superexpressos na combinação NC x BNC. As combinações analisadas mostram a expressão diferencial de um grande número de famílias de fatores de transcrição, muitas delas já descritas como responsivas ao estresse por frio, que apresentam um padrão difuso de expressão entre as subespécies *indica* e *japonica*.

Palavras-chave: arroz; frio; plântula; transcriptoma; RNA-Seq; genes diferencialmente expressos; fatores de transcrição responsivos ao frio.

Abstract

WOYANN, Leomar Guilherme. **Transcriptome analysis of rice genotypes under cold stress**. 2014. 107p.

Dissertation (Master Degree in Agronomy – Concentration's area: Plant Breeding) - Program of Post-Graduation in Agronomy, Faculty of Agronomy Eliseu Maciel, Federal University of Pelotas – Brazil, Pelotas, 2014.

Chilling stress can cause significant economic losses in different phases of the cycle of rice. There are a wide of genetic variability for cold tolerance in rice (*Oryza sativa* L.), *japonica* subspecies is considered tolerant and subspecies *indica* is considered sensitive. RNA sequencing (RNA-Seq) was performed to identify differentially expressed genes among the cultivar BRS Atalanta (*indica*) (A), Nipponbare (*japonica*) (N) and bulks composed by RILs from crosses between them. Bulks (B) with the 15 closest RILs to each parent (BA and BN) RILs were composed based on SNPs-genotyping data. To analyze the transcriptome, plants of the treatment condition (T), 7d after imbibition were exposed to a chilling temperature of 13°C per 24h and plants of the control condition (C) remained 24h more at 28°C. After this time was performed the collection of seedlings, total RNA extraction, cDNA synthesis and preparation of libraries for sequencing (RNA-Seq) in a HiSeq 2000 sequencer, with 100bp single-end reads. Several combinations were analyzed to obtain the number of differentially expressed genes involving BRS Atalanta conditions (control and treatment), Nipponbare (control and treatment), bulk of RILs closest to BRS Atalanta – BA (control and treatment) and bulks of RILs closest to the cultivar Nipponbare - BN (control and treatment). A significant number of differentially expressed genes are shown in each condition, being generally founded more down-regulated than up-regulated genes, except for the combinations AC x AT x NT x BAT and BAC x BNT. A dendrogram using the Jensen-Shannon distance shows the formation of two branches of which one is composed by BRS Atlanta cultivar and the bulks composed by the genotypic similarity with this cultivar. In the other branch is the cultivar Nipponbare and the bulks composed by similarity with the parent. Several differentially expressed genes from transcription factors families are present in different combinations, and for many of these families has been demonstrated its involvement in the response to chilling stress. The conclusions of this study indicate that the number of differentially expressed genes ($\log_2$ fold-change $\geq |1.5|$) greatly changed between combinations, with an upper limit of 1481 down-regulated genes in the combination BAC x BNC and 1017 up-regulated genes in the combination NT x AT, six genes are up-regulate in the combination NT x BNT and five genes are up-regulated in NC x BNC combination. The combinations analyzed showed differential expression of a large number of families of transcription factors, many of which have been described as responsive to cold stress, showing a diffuse pattern of expression between *indica* and *japonica* groups.

Key-words: rice; chilling; seedling; transcriptome; RNA-Seq; differentially expressed genes, chilling responsive transcription factors.

Lista de Figuras

- Figura 1 Modelo hipotético da tolerância ao frio não-congelante. Espécies reativas de oxigênio, cálcio e fitormônios como moléculas de sinalização podem desempenhar um papel crítico na adaptação de plantas ao estresse pelo frio. A modulação destas vias de sinalização ativa a regulação da transcrição. A expressão diferencial de fatores de transcrição faz com que muitos grupos funcionais se mostrem diferencialmente expressos entre genótipos tolerantes e sensíveis, tais como proteínas funcionais, modificações pós-traducionais de proteínas, e genes associados com processos fisiológicos e metabólicos, como a fotossíntese, a produção de ERO ou sistema de detoxificação e enzimas metabólicas. A modulação da expressão dos genes associados com modificações pós-traducionais, ERO, e fitormônios, em seguida, retroalimentou e fez o ajuste fino de fatores de transcrição, ERO, e redes de sinalização de fitormônios. ERO: espécies reativas de oxigênio, AO: amina oxidase, PODs: peroxidases, GST: glutathione S- transferases, CAT: catalase, RBCS3B: pequena subunidade 3B de rubisco, FNR: ferredoxina-NADP +- oxirredutase ; CABS: proteínas de ligação de clorofila a/b ; GOXs: glicolato oxidases, LOXs: lipoxigenases, ABA3: ABA deficiente 3, UGTs: UDP- glucosiltransferases, CKX: citocinina oxidase/desidrogenase, PAL: fenilalanina amônia-liase, PRS, proteínas relacionadas à patogênese, BAM: beta-amilase ; Hsps: proteínas de choque de calor, ERD: resposta precoce à desidratação, HSFS: fatores de choque térmico, SKP2A: proteína 2A associada a quinase fase S, RLK: receptor proteína quinase, LAP-A1: leucina aminopeptidase A1. Adaptado de Liu, et al., (2012)..... 25
- Figura 2 Visão geral de processamento do mRNA em eucariotos. Pouco tempo depois da RNA polimerase II iniciar a transcrição do primeiro nucleotídeo, do primeiro éxon de um gene, a extremidade 5' do RNA nascente recebe um CAP formado por com 7-metilguanilato. A transcrição pela RNA polimerase II termina em qualquer um dos vários locais de terminação a jusante do local da poli(A), que está localizado na extremidade 3' do éxon final. Após, a transcrição primária é clivada no local da poli(A), uma cadeia de resíduos de adenina (A) é adicionada. A cauda poli(A) contém resíduos ≈ 250 A em mamíferos, ≈ 150 em insetos, e ≈ 100 em leveduras. Para transcrições primárias curtas com alguns íntrons, a poliadenilação, clivagem e splicing geralmente segue-se a terminação, como mostrado. Para grandes genes com vários íntrons, os íntrons são frequentemente retirados do RNA nascente antes da transcrição do gene ser completada. O CAP na extremidade 5' é mantido em mRNAs maduros. Obtida de Lodish et al., (2000)..... 27

Figura 3	Preparação de biblioteca de RNA para plataforma Illumina. RNA com cauda poli(A) é enriquecida utilizando beads oligo(dT) seguido por fragmentação e transcrição reversa. As regiões 5' e 3' dos fragmentos de cDNA são em seguida preparados para permitir a ligação eficiente dos adaptadores "Y", que contêm um código de barras único e sítios de ligação de primers. Finalmente, os cDNAs ligados são amplificadas por PCR, e estão prontos para o sequenciamento. Fonte: http://www.labome.com/method/RNA-seq-Using-Next-Generation-sequencing.html	39
Figura 4	Padrão de informações presentes nos arquivos com formato .fastq. Figura obtida de Cock et al., 2010.....	41
Figura 5	Representação da qualidade de <i>reads</i> obtida pelo uso da ferramenta fastQC. Figura obtida de: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc/fastqc_report.html	42
Figura 6	Em um experimento envolvendo duas condições, as <i>reads</i> são mapeadas contra o genoma com TopHat. As <i>reads</i> para cada replicata biológica são mapeados de forma independente. Estas <i>reads</i> mapeadas são fornecidos como entrada para o programa Cufflinks, que produz um arquivo de transfrags montadas para cada repetição. Os arquivos de montagem são mescladas, por Cuffmerge, com a anotação do transcrito de referência, em uma anotação unificada para análise posterior. Esta anotação mesclada é quantificada para cada condição por Cuffdiff, que produz dados de expressão em um conjunto de arquivos tabulares. Esses arquivos são indexados e visualizado com CummeRbund. Para facilitar a exploração de genes identificados por Cuffdiff como genes diferencialmente expressos, apresentando splicing alternativo ou sendo transcricionalmente regulados. Fonte: Trapnell et al., 2012.....	44
Figura 7	A pipeline de TopHat. <i>Reads</i> de RNA-Seq são mapeadas contra o genoma de referência, e aquelas <i>reads</i> que não são mapeadas são separadas. Um consenso inicial das regiões mapeadas é calculado utilizando o modulo de montagem Maq (Li et al., 2008). Sequências de flanqueamento com potencial de doador / receptor de locais de splicing dentro de regiões vizinhas são unidas para formar potenciais junções de processamento alternativo. As <i>reads</i> inicialmente não mapeadas (IUM) são indexadas e alinhadas para estas sequências de junção de emendas. Fonte: Trapnell, et al., 2009.....	47
Figura 8	Esquema mostrando as combinações par a par analisadas para obtenção dos genes diferencialmente expressos entre cada combinação. A origem da seta indica a primeira condição e o destino da seta indica a segunda condição utilizada na	

comparação. nippon_ctrl: cultivar Nipponbare na condição controle; nippon_trat: cultivar Nipponbare na condição tratamento por frio; bulk_nippon_ctrl: *bulk* de RILs mais próximos da cultivar Nipponbare, na condição controle; bulk_nippon_trat: *bulk* de RILs mais próximos da cultivar Nipponbare, na condição tratamento; ata_ctrl: cultivar BRS Atalanta na condição controle; ata_trat: cultivar BRS Atalanta na condição tratamento por frio; bulk_ata_ctrl: *bulk* de RILs mais próximos da cultivar BRS Atalanta, na condição controle; bulk_ata_trat: *bulk* de RILs mais próximos da cultivar BRS Atalanta, na condição tratamento..... 49

Figura 9 Figura 11: Gráfico do tipo boxplot onde são apresentados os valores mínimos e máximo, os quartis superiores e inferiores, e a mediana para cada uma das 24 bibliotecas analisadas. No eixo y estão representados os valores de expressão, após a normalização por meio do método FPKM, em escala log10. No eixo x são apresentadas cada condição com suas replicatas biológicas. nip_trat_0 equivale a repetição biológica 1. nip_trat_1 equivale a repetição biológica 2 e nip_trat_2 equivale a repetição biológica 3. nip_trat: Nipponbare tratamento; ata_trat: BRS Atalanta tratamento; bnip_trat: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição tratamento; bata_trat: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição tratamento; nip_ctrl: Nipponbare controle; ata_ctrl: BRS Atalanta controle; bnip_ctrl: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição controle; bata_ctrl: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição controle. Criado com o pacote CummeRbund.....52

Figura 10 Dendrograma de distâncias de Jensen-Shannon mostrando as distâncias entre as condições em estudo. nip_trat: Nipponbare tratamento; ata_trat: BRS Atalanta tratamento; bnip_trat: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição tratamento; bata_trat: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição tratamento; nip_ctrl: Nipponbare controle; ata_ctrl: BRS Atalanta controle; bnip_ctrl: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição controle; bata_ctrl: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição controle. Criado com o pacote CummeRbund.....57

Figura 11 Diagrama de Venn apresentando genes subexpressos para a comparação entre a condição controle e a condição tratamento para os genótipos BRS Atalanta, Nipponbare e para os *bulks* de RILs que se mostraram mais próximos de cada genitor. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento..... 60

Figura 12 Diagrama de Venn apresentando genes superexpressos para a

comparação entre a condição controle e a condição tratamento para os genótipos BRS Atalanta, Nipponbare e para os *bulks* de RILs que se mostraram mais próximos de cada genitor. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento..... 61

Figura 13 Fatores de transcrição superexpressos envolvendo a condição controle e a condição tratamento para as combinações: AC x AT, BAC x BAT, BNC x BNT e NC x NT. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento.....62

Figura 14 Fatores de transcrição subexpressos envolvendo a condição controle e a condição tratamento para as combinações: AC x AT, BAC x BAT, BNC x BNT e NC x NT. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento.....63

Figura 15 Anotação da Ontologia gênica (GO) para os genes diferencialmente expressos na combinação AC x AT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação AC x AT..... 71

Figura 16 Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação NC x NT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação NC x NT.....73

- Figura 17 Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação BNC x BNT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação BNC x BNT..... 74
- Figura 18 Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação BAC x BAT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação BAC x BAT..... 76
- Figura 19 Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação AT x BAT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação AT x BAT..... 78

Lista de Tabelas

Tabela 1	Lista de <i>softwares</i> , com a versão, autor e link, utilizados nas análises realizadas neste trabalho.....	50
Tabela 2	Número de reads usadas pelo programa TopHat para realizar o alinhamento com o genoma e o transcrito de referência. Total de reads representa o input no programa, número de reads mapeadas e porcentagem frente ao total de reads, alinhamentos únicos representam reads que alinham em um único local no genoma e a porcentagem frente ao total de reads. Alinh. multi-posições e a sua porcentagem representam reads que alinham em mais de um local no genoma. Leituras não mapeadas e sua porcentagem representam reads que não alinham com o genoma e o transcrito de referência.....	53
Tabela 3	Genes diferencialmente expressos entre as diversas combinações. Número total de genes diferencialmente expressos obtidos com o uso do protocolo Tuxedo (Trapnell et al., 2012). Número de genes que apresentaram uma expressão diferencial com log2-fold-change maior ou igual a 1,5 e número de genes que apresentam um nível de expressão diferencial com log2-fold-change menor ou igual a -1,5. Número de genes que apresentam identificação própria (ID), por meio do padrão Osxxgxxxxxx, para o genoma depositado no banco de dados de arroz RAP-DB.....	58

Lista de Abreviaturas e Siglas

DNA	Ácido desoxirribonucleico
RNA	Acido ribonucleico
mRNA	RNA mensageiro
tRNA	RNA total
CRW	Arroz selvagem comum
RBP	Proteínas de ligação ao RNA
GA	Ácido giberélico
ABA	Ácido abscísico
rRNA	RNA ribossômico
UTR	Regiões não traduzidas
cDNA	DNA complementar
NGS	Sequenciadores de nova geração
SBS	Sequenciamento por síntese
dNTP	Desoxirribonucleotídeos fosfatados
ASCII	Código Padrão Americano para o Intercâmbio de Informação
SNP	Polimorfismo de nucleotídeo único
RIL	Linhas endogâmicas recombinantes
NaClO	Hipoclorito de sódio
DBC	Delineamento de blocos casualizados
RPKM	Leituras por quilo base por milhão de <i>reads</i> mapeadas
FPKM	Fragmentos por quilo base por milhão de fragmentos mapeados
IUPAC	União Internacional de Química Pura e Aplicada
GO	Ontologia Gênica
HSICT	Tolerância a baixas temperaturas induzida por choque térmico
HSF	Fatores de choque térmico
GBS	Genotipagem por sequenciamento
AC	BRS Atalanta controle
AT	BRS Atalanta tratamento
NC	Nipponbare controle
NT	Nipponbare tratamento
BAC	<i>Bulk</i> de RILs BRS Atalanta controle
BAT	<i>Bulk</i> de RILs BRS Atalanta tratamento
BNC	<i>Bulk</i> de RILs BRS Nipponbare controle
BNT	<i>Bulk</i> de RILs BRS Nipponbare tratamento
t	Toneladas
TF	Fator de transcrição

Sumário

1 Introdução.....	14
1.1 A cultura do arroz.....	14
1.2 Estresse por frio em arroz.....	15
1.3 Germinação e estágios iniciais de desenvolvimento do arroz.....	18
1.4 Tolerância a frio em arroz.....	20
1.5 Mecanismos envolvidos na tolerância ao frio em plantas.....	24
1.6 Análise de transcritomas.....	26
1.6.1 O processamento do RNA mensageiro.....	26
1.6.2 O transcritoma em eucariotos.....	29
1.6.3 Sequenciadores de nova geração (NGS).....	30
1.7 Análise dos dados obtidos por meio de RNA-Seq.....	33
2 Objetivos.....	34
3 Material e Métodos.....	34
3.1 Material genético.....	34
3.2 Tratamento por frio.....	36
3.2.1 Condição controle.....	37
3.2.2 Condição tratamento.....	37
3.3 Delineamento experimental.....	37
3.4 Extração de RNA total.....	37
3.5 Preparação das bibliotecas de cDNA.....	38
3.6 Sequenciamento das bibliotecas de cDNA.....	40
3.7 Análise dos dados obtidos pelo sequenciamento das bibliotecas.....	42
4 Resultados e discussão.....	51
5 Conclusões.....	79
6 Referências bibliográficas.....	80
Apêndice A. Lista dos TFs superexpressos nas 12 combinações analisadas...	94
Apêndice B. Lista dos TFs subexpressos nas 12 combinações analisadas.....	100

1. Introdução

1.1 A cultura do arroz

O arroz (*Oryza sativa* L.) é uma das espécies mais importantes para a humanidade uma vez que alimenta, diariamente, mais de 3,5 bilhões de pessoas, sendo que estas dependem do arroz para suprir mais de 20% de sua ingestão diária de calorias (IRRI, 2010). Mais de 90% do arroz produzido no mundo é oriundo da Ásia, sendo que a China e a Índia representam quase a metade desse total. No ano de 2012, a produção mundial do cereal foi de 718 milhões de toneladas em uma área de aproximadamente 163 milhões de hectares. Os maiores produtores mundiais são: China (206 milhões de toneladas), Índia (152 milhões de t), Indonésia (69 milhões de t), Vietnã (43 milhões de t), Tailândia (38 milhões de t), Bangladesh (34 milhões de t), Mianmar (33 milhões de t), Filipinas (18 milhões de t), Brasil (11,3 milhões de t) (FAO, 2013). O Brasil é o nono maior produtor mundial e coloca-se como o maior produtor fora do continente asiático. No Brasil, na safra 2012/2013 foram colhidas 11,3 milhões de toneladas, numa área de 2,39 milhões de hectares. Os estados do Rio Grande do Sul e Santa Catarina são os principais estados produtores, sendo responsáveis por, aproximadamente, 60% e 10% da produção nacional, respectivamente. O cultivo nestes dois estados é, basicamente, irrigado sob inundação proporcionando a obtenção de produtividades médias superiores a sete toneladas por hectare. Esta produtividade é muito superior ao obtido no cultivo de sequeiro, que consiste no sistema de produção adotado nas demais regiões produtoras de arroz do país (CONAB, 2014). No mundo, cerca de 85-90 milhões de hectares de arroz são cultivados em áreas de várzea, sob irrigação, fornecendo 75% da produção mundial (IRRI, 2010).

As previsões de aumento na demanda por arroz continuam elevadas. Para o ano de 2035, um adicional de 116 milhões de toneladas de arroz será necessário para alimentar as populações que continuam crescendo. Na África, onde o arroz é a fonte de alimento cujo consumo mais rapidamente cresce, cerca de 30 milhões de toneladas a mais de arroz serão necessários em 2035, representando um aumento

de 130% no consumo de arroz a partir de 2010. A nível mundial, a produção deste cereal terá de aumentar cerca de 8-10 milhões de toneladas por ano ao longo da próxima década o que, mantendo a atual área de cultivo, exigirá um aumento anual na produtividade de cerca de 1,2-1,5%, equivalente a um aumento de rendimento de 0,6 t ha⁻¹ em todo o mundo (Sech et al., 2012).

Os programas de melhoramento genético de arroz aplicam um esforço contínuo que visa a redução das disparidades de rendimento causadas por restrições bióticas e abióticas, além de manter ou mesmo aumentar o potencial de rendimento das novas cultivares por meio do uso de técnicas de melhoramento clássico e de novas técnicas que podem consistir do uso de seleção assistida por marcadores para atualizar variedades existentes com um ou mais genes que lhes dará uma vantagem nas condições de desenvolvimento presentes nas áreas de cultivo. As mudanças climáticas prognosticadas podem ameaçar seriamente os níveis de produção necessários para alimentar as futuras gerações na Ásia e em outros continentes. As alterações climáticas têm muitas facetas, incluindo mudanças nas tendências de longo prazo na temperatura e nos regimes de chuvas, bem como no aumento da variabilidade de eventos extremos. As mudanças climáticas irão agravar uma série de estresses para as plantas de arroz, entre as quais podem ser citados o calor, o frio, a seca, a salinidade e a submersão (Wassmann, et al., 2009). A melhoria nos níveis de tolerância a esses estresses abióticos tem sido um dos principais objetivos dos diferentes componentes da cadeia produtiva do arroz.

1.2 Taxonomia da espécie *Oryza sativa* L.

A espécie *Oryza sativa* L. pertence à divisão Magnoliophyta, classe Liliopsida, subclasse Commelinidae, ordem Cyperales, família Poaceae, subfamília Pooideae, a tribo Oryzae e ao gênero *Oryza* (USDA, 2006). A adaptação às condições aquáticas ou de inundação parece ter sido o principal fator na evolução da tribo Oryzae, uma tribo isolada de gramíneas com cerca de quatorze gêneros (Chandraratna, 1964).

Existem dois tipos distintos de arroz domesticado: *Oryza sativa*, também chamado de arroz asiático e *Oryza glaberrima*, conhecido como arroz africano, sendo que ambos possuem histórias únicas de domesticação. O gênero *Oryza*

contém 21 parentes selvagens dos arroz domesticados (Vaughan et al., 2003). O gênero é dividido em quatro complexos de espécies, sendo eles: *O. sativa*, *O. officinalis*, *O. ridelyi* e *O. granulata*. Todos os membros do gênero *Oryza* tem $n=12$ cromossomos e enquanto cruzamentos interespecíficos são possíveis dentro de cada complexo, é difícil recuperar descendentes férteis a partir de cruzamentos entre espécies pertencentes a complexos diferentes (Vaughan et al., 2003). O complexo *O. sativa* contém duas espécies domesticadas: *O. sativa* e *O. glaberrima*, além de cinco ou seis espécies selvagens: *O. rufipogon*, *O. nivara* (também considerado um ecótipo de *O. rufipogon*), *O. barthii*, *O. longistaminata*, *O. meridionalis* e *O. glumaepatula*, sendo que todas elas são espécies diplóides. *Oryza sativa* é distribuída globalmente, com uma alta concentração na Ásia enquanto *O. glaberrima* é cultivada no oeste da África. *Oryza rufipogon* pode ser encontrado na Ásia e na Oceania. *Oryza barthii* e *O. longistaminata* são espécies africanas. *O. barthii* é endêmica no oeste da África e *O. longistaminata* é encontrada por toda a África. *Oryza meridionalis* é nativa da Austrália e *O. glumaepatula* é endêmica na América do Sul e na América Central. Dada esta distribuição é possível localizar os *pools* ancestrais dos quais o arroz moderno foi extraído. Os genótipos africanos foram domesticados de *O. barthii* (formalmente chamado de *O. breviligulata*) e *O. sativa* foi domesticada de *O. rufipogon* (Sweeney e McCouch, 2007).

A espécie *Oryza sativa* foi domesticada do seu comumente reconhecido progenitor, o arroz selvagem comum (CRW) *O. rufipogon* Griff. (Chang 1976; Yamanaka et al. 2004). Ao contrário do bem conhecido processo de evolução e domesticação do milho moderno (Doebley, 2004), a origem de *O. sativa* tem sido um problema não resolvido por décadas. O debate de longa data diz respeito a se o arroz asiático cultivado se originou monofilética ou polifileticamente. A hipótese da origem monofilética sugere que as populações selvagens de *O. rufipogon* poderiam ter evoluído para a subespécie *indica* no Sul e/ou Sudeste da Ásia, a partir do qual a subespécie *japonica* evoluiu através da adaptação durante sua dispersão para maiores latitudes e também maiores altitudes (Chang, 1976; Lu et al., 2002). Em contraste, a hipótese polifilética postula que a subespécie *indica* e a subespécie *japonica* foram originadas a partir de diferentes populações ancestrais do CWR (Yamanaka et al, 2004; Londo et al 2006). Uma terceira teoria considera que a domesticação do arroz foi um processo dinâmico (Sang e Ge, 2007), onde ocorreu uma domesticação independente entre as subespécies *indica* e *japonica*, seguida

por uma contínua troca de genes entre os diversos grupos de *O. sativa* (Yang et al., 2011; Sweedey et al., 2007; Zhao et al., 2010) e também entre *O. sativa* e *O. rufipogon*, facilitada pela ausência de uma aparente barreira reprodutiva entre elas (Oka, 1988).

As subespécies *indica* e *japonica* se diferenciam por uma série de características morfológicas, fisiológicas e genéticas. Quando tipos *indica* e *japonica* típicos foram comparados, muitos genes e caracteres mostraram uma associação não casualizada com cada um para claramente separar as duas subespécies (Morishima, 2001). Em todos os níveis de análise, as diferenças entre as subespécies *indica* e *japonica* são muito aparentes. Diferenças entre estas duas subespécies são documentados na literatura chinesa, já em 100 antes de Cristo (Matsuo et al., 1997). Em termos ecogeográficos, a subespécie *indica* é primeiramente conhecida como arroz de várzea que é cultivado em toda a Ásia tropical, enquanto *japonica* é normalmente encontrada em regiões temperadas da Ásia, nas regiões de montanha do sudeste e nas elevações no sul da Ásia. As características que têm sido utilizadas para classificar *japonica* e *indica* incluem o formato dos grãos, a reação de fenol, a sensibilidade ao clorato de potássio, a cor da folha, e o comprimento do apículo da folha, embora o espectro de variação de qualquer uma dessas características individuais sobreponha-se nas duas subespécies (Oka, 1988; Garris et al., 2005).

Além disso, caracteres morfológicos permitem diferenciar os indivíduos entre as duas subespécies destacando-se o formato do grão, sendo que em geral indivíduos *japonica* possuem grãos curtos, largos, espessos e seção transversal arredondada. Os indivíduos *indica* possuem grãos longos, estreitos e levemente planos (Chandraratna, 1964; Watanabe, 1997). Os caracteres agrônômicos podem ser considerados na diferenciação entre as duas subespécies. A ausência de arista, fácil debulha, cor verde claro das plantas, grande número de afilhos, afilhos com hábito de crescimento mais aberto e sensibilidade a baixas temperaturas são caracteres encontrados na subespécie *indica*. Já a subespécie *japonica* se caracteriza pela presença de arista, resistência a debulha, cor verde escura das plantas, menor número de afilhos, afilhos com hábito de crescimento mais ereto e tolerância a baixas temperaturas (Takahashi, 1984).

1.3 Germinação e estágios iniciais de desenvolvimento do arroz

A semente é uma forma condensada de planta. Na presença de condições ambientais adequadas, ela pode retomar a atividade metabólica do estado de repouso fisiológico, e mobilizar as reservas, biossintetizar novas proteínas, regenerar organelas e membranas celulares e, finalmente, projetar a radícula e se estabelecer como plântula (He e Yang, 2013). Sementes secas de arroz contêm mais de 17.000 mRNAs armazenados. As proteínas de ligação a RNA (RBP) desempenham um papel importante em manter a estabilidade e a regulação das funções destes mRNAs de vida longa (Howell et al., 2009). As características hereditárias das sementes são o principal fator que influencia a germinação. As cultivares de arroz híbrido geralmente têm alto vigor na germinação. A comparação dos perfis do proteoma de embrião maduro de arroz foi conduzida em cultivar híbrida e as suas linhas genitoras. A maioria das proteínas de reserva exibiu efeitos gênicos de sobredominância e proteínas induzidas por estresses exibiram efeitos aditivos, o que pode contribuir para a tolerância a estresses bióticos e abióticos (Ge et al., 2008).

Por definição, a germinação compreende os eventos que iniciam com a absorção de água pelas sementes quiescentes e termina com o alongamento do eixo embrionário (Bewley e Black, 1994). O sinal visível de que a germinação está completa é, geralmente, a penetração das estruturas em torno do embrião pela radícula. O resultado é muitas vezes chamado de germinação visível. Eventos subsequentes, incluindo a mobilização das substâncias de reserva, estão associados com o crescimento da plântula. Praticamente todos os eventos celulares e metabólicos que são conhecidos por ocorrer antes da conclusão da germinação de sementes não-dormentes também ocorrem em sementes dormentes que foram embebidas em água. Na verdade, as atividades metabólicas das sementes dormentes são frequentemente apenas sutilmente diferentes daquelas de sementes que não apresentam dormência. Assim, uma semente dormente pode atingir praticamente todos os passos metabólicos necessários para completar a germinação, contudo por algum motivo desconhecido, o eixo embrionário (isto é, a radícula) não se alonga (Bewley, 1997).

A germinação das sementes é um processo fisiológico e bioquímico complexo que envolve uma série de transdução de sinais e regulação da expressão gênica. Nela, os fitormônios presentes nas sementes apresentam um papel decisivo. Uma série de fatores exógenos e endógenos participam na regulação da germinação, incluindo a água, a temperatura, a luz, o ritmo circadiano, e fitormônios (Penfield et al, 2005; Holdsworth et al, 2008). Os ácidos giberélico (GA) e abscísico (ABA) são os dois principais fitormônios que antagonicamente regulam a germinação das sementes pela modulação de proteínas funcionais GA e ABA-responsivas. GA pode superar a dormência e induzir a germinação de sementes, enquanto ABA pode induzir à dormência na fase posterior a maturação das sementes. As vias GA e ABA podem interagir uma com a outra, inibindo diretamente a síntese de algumas proteínas ou regular as proteínas-alvo comuns, tais como as proteínas DELLA (He e Yang, 2013). A dormência das sementes nas cultivares de arroz atualmente cultivado é reduzida ou mesmo nula. Em algumas cultivares, a semente madura pode germinar nas panículas, sendo esta característica denominada de germinação na pré-colheita, desde que a semente encontre condições climáticas adequadas. A ocorrência de germinação na pré-colheita reduz a produção e a qualidade das sementes. Genes relacionados com a síntese de ABA tem provado seu envolvimento na germinação pré-colheita de sementes de arroz (Fang e Chu, 2008).

Nas sementes de arroz, o endosperma, composto basicamente por amido, ocupa grande parte do seu volume e serve como reservatório de nutrientes. Ele é cercado por uma camada de aleurona e na parte basal encontra-se o embrião. Entre o endosperma e embrião situa-se o escutelo, que pode ser considerado como uma forma alternativa dos cotilédones. O embrião e o endosperma desempenham papéis diferentes na germinação das sementes de arroz. O embrião contém a maior parte da informação genética que controla a germinação. Após a embebição das sementes, a presença de substrato e a falta de energia para a sua manutenção irão ativar o embrião e este passará a produzir fitormônios, principalmente o ácido giberélico (GA). Este fitormônio, que se apresenta basicamente na forma gasosa, pode se difundir pela camada de aleurona e iniciar uma cascata de sinalização que leva à síntese de α -amilases e outras enzimas hidrolíticas (He e Yang, 2013). Estas enzimas serão secretadas no endosperma para conduzir à degradação de compostos de armazenamento, incluindo o amido, lipídeos e proteínas para o estabelecimento das plântulas (Jacobsen et al, 1995; Bethke et al, 1997).

1.4 Tolerância a frio em arroz

O arroz é o alimento básico para metade da população humana. Ao contrário de outros cereais, tais como trigo e cevada, as plantas de arroz são sensíveis ao estresse por frio, o que muitas vezes resulta na diminuição da produtividade, especialmente em regiões onde a subespécie *indica* é cultivada. A tolerância ao frio durante a germinação é importante para garantir um rápido e uniforme estabelecimento da cultura no início da estação de cultivo. No entanto, a avaliação desta característica em condições de campo é limitada pela variação do ambiente, o que torna difícil identificar linhas geneticamente superiores. A avaliação da tolerância ao frio sob condições controladas de temperatura pode ser realizada por meio da avaliação da percentagem de redução no comprimento do coleóptilo e crescimento do coleóptilo (da Cruz et al., 2006). A germinação é um processo fisiológico complexo que é afetado pela temperatura e pelo potencial de água do solo (Alvarado e Bradford, 2002). Este processo inclui os eventos de ativação do embrião a partir de um estágio dormente para uma forma dinâmica. A germinação é um processo importante no estabelecimento do número final de plantas na lavoura e uma densidade ótima é conseguida quando as sementes são completamente saudáveis e germinadas dentro de uma taxa adequada (Albuquerque et al., 2003).

As baixas temperaturas podem ter impactos negativos sobre as plantas de arroz durante a germinação, crescimento vegetativo e estágios reprodutivos. O arroz é cultivado em quase todo o mundo, desde a latitude 50°N até 40°S e com altitudes que vão do nível do mar até 3000m (Juliano, 1993), fazendo com que a tolerância ao frio seja uma característica necessária em muitas destas regiões. Anualmente, cerca de 15 milhões de hectares de lavouras de arroz tem sua produção ameaçada pela ocorrência de frio (Zhang et al., 2014).

No Brasil, a cultura pode sofrer danos causados pela ocorrência de frio principalmente no estado do Rio Grande do Sul. Dentre os estados produtores, este é o estado que apresenta as maiores latitudes. Nesta região, bem como no Uruguai e na Argentina, o arroz *indica* é responsável pela quase totalidade do arroz cultivado. As baixas temperaturas podem reduzir em 25% o rendimento final da cultura (Bierlen

et al. 1997; Clayton e Neves 2011; Lima et al. 2012). Para o arroz, temperaturas inferiores a 20°C diminuem a velocidade e a porcentagem de germinação (Yoshida, 1981), resultando em menor número de plantas por metro quadrado e custos de produção mais elevados. No sul do estado brasileiro do Rio Grande do Sul, a temperatura mínima média em Outubro, quando o arroz é semeado, é de cerca de 12 °C, sendo necessário tolerância ao frio para garantir um estabelecimento rápido e uniforme da cultura do arroz. Experimentos como o de Tilebeni et al. 2012 não obtiveram germinação quando a semente foi submetida a temperaturas de 12°C. Porém, a temperatura inferior crítica para a germinação pode ser extremamente genótipo-dependente.

As baixas temperaturas podem ser diferenciadas de congelamento em termos da gama de temperaturas que provocam os danos relacionados (Iba, 2002). Muitas espécies de origem tropical ou subtropical são danificadas ou mortas por temperaturas baixas mas não congelantes, e apresentam vários sintomas de injúria por frio, tais como clorose, necrose, ou retardo do crescimento. Em contraste, as espécies tolerantes ao frio são capazes de crescer na presença de baixas temperaturas (Sanghera et al., 2011). Os sintomas da sensibilidade ao frio e os danos podem variar de acordo com a fase de desenvolvimento da planta do arroz (Yoshida 1981). Na fase de germinação, os sintomas mais comuns são o atraso e uma menor porcentagem de germinação (da Cruz e Milach 2000). Em geral, os genótipos pertencentes à subespécie *japonica* apresentam maior tolerância ao frio do que a subespécie *indica*, mas trabalhos constataram variabilidade dentro destas subespécies (da Cruz e Milach, 2004).

Melhoramento de plantas exige a existência de variabilidade genética. Neste sentido, *Oryza sativa* L. tem grande capacidade de adaptação ao frio e ecótipos tolerantes estão disponíveis para utilização em hibridações artificiais. A subespécie *indica* inclui cultivares melhor adaptadas a ambientes tropicais, como Índia, China e Indonésia, enquanto cultivares *japonica* são mais adaptados a climas temperados, como os do Japão, Coreia e das ilhas Java (Takahashi, 1984).

O grau da lesão no arroz depende geralmente do período de ocorrência (fase de crescimento), da intensidade do frio, e da duração das baixas temperaturas (Li et al. 1981). Baixa temperatura tem o potencial de afetar o crescimento e desenvolvimento de plantas de arroz em qualquer fase de desenvolvimento, desde a germinação até o enchimento de grãos (Ye et al. 2009). No entanto, Yoshida (1981)

mostrou que a sensibilidade ao frio varia entre os estágios de desenvolvimento. O limite inferior da temperatura nos estádios de germinação e nos estádios vegetativos situam-se entre 10 e 13°C. Já no estágio reprodutivo esse limite inferior de temperatura situa-se entre 18 e 20°C. Dessa forma percebe-se que as plantas de arroz apresentam uma maior sensibilidade ao frio durante a fase reprodutiva.

Estudos com grande número de cultivares pertencentes a estas duas subespécies mostraram que os genótipos *japonica* tem maior grau de tolerância ao frio na fase de germinação (Mertz et al. 2009), bem como nas fases vegetativa e reprodutiva (Li et al, 1981; Mackill e Lei 1997). da Cruz e Milach (2004) também concluíram que genótipos *japonica* apresentam maior tolerância ao frio no estágio de germinação do que genótipos *indica*, embora eles encontrassem variabilidade para essa característica dentro de ambas as subespécies. Isto corrobora com resultados anteriores de que alguns genótipos de *indica* de regiões com alta latitude podem apresentar nível moderado de tolerância ao frio (Jennings et al., 1979). Estudos também demonstraram que algumas variedades pertencentes a subespécie *javanica* (também caracterizada como uma sub-população tropical ou um ecótipo de *japonica*) apresentam tolerância ao frio (Sweeney e McCouch 2007; Brar e Singh, 2011; Li et al., 2012). Por meio de cruzamentos artificiais vários genes que conferem tolerância a frio, presentes nesta subespécie, foram transferidos para linhagens da subespécie *japonica* (Saito et al., 2001).

Correlações de tolerância ao frio entre os diferentes estágios de crescimento para o arroz foram relatados, e foi sugerido que variedades com alta germinação e vigor de plântulas sob condições de baixa temperatura também são mais propensos a serem mais tolerantes à exposição a baixa temperatura na diferenciação floral e na fase de floração (Ye et al. 2009). Na fase de germinação, a tolerância ao frio em arroz apresenta uma herança complexa (da Cruz et al., 2013). da Cruz et al. (2006a) determinaram a herança e a herdabilidade de tolerância ao frio no estágio de germinação em cruzamentos entre seis genótipos de arroz. Eles mostraram que ações gênicas aditivas e não aditivas estavam envolvidas, mas que o componente não-aditivo foi relativamente mais importante para a porcentagem de redução no comprimento do coleóptilo e no crescimento do coleóptilo. A epistasia, um efeito não-aditivo, também mostrou-se importante para a capacidade de germinação de arroz em condições de baixa temperatura (Chen et al., 2006).

Considerando-se as elevadas estimativas de herdabilidade que foram obtidas

para germinação à baixas temperaturas em arroz (Sthapit e Witcombe 1998; da Cruz et al. 2006a), a seleção de genótipos tolerantes ao frio é provável que seja bem sucedida. No entanto, a grande importância dos efeitos não aditivos sugere que a seleção deveria ser aplicada em gerações avançadas de programas de melhoramento (F4 ou F5), quando os efeitos de dominância foram diminuídos devido ao aumento da homozigose dos indivíduos (da Cruz et al., 2013). No entanto, os principais avanços têm sido obtidos nas cultivares *japonica*. Portanto, o desafio ainda continua sendo desenvolver cultivares tipo *indica* com tolerância ao frio, adequada para o regiões de elevada latitude, como o sul do Brasil, onde os genótipos *indica* são mais produtivos que os genótipos *japonica*, apresentam tipo de grão, composição de grão e qualidade de grãos adaptados aos mercados-alvo, mas são propensos a danos causados pelo frio em diversos estádios de desenvolvimento da cultura (Bierlen et al. 1997). Uma solução para esta questão pode residir no cruzamento entre genótipos contrastantes para esta característica, ou seja, cruzar genótipos *indica* com genótipos *japonica*, a fim de transferir genes para tolerância ao frio presentes na subespécie *japonica* para a subespécie *indica*. No entanto, as diferenças entre estes dois grupos de arroz torna difícil manter as características agrônômicas que caracterizam cada subespécie tais como o formato do grão, a qualidade de cocção, a relação amilose/amilopectina entre outros, extremamente necessários para a sua aceitação pelo consumidor.

Após a emergência das plântulas, a incidência de frio pode continuar a trazer problemas para a cultura. O desenvolvimento de variedades que apresentem tolerância ao frio é a maneira mais econômica e eficaz para evitar os danos causados por este estresse ao arroz. No entanto, o melhoramento para tolerância ao frio na fase de plântulas tem sido difícil devido à sua natureza poligênica e conhecimento inadequado sobre a base genética da tolerância ao frio. Baixas temperaturas na fase de plântula pode levar à descoloração ou amarelecimento, enrolamento da folha e murcha, redução do crescimento, retardo na maturação, problemas no estabelecimento das plantas e, conseqüentemente, redução na produtividade da cultura (Zhang et al., 2014). A fim de desenvolver cultivares de arroz que apresentem um adequado estabelecimento de plântulas sob baixas temperaturas, é necessário investigar os recursos genéticos provenientes de várias regiões do mundo, fornecendo conhecimentos básicos para permitir a introdução de genes envolvidos na capacidade de germinação em baixa temperatura de cultivares

e variedades tradicionais em cultivares elite (Miura et al., 2001).

1.5 Mecanismos envolvidos na tolerância ao frio em plantas

Estresses abióticos afetam negativamente o crescimento, a produtividade e desencadeiam uma série de alterações fisiológicas, morfológicas, bioquímicas e moleculares nas plantas. Baixas temperaturas tem um enorme impacto sobre a sobrevivência e distribuição geográfica das plantas. O efeito dessas baixas temperaturas sobre o metabolismo celular dependem da intensidade e da duração do estresse. Como resposta da planta a esta condição estão mudanças no seu padrão de expressão gênica e nos produtos proteicos (Sanghera et al., 2011). A tolerância às baixas temperaturas nas plantas é uma característica muito complexa, envolvendo muitas vias metabólicas e diferentes compartimentos celulares (Hannah et al., 2005).

Entre os mecanismos envolvidos na tolerância ao frio podem ser citados: estabilidade (Steponkus et al., 1993) e fluidez da membrana celular (Orvar et al., 2000); eficiência do fotossistema II (PSII) (Routaboul et al., 2000); acúmulo de açúcares (Strauss e Hauser, 1986), síntese de osmoprotetores, entre inúmeros outros. Além disso, por meio de transformação genética, a participação de inúmeros genes na tolerância a frio foi documentada e estão reunidos em revisão de Sanghera et al. (2011). Uma classe de genes que está recebendo grande atenção devido a sua grande importância na regulação da expressão gênica são os TFs. Genomas de plantas atribuem aproximadamente 7% das suas sequências de codificação para TF, o que comprova a complexidade da regulação transcricional (Udvardi et al., 2007). Os TFs são moléculas adaptadoras que detectam sequências regulatórias no DNA e visam à montagem de complexos de proteínas que controlam a expressão dos genes (Zaret e Carroll, 2011). Como moduladores da expressão gênica estes TFs agem em todos os sistemas bioquímicos eucarióticos, dirigindo "redes" ou "programas de regulação" que definem os estágios de desenvolvimento e zelam pela manutenção das células em microambientes que mudam dinamicamente (Yusuf, et al., 2012). Os fatores de transcrição interagem com cis-elementos das regiões promotoras dos vários genes relacionados com o estresse e, assim, sobrerregulam a expressão de muitos genes a jusante, resultando na transmissão de tolerância ao

estresse abiótico (Agarwal e Jha, 2010). Liu et al. (2012) desenvolveram um modelo hipotético de resposta ao frio não congelante a partir de resultados obtidos com linhas introgridas desenvolvidas a partir de *Solanum lycopersicum* e *Solanum habrochaites* (Figura 1).

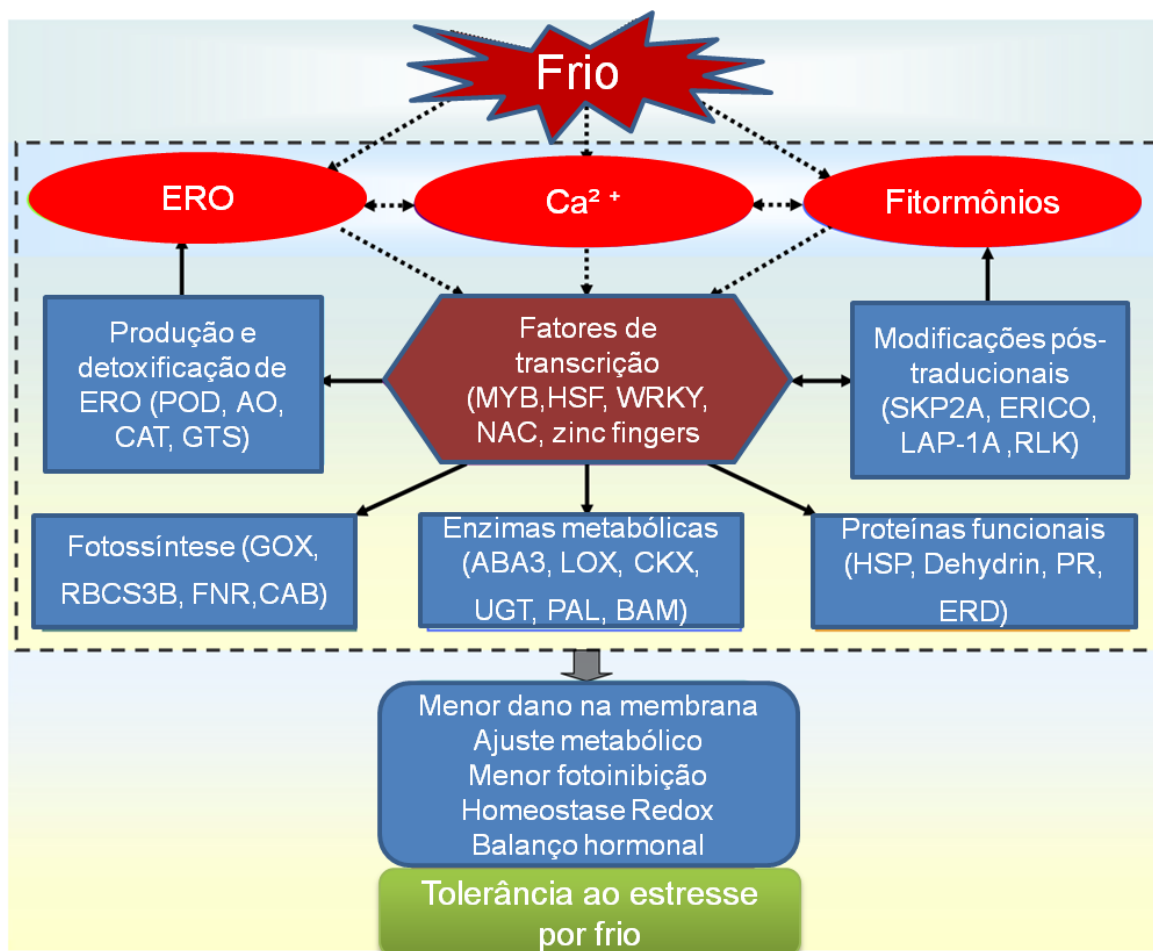


Figura 1. Modelo hipotético da tolerância ao frio não-congelante. Espécies reativas de oxigênio, cálcio e fitormônios como moléculas de sinalização podem desempenhar um papel crítico na adaptação de plantas ao estresse pelo frio. A modulação destas vias de sinalização ativa a regulação da transcrição. A expressão diferencial de fatores de transcrição faz com que muitos grupos funcionais se mostrem diferencialmente expressos entre genótipos tolerantes e sensíveis, tais como proteínas funcionais, modificações pós-traducionais de proteínas, e genes associados com processos fisiológicos e metabólicos, como a fotossíntese, a produção de ERO ou sistema de detoxificação e enzimas metabólicas. A modulação da expressão dos genes associados com modificações pós-traducionais, ERO, e fitormônios, em seguida, retroalimentou e fez o ajuste fino de fatores de transcrição, ERO, e redes de sinalização de fitormônios. ERO: espécies reativas de oxigênio, AO: amina oxidase, PODs: peroxidases, GST: glutathione S- transferases, CAT: catalase, RBCS3B: pequena subunidade 3B de rubisco, FNR: ferredoxina-NADP +- oxirredutase; CABS: proteínas de ligação de clorofila a/b; GOXs: glicolato oxidases, LOXs: lipoxigenases, ABA3: ABA deficiente 3, UGTs: UDP- glucosiltransferases, CKX: citocinina oxidase/desidrogenase, PAL: fenilalanina amônia-liase, PRs, proteínas relacionadas à patogênese, BAM: beta-amilase ; Hsps: proteínas de choque de calor, ERD: resposta precoce à desidratação, HSFS: fatores de choque térmico, SKP2A: proteína 2A associada a quinase fase S, RLK: receptor proteína quinase, LAP-A1: leucina aminopeptidase A1. Adaptado de Liu, et al., (2012)

ERO, cálcio e fitormônios como moléculas de sinalização podem desempenhar um papel crítico na adaptação ao estresse pelo frio. A modulação dessas vias de sinalização pode ativar diretamente genes relacionados à resposta e interagir direta ou indiretamente com várias outras redes de sinalização para regular a transcrição. Os TFs ativados (MYBs, HSFS, WRKYs, NAC e os dedos de zinco, por exemplo) causam em seguida a expressão diferencial de muitos transcritos entre genótipos tolerantes e sensíveis, tais como proteínas funcionais (HSPs, PRS e dehydrin, por exemplo), proteínas de modificações pós-transducionais (SKP2A, LAP-A1 e XERICOs, por exemplo) e genes associados com processos fisiológicos e metabólicos, como a fotossíntese (FNR, GOXs e RBCS3B, entre outras), a produção de ERO ou do sistema de detoxificação (PODs, AO, e GST, entre outras), a biossíntese e metabolismo de fitormônios (ABA3, LOXs, e UGTs), e outras enzimas metabólicas tais como PAL e BAM. Estas modificações específicas fizeram com que as linhas introgrididas se mostrassem mais tolerantes ao frio do que o genitor da espécie *S. lycopersicum*, ao reduzir o dano à membrana celular e fotoinibição, regular o metabolismo e manter o hormonal e a homeostase de ERO.

1.6 Análise de transcritomas

1.6.1 O processamento do RNA mensageiro

Inúmeros tipos de RNAs são produzidos nas células. Os diferentes RNAs recém-sintetizados no processo de transcrição denominam-se transcritos primários (pré-RNAs). A forma funcional de uma molécula de RNA é gerada após o seu processamento, em uma série de modificações pós-transcricionais (Figura 2). Estas alterações podem ser de deleção, adição ou modificação de nucleotídeos, ou de regiões maiores do transcrito primário. Em eucariotos, os RNA nucleares que dão origem aos mRNAs são bastante alterados, desde o início de sua transcrição até o momento de serem transportados ao citoplasma. As modificações são de vários tipos e incluem a adição de nucleotídeos nas extremidades 5' (CAP) e a poliadenilação da extremidade 3' e a retirada de fragmentos que interrompem as sequências funcionais do mRNA (excisão de íntrons ou *splicing*) (Zaha et al., 2012).

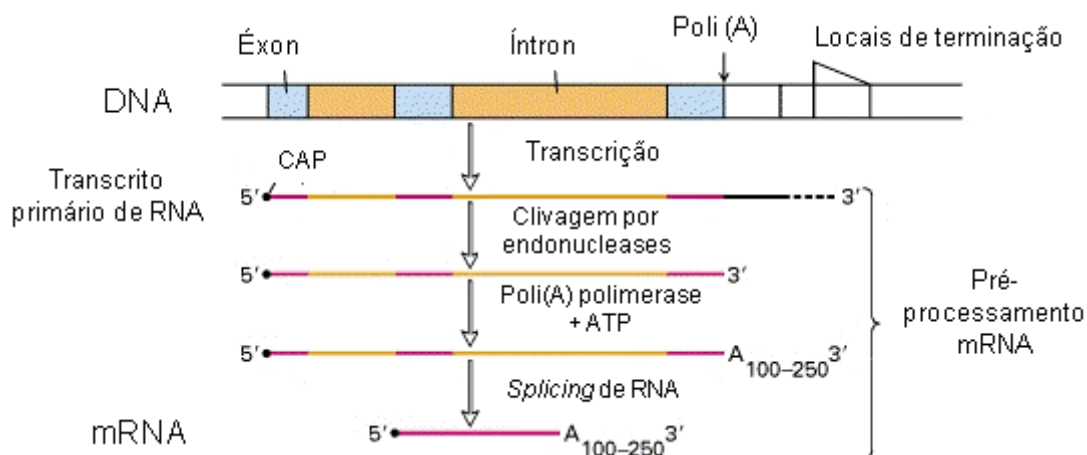


Figura 2. Visão geral de processamento do mRNA em eucariotos. Pouco tempo depois da RNA polimerase II iniciar a transcrição do primeiro nucleotídeo, do primeiro éxon de um gene, a extremidade 5' do RNA nascente recebe um CAP formado por com 7-metilguanilato. A transcrição pela RNA polimerase II termina em qualquer um dos vários locais de terminação a jusante do local da poli(A), que está localizado na extremidade 3' do éxon final. Após, a transcrição primária é clivada no local da poli(A), uma cadeia de resíduos de adenina (A) é adicionada. A cauda poli(A) contém resíduos ≈ 250 A em mamíferos, ≈ 150 em insetos, e ≈ 100 em leveduras. Para transcrições primárias curtas com alguns íntrons, a poliadenilação, clivagem e *splicing* geralmente segue-se a terminação, como mostrado. Para grandes genes com vários íntrons, os íntrons são frequentemente retirados do RNA nascente antes da transcrição do gene ser completada. O CAP na extremidade 5' é mantido em mRNAs maduros. Obtida de Lodish et al., (2000)

Na extremidade 5' de mRNAs nucleares ocorre a adição de uma 7-metilguanossina, denominada 5'-CAP, no primeiro nucleotídeo transcrito. Esta modificação é a primeira que ocorre nos pré-mRNAs, sendo necessária para a tradução da maioria dos mRNAs, pois é reconhecida pelos ribossomos no início da síntese de proteínas. Esta adição é essencial para a expressão eficiente de genes e a viabilidade celular e é, também, necessária para a maioria dos mRNAs celulares em todos os organismos eucariotos em que foi investigada. Em alguns sistemas experimentais também foi demonstrado que 5'-CAP atua na estabilização do mRNA contra o ataque de exonucleases, promove a transcrição, excisão de íntrons, poliadenilação e o transporte dos mRNA através da membrana nuclear (Cowling, 2010). O 5'-CAP ocorre em todos os mRNAs eucarióticos presentes no citoplasma, mas não ocorre nos mRNAs de mitocôndrias e de cloroplastos (Zaha et al., 2012).

A poliadenilação da extremidade 3' de mRNAs de eucariotos consiste da adição, nesta extremidade, de uma sequência de aproximadamente 200 resíduos de adenina que é denominada de cauda poli(A). Esta cauda poli(A) não está codificada no DNA e não existe nos rRNAs e nos tRNAs. Ela é adicionada aos pré-RNAs pela ação da enzima poli(A)-polimerase. A clivagem pós-transcricional do precursor do

mRNA é um passo essencial na maturação do mRNA. Após a clivagem, a maioria dos mRNAs eucarióticos, com a exceção de transcritos de replicação histona-dependentes em alguns organismos, adquirem uma cauda poli(A) nas suas extremidades 3'OH. O processo de formação da extremidade 3' promove a terminação da transcrição e o transporte do mRNA a partir do núcleo. A cauda poli(A) também melhora a estabilidade e a tradução de mRNA, mais provavelmente por providenciar um sítio de ligação para proteínas de ligação poli(A) (Zhao et al., 1999). Quase todos os sinais de poliadenilação em eucariotos contêm um elemento de núcleo a montante e uma sequência consenso AAUAAA (ou uma variante) localizada entre 10 e 35 nucleotídeos a montante do local real da adição da cauda poli(A). Além disso, as sequências entre 14-70 nucleotídeos a jusante (elementos a jusante do núcleo) são conhecidas por estarem envolvidas em dirigir a poliadenilação. Elementos auxiliares a montante e a jusante da sequência AAUAAA também têm sido caracterizados. Estes elementos auxiliares podem melhorar a eficiência de poliadenilação, muitas vezes mediada pela ligação específica de fatores de trans-regulação adicionais (Lutz e Moreira, 2011).

A cauda poli(A), presente na grande maioria dos mRNAs, é uma característica de grande importância para a realização de análises de RNA-seq. Atualmente, a técnica mais utilizada é realizada através da seleção da cauda poli(A), que usa a afinidade desta com o oligo-dT para selecionar transcritos que apresentem esta cauda. Na maioria das células, entre 60-90% das espécies de RNA consistem em RNAs estruturais, tais como rRNA e tRNA. Deste modo, quando este tipo de RNAs não constitui o foco do estudo, necessita-se de uma estratégia para evitar que estes RNAs dominem os dados do sequenciamento. Duas abordagens são utilizadas para o enriquecimento relativo de mRNA presente nas amostras. A primeira abordagem começa com o RNA total, de onde o rRNA é deplecionado usando-se um conjunto de oligonucleotídeos que se ligam ao rRNA. O segundo método seleciona para transcritos por meio do isolamento de RNA que apresente uma cauda poli-A na sua extremidade 3'. Estes constituirão o material de partida para a construção de bibliotecas de transcrito (Tariq et al., 2011). No entanto, este protocolo requer alta qualidade (tecido fresco ou rapidamente congelado) e uma quantidade suficiente de RNA a partir de uma amostra a ser sequenciada (Sun et al., 2013).

Durante o passo final na formação de um mRNA maduro e funcional, os íntrons são removidos e os éxons são unidos em conjunto. O processo de excisão

de íntrons é denominado *splicing*. Os íntrons spliceossômicos representam os íntrons presentes na grande maioria dos mRNAs, cujo *splicing* é realizado com o auxílio de ribonucleoproteínas que formam um complexo chamado spliceossomo. Após a excisão dos íntrons, as porções restantes denominadas éxons, além das sequências das regiões não traduzidas (UTRs) localizadas nas regiões 5' e 3' (UTRs 5' e 3') permanecem no RNA maduro, porém não são traduzidas (Zaha et al., 2012). Para as unidades de transcrição curtas, o *splicing* de RNA geralmente segue a clivagem e poliadenilação da extremidade 3' do transcrito primário. Mas para as longas unidades de transcrição contendo vários éxons, o *splicing* de éxons no RNA recém produzido geralmente começa antes que a transcrição do gene seja completada.

A localização de junções éxon-íntron (ou seja, os locais de *splicing*), em um pré-mRNA pode ser determinada pela comparação da sequência do DNA genômico com a do cDNA preparado a partir da sua sequência correspondente. As sequências que estão presentes no DNA genômico, mas ausente do cDNA representam íntrons e indicam as posições dos locais de *splicing*. A análise de um grande número de diferentes mRNAs revelaram curtas sequências consenso moderadamente conservadas nas fronteiras éxon-íntron em pré-mRNA eucarióticos; em organismos superiores, uma região rica em pirimidina, imediatamente a montante do local de *splicing* na região 3' é também comum. Os nucleotídeos mais conservados são os (5') GU e (3') AG encontrados nas extremidades da maioria dos íntrons. Em análises de supressão da porção central de íntrons em vários pré-mRNA foi demonstrado que, geralmente, apenas 30-40 nucleotídeos em cada extremidade de um íntron são necessários para que o *splicing* ocorra em taxas normais (Lodish et al., 2000).

1.6.2 O transcrito em eucariotos

O transcrito pode ser definido como a completa gama de moléculas de RNA mensageiro (mRNA) expressas por um organismo. O termo transcrito também pode ser usado para descrever o conjunto de transcritos de mRNA produzidos por uma célula ou por um tipo de tecido em particular. Em contraste com o genoma, o qual é caracterizado pela sua estabilidade, o transcrito muda ativamente. Na verdade, o transcrito de um organismo varia dependendo de

muitos fatores, incluindo a fase de desenvolvimento e as condições ambientais (Brown, 2002). Embora o mRNA caracterize o transcrito, ele corresponde a menos de 4% do total de RNA das células. Mesmo assim, constitui o componente mais significativo, pois contém os RNAs de codificação que especificam a composição do proteoma e, portanto, determinam a capacidade bioquímica da célula. Um ponto importante a se notar é que o transcrito nunca é sintetizado de novo. Cada célula recebe parte do transcrito do seu genitor quando ele é trazido primeiramente à existência por divisão celular, e mantém um transcrito durante toda a sua vida. Mesmo células quiescentes em esporos bacterianos ou nas sementes de plantas têm o transcrito, embora a tradução do transcrito em proteína possa ser completamente desligado. Transcrição não resulta na síntese do transcrito mas o mantém, substituindo mRNAs que foram degradados, e traz alterações na composição do transcrito através do processo de ligar e desligar diferentes conjuntos de genes (Brown, 2002).

A análise de expressão gênica é amplamente utilizada para desvendar mecanismos reguladores que controlam os processos celulares em plantas, animais e micróbios. As análises de micro-arranjo foram muito utilizadas na análise do perfil de expressão global de genes, embora esta tecnologia à base de hibridização tenha sua utilização grandemente restringida aos genes conhecidos e por apresentar uma gama limitada de quantificação. A tecnologia de RNA-Seq estende as possibilidades de estudos do transcrito para a análise de genes previamente não identificadas e de *splicing* alternativo (Wang et al., 2009). Além disso, o RNA-Seq dispõe de uma gama dinâmica de quantificação e reduzida variabilidade técnica. Essas vantagens, juntamente com a diminuição do custo do sequenciamento, fazem do RNA-Seq um método cada vez mais usado para estudos de expressão de todo o genoma em muitos sistemas biológicos, incluindo espécies com genomas não sequenciados (Van Verk et al., 2013).

1.6.3 Sequenciadores de nova geração (NGS)

Com o surgimento dos sequenciadores de nova geração (NGS) a capacidade de sequenciamento aumentou drasticamente nos últimos anos. Diversas

plataformas de sequenciamento surgiram, entre as quais podemos citar SOLiD e Ion Torrent PGM (Life Sciences), Genome Analyzer, HiSeq 2000, MiSeq (Illumina), GS FLX Titanium e GS Junior (Roche), entre outras. O sequenciamento de RNA (RNA-Seq) por meio de NGS vem se popularizando rapidamente e faz com que outras tecnologias usadas na análise de transcritomas sejam deixadas de lado, tanto pelo aumento da confiabilidade dos resultados produzidos pelos NGS bem como pela redução de custo no sequenciamento. O sequenciador HiSeq 2000 (Illumina) adota a tecnologia de sequenciamento por síntese (SBS). A tecnologia SBS permite sequenciamento paralelo em massa de milhões de fragmentos utilizando um método baseado em um terminador reversível que detecta bases únicas quando elas são incorporados nas fitas de DNA em crescimento. Um terminador fluorescente marcado é visualizado quando cada dNTP é adicionada e em seguida clivada para permitir a incorporação da base seguinte. Uma vez que todos os quatro dNTPs ligados a terminadores reversíveis estão presentes durante cada ciclo de sequenciamento, a concorrência natural minimiza a incorporação de vieses relacionados à adição das bases. A identificação é feita diretamente a partir de medições de intensidade de sinal durante cada ciclo, o que reduz as taxas de erro em comparação com outras tecnologias. O resultado final é obtido com uma acurácia superior a 98% (Liu et al., 2012) no sequenciamento base-por-base, que elimina erros específicos de contexto da sequência permitindo uma identificação robusta em todo o genoma, incluindo regiões de sequências repetitivas e dentro de homopolímeros.

O fluxo de trabalho do sequenciamento Illumina é baseado em três passos: as bibliotecas são preparadas a partir de praticamente qualquer amostra de ácido nucleico amplificado para produzir aglomerados clonais, e são sequenciados utilizando síntese paralela massiva. A preparação das bibliotecas pode ser realizada por meio de *kits* de preparação de amostras TruSeq (Illumina) ou pelo uso de *kits* de preparação de amostras de DNA compatíveis com sequenciadores Illumina. A geração dos *clusters* ocorre no sistema automatizado de geração de *clusters* cBot (Illumina). cBot amplifica isotermicamente os fragmentos de cDNA que foram capturados por oligonucleotídeos complementares ao adaptador de ligação covalente à superfície das células de fluxo Illumina. As células de fluxo facilitam o acesso do DNA ligado às enzimas, garantindo alta estabilidade do molde ligado à superfície e baixa ligação de nucleotídeos fluorescentes rotulados não-específicos.

Fragmentos de DNA anexados são estendidos e amplificados por ponte para criar centenas de milhões de agrupamentos, cada um dos quais contém aproximadamente 1.000 cópias idênticas de uma única molécula do molde. Por último, os reagentes de sequenciamento pré-configurados são depositados na bandeja de reagentes do instrumento antes do início da corrida.

Antes de sequenciamento, as bibliotecas são juntadas em cadeias simples com a ajuda de enzimas de linearização (Mardis, 2008) e, em seguida, quatro tipos de nucleotídeos (ddATP, ddGTP, ddCTP, ddTTP), que contêm um corante fluorescente clivável diferente cada um, e um grupo de bloqueio removível pode complementar o modelo em uma base a cada vez, e o sinal pode ser captado por um dispositivo de carga acoplada (charge-coupled device – CCD). Com a multiplexação incorporada em primers P5/P7 e adaptadores, o sequenciador pode lidar com milhares de amostras simultaneamente. HiSeq 2000 necessita de um software de controle HiSeq (HCS) para o controle do programa, um software analisador em tempo real (RTA) para fazer a identificação no instrumento, e o software CASAVA para a realização de análises secundárias (Liu et al., 2012).

Os softwares sistema de controle HiSeq (HCS) e analisador em tempo real (RTA) são adotados por HiSeq 2000. Estes dois softwares podem calcular o número e a posição de agrupamentos com base em suas 20 primeiras bases, de modo que as primeiras 20 bases de cada sequenciamento decidiriam a produção e qualidade de cada sequência. HiSeq 2000 utiliza dois *lasers* e quatro filtros para detectar quatro tipos de nucleotídeos (A, T, G e C). Os espectros de emissão destes quatro tipos de nucleotídeos apresentam *cross-talk*, pois as imagens dos quatro nucleotídeos não são independentes e a distribuição de bases pode afetar a qualidade do sequenciamento. Os arquivos de saída do sequenciamento padrão do HiSeq 2000 consistem em arquivos bcl*, que contêm as chamadas de base e índices de qualidade em cada ciclo. Por meio do programa CASSAVA os arquivos no formato bcl* são convertidos para o formato fastQ. Neste formato as bases e os valores de qualidade são representados em um mesmo arquivo. Além disso, as qualidades são codificadas por meio de caracteres ASCII ao invés de uma sequência de números. Isto permite que a codificação seja mais eficiente e ocupe um menor espaço nos arquivos criados.

1.7 Análise dos dados obtidos por meio de RNA-Seq

A tecnologia de RNA-Seq está se consolidando como o método padrão para medir os níveis de expressão do RNA. Ela substituiu com inúmeras vantagens a abordagem de microarranjos que constituiu, por aproximadamente uma década, o principal método para verificação da expressão diferencial de genes. Entre as vantagens que RNA-seq oferece sobre os microarranjos está o fato de que a tecnologia de RNA-Seq não se limita a detecção de transcritos que correspondam a sequências genômicas preexistentes. Isso faz com que o RNA-Seq se torne particularmente útil para organismos que ainda não tenham um genoma e um transcriptoma de referência. RNA-Seq pode revelar a localização exata de limites de transcrição, numa resolução de uma única base. *Reads* de 30 pb provenientes de RNA-Seq fornecem informações sobre como dois éxons estão conectados, enquanto que *reads* maiores ou *reads* curtas *paired-end* revelam a conectividade entre vários éxons (Wang et al., 2009). Esses fatores tornam o RNA-Seq muito útil no estudo de transcriptomas complexos e na detecção de SNPs nas regiões transcritas (Quinn et al., 2013). Comparando com a técnica de microarranjo, o RNA-Seq possui um baixo nível de ruído, pois as sequências de DNA podem ser inequivocadamente mapeadas em regiões únicas do genoma. A tecnologia RNA-Seq também demonstrou ser altamente precisa para quantificar os níveis de expressão, tal como determinado utilizando PCR quantitativa (qPCR) (Nagalakshmi et al., 2008) e quando adicionados controles com concentrações conhecidas de RNA (Mortazavi et al., 2008). Além disso, os resultados do RNA-Seq também apresentam níveis mais elevados de reprodutibilidade, tanto para replicatas técnicas quanto para replicatas biológicas, quando comparado com outras técnicas de análise de expressão diferencial (Nookaew et al., 2012).

Atualmente, ainda não há um consenso sobre o melhor método de análise dos dados obtidos pela técnica de RNA-Seq. Algoritmos estão sendo desenvolvidos para as diversas etapas envolvidas na análise dos dados. Artigos comparando os estes algoritmos utilizados em cada etapa mostram que um consenso ainda está longe de ser alcançado. Para a etapa de alinhamento das *reads*, Engström et al. (2013) compararam a performance de 26 protocolos baseados em 11 programas de alinhamento e concluem que cada um deles possui pontos fortes e fracos, e não

encontraram nenhum algoritmo que apresentasse uma boa performance em todos os parâmetros analisados. A comparação de métodos para análise da expressão diferencial dos dados obtidos por meio de RNA-Seq não é conclusiva sobre o melhor algoritmo e mostram que não há, atualmente, um algoritmo que se saia melhor em todos os parâmetros importantes na determinação dos genes diferencialmente expressos (Soneson e Delorenzi, 2013; Rapaport et al., 2013; Wesolowski et al., 2013).

2. Objetivos

O objetivo geral deste trabalho consiste em analisar a expressão de genes em arroz sob estresse por frio. Os objetivos específicos são: Identificar os genes diferencialmente expressos entre os genitores Nipponbare, BRS Atalanta e os *bulks* de RILs formados a partir de genotipagem frente ao estresse por frio na fase de plântula; e identificar fatores de transcrição diferencialmente expressos em cada uma das combinações analisadas.

3. Material e Métodos

3.1. Material genético

Para a execução deste trabalho foram utilizados cruzamentos artificiais entre genótipos contrastantes para o caráter tolerância a frio em arroz. Na literatura, a subespécie *japonica* é descrita como tolerante ao frio na germinação, na fase vegetativa e na fase reprodutiva. Este fato decorre da região onde esta subespécie foi domesticada, que corresponde às regiões temperadas da Ásia e com maiores altitudes. Já a subespécie *indica* mostra-se sensível ao frio na germinação, na fase vegetativa e na fase reprodutiva devido a sua domesticação ter ocorrido em regiões com maiores temperaturas e menores altitudes. Para a subespécie *japonica* foi utilizada o genótipo Nipponbare, pois dessa cultivar se dispõe de grande quantidade

de informações em âmbito genético-molecular, por ter servido como modelo para o sequenciamento da espécie *Oryza sativa* L. (GOFF et al., 2002; IRGSP, 2005; RAP-DB, 2013). Como representante da subespécie *indica* foi utilizada a cultivar BRS Atalanta. Esta cultivar foi utilizada nos cruzamentos artificiais devido a sua sensibilidade ao frio. Além disso, esta cultivar apresenta características agronômicas de interesse, tais como: alta produtividade, elevada qualidade de grãos, elevado teor de amilose e ciclo super-precoce.

As RILs foram desenvolvidas a partir do cruzamento entre o genótipo Nipponbare e a cultivar BRS Atalanta. Os cruzamentos artificiais foram realizados no ano de 2008. No ano de 2009 as sementes F1 obtidas foram semeadas, dando origem às sementes F2. Em agosto de 2010, 450 sementes F2 foram semeadas na área experimental pertencente a Embrapa Arroz e Feijão, no município de Brazabrantes, estado de Goiás/Brasil. A colheita das sementes F3 foi realizada em novembro de 2010. Em dezembro do mesmo ano, 250 famílias F3 foram semeadas na área experimental da Estação Experimental Terras Baixas, pertencente a Embrapa Clima Temperado, no município de Capão do Leão, Rio Grande do Sul/BR. Em abril de 2011 as sementes F4 foram colhidas. Em junho de 2011, 237 famílias F4 foram semeadas em área experimental pertencente à empresa RiceTec, no município de Normandia, estado de Roraima/BR. Em setembro de 2011 foram colhidas as sementes F5. Em outubro de 2011, sementes de 237 F5 foram semeadas na área experimental da Estação Experimental Terras Baixas, pertencente a Embrapa Clima Temperado, no município de Capão do Leão, Rio Grande do Sul/BR. Em março de 2012 foram colhidas as sementes F6 de 210 famílias.

Uma vez obtidas as sementes na geração F6, ou seja, com aproximadamente 97% de homozigose, considerou-se esta taxa restante de heterozigose como desprezível. Assim, procedeu-se a seleção das 15 famílias que se mostraram mais próximas geneticamente aos pais utilizados nos cruzamentos. Esta seleção foi realizada tendo-se em mãos os resultados da genotipagem por sequenciamento (GBS) das 210 famílias que constituem a geração F6 provenientes do cruzamento inicial (Ahlert et al., 2014 (dados não publicados)). A genotipagem dos pais e das famílias F6 foi realizado por meio de um método de alto rendimento utilizando os dados gerados a partir do resequenciamento de todo o genoma dos pais e das 210 famílias. A estratégia utilizada para a realização da genotipagem por

sequenciamento (GBS) segue o protocolo proposto por Elshire et al. (2011), onde enzimas de restrição sensíveis à metilação são utilizadas para reduzir a complexidade do genoma, uma vez que evitam o acesso a regiões repetitivas. A filtragem e seleção de SNPs foi realizada por meio da ferramenta *GBS analysis*, uma extensão do programa TASSEL (Bradbury et al., 2007). Com base nos resultados obtidos foram selecionadas as 15 RILs que se mostraram mais próximas ao genitor Nipponbare para compor o *bulk* de RILs tolerantes ao frio. As 15 RILs mais próximas ao genitor BRS Atalanta formaram o *bulk* de RILs sensíveis ao frio.

3.2 Tratamento por frio

As sementes dos genótipos Nipponbare e BRS Atalanta e das RILs selecionadas para compor cada *bulk* foram tratadas com hipoclorito de sódio (NaClO), numa concentração de 1%, durante 3 minutos e em seguida lavadas por três vezes com água destilada, para evitar o desenvolvimento de patógenos durante a condução do experimento. De cada genitor foram colocadas para germinar três repetições com 50 sementes cada. De cada RIL foram utilizadas cinco sementes por repetição, totalizando 75 sementes por repetição. Estas sementes foram colocadas para germinar em placas de Petri contendo duas folhas de papel germinador. Foram adicionados 15 mL de água ultrapura para embebição durante as primeiras 24h de condução do experimento. Após este período, o excesso de água foi retirado. A manutenção dos níveis de umidade foi feita a cada dois dias. As sementes foram postas para germinar em câmaras de germinação do tipo BOD (*Biochemical Oxygen Demand*), a 28°C. Todas as etapas do experimento foram conduzidas sem a presença de luz.

3.2.1 Condição controle

A condição controle dos genótipos (Nipponbare e BRS Atalanta) e dos *bulks* foi composta pela manutenção das sementes por sete dias a 28°C (da Cruz e Milach, 2004; Bosetti, et al., 2012), que é considerada a temperatura ótima de germinação, em câmara de germinação. Posteriormente, as sementes da condição controle permaneceram por mais 24h na condição de 28°C, sem a presença de luz.

3.2.2 Condição tratamento

Na condição tratamento, as sementes permaneceram durante sete dias a 28°C em câmara de germinação, sem a presença de luz. Após este período foi aplicado o tratamento por frio. Este tratamento consistiu na exposição da condição tratamento a temperatura de 13°C pelo período de 24h.

3.3 Delineamento Experimental

O delineamento experimental utilizado na condução do experimento foi o de blocos casualizados (DBC), com três repetições.

3.4 Extração de RNA total

Todos os utensílios utilizados na coleta do material biológico foram limpos com o reagente RNase AWAY (Life Technologies, Carlsbad, CA, Estados Unidos) para a eliminação de RNases e de DNA que possam constituir contaminantes às amostras coletadas. As plântulas foram coletadas removendo-se a casca das sementes e o endosperma das mesmas. Logo após estas plântulas foram lavadas

com água ultrapura, secas levemente com *Kimwipes*[™] (Kimberly-Clark, Irving, TX, Estados Unidos) e colocadas em tubos para centrifugação do tipo Falcon, em nitrogênio líquido. O material foi armazenado em ultrafreezer, a -80°C, até posterior uso.

O RNA total foi extraído das plântulas usando-se o reagente TRI Reagent® (Sigma-Aldrich, St. Louis, MO, Estados Unidos) de acordo com instruções do fabricante. A purificação do RNA total foi realizada usando-se o *kit RNeasy Plant Mini* (Qiagen, Valencia, CA, Estados Unidos) seguindo as instruções do fabricante. A quantificação do RNA total foi realizado com o espectrofotômetro NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). O preparo das bibliotecas começou com o uso de 4µg de RNA total. A qualidade do RNA total foi verificada por eletroforese em gel de agarose a 1% (2g de agarose + 200 mL de tampão TBE 0,5X + 5µL de brometo de Etídio (EtBr)).

Posteriormente, as amostras foram diluídas em água Nuclease-Free (Life Technologies, Carlsbad, CA, Estados Unidos) até completar o volume de 10µL. A concentração final das amostras foi ajustada para 400ng por amostra (400ng de RNA por amostra + 2µL de 6X Loading buffer + água para completar 10µL).

3.5 Preparação das bibliotecas de cDNA

Para a confecção das bibliotecas de cDNA foi utilizado somente o mRNA, sendo reduzida a presença dos demais tipos de RNAs (Figura 3). Realizou-se este preparo por meio do uso do *kit TruSeq RNA sample preparation v2* (Illumina, San Diego, CA, Estados Unidos) de acordo com as instruções do fabricante. Começando com o RNA total, o mRNA foi primeiramente purificado usando a seleção para a cauda poli(A), que é uma característica exclusiva do mRNA e assim somente RNAs que contém esta cauda são selecionados. O primeiro passo do fluxo de trabalho envolve a purificação moléculas de mRNA que apresentam a cauda poli(A) por meio de BeadChips magnéticas que fazem a junção do oligonucleotídeo poli-T com as bases complementares presentes na cauda poli(A).

Após a purificação, o mRNA é fragmentado em pequenos pedaços, por meio de cátions divalentes a temperatura elevada. Fragmentos de RNA segmentados são

copiados para a primeira cadeia de cDNA utilizando transcriptase reversa e primers aleatórios. Isto é seguido por síntese da segunda cadeia de cDNA, utilizando DNA-polimerase I e RNase H. Estes fragmentos de cDNA foram então submetidas a um processo de reparação das extremidades 5' e 3' dos fragmentos preparando-as para uma eficiente ligação dos adaptadores “Y” contendo um código de barras único (*barcode*) e locais de junção dos *primers*. Em seguida, os produtos são purificados e enriquecidos por meio da reação em cadeia da polimerase (PCR) para criar a biblioteca de cDNA definitiva. Foram realizadas duas seleções por meio do uso de oligo(dT) para a cauda poli(A) do mRNA visando reduzir os níveis de rRNA a valores aceitáveis em experimentos de RNA-Seq (Nagalakshmi et al., 2010). O *kit* de preparo gera fragmentos de aproximadamente 350 pares de bases (bp) e por meio dos beads foram removidos aquelas sequências que apresentaram um comprimento menor que 100 bp.

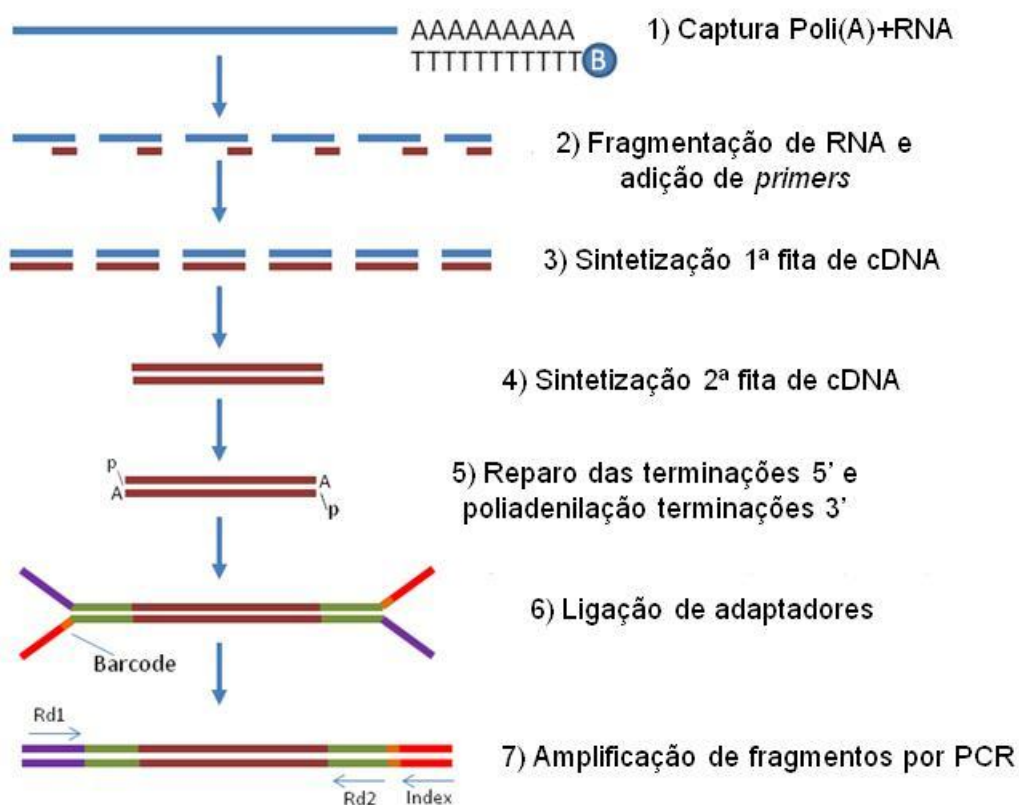


Figura 3. Preparação de biblioteca de RNA para plataforma Illumina. RNA com cauda poli(A) é enriquecida utilizando beads oligo(dT) seguido por fragmentação e transcrição reversa. As regiões 5' e 3' dos fragmentos de cDNA são em seguida preparados para permitir a ligação eficiente dos adaptadores "Y", que contêm um código de barras único e sítios de ligação de primers. Finalmente, os cDNAs ligados são amplificadas por PCR, e estão prontos para o sequenciamento. Fonte: <http://www.labome.com/method/RNA-seq-Using-Next-Generation-sequencing.html>

Para purificação da PCR foram utilizados *BeadsChips* magnéticos Agencourt AMPure XP (Beckman Coulter, Pasadena, CA, Estados Unidos). Este reagente tem como função a recuperação dos amplicons, remoção de dNTPs não incorporados, *primers*, dímeros de *primers*, sais e outros contaminantes. Atua, também, na estabilização dos produtos da PCR após sua limpeza e possibilita a recuperação de DNAs de fita dupla e também de fita simples.

Neste trabalho foi utilizado o conjunto “A” de adaptadores disponibilizados pela Illumina. Após a finalização da confecção das bibliotecas de cDNA, foi verificada a qualidade e a quantidade de cDNA obtido por meio do aparelho Bioanalyzer 2100, modelo Agilent DNA 7500 (Agilent, Santa Clara, CA). Além disso, a qualidade e quantidade de DNA obtido foi mensurado pelo aparelho Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, Estados Unidos).

Após a validação das bibliotecas, fez-se a normalização das mesmas para uma concentração de 10nM usando-se o tampão EB (10 mM Tris·Cl, pH 8.5)(Qiagen, Venlo, Holanda), com o uso de tensoativo hidrofílico (Tween 20) a 0,1%. Para a transformação dos valores das bibliotecas para uma concentração de 10nM fez-se a conversão de micrograma (μ) para picomol (pmol) de regiões terminais de DNA linear por meio da fórmula:

$$pmol\ de\ DNA = \mu g\ DNA \times \frac{pmol}{660\ pmol} \times \frac{10^6 pg}{1\ \mu g} \times \frac{1}{N} \times 2 \times \frac{kb}{1000\ pb}$$

Onde N representa o número de nucleotídeos (em kb), 660pg/pmol é o peso molecular médio de um único par de nucleotídeos, 2 é o número de terminais de uma molécula linear de DNA, e kb/1000bp é o fator de conversão de quilo bases para par de bases.

3.6 Sequenciamento das bibliotecas de cDNA

O sequenciamento das bibliotecas foi realizado em sequenciador de nova geração (NGS) HiSeq2000 (Illumina, San Diego, CA, Estados Unidos), sendo as bibliotecas submetidas a 100 ciclos de sequenciamento *single-end* (1x100bp). Foram utilizadas quatro canaletas de sequenciamento. Em cada uma delas foram colocadas seis bibliotecas, totalizando 24 bibliotecas sequenciadas. A qualidade da

read de cada base foi representada usando-se caracteres ASCII para Phred+33. O índice de qualidade $Q_{\text{PHRED}} = -10 \times \log_{10}(\text{Pe})$ foi utilizado primeiramente nas *reads* obtidas pelo método de Sanger e a partir da versão 1.8 do software CASSAVA (Illumina) passou a ser o padrão para os arquivos .fastq gerados a partir de sequenciamentos realizados com a plataforma Illumina. Os índices de qualidade Phred são logaritmicamente ligados à probabilidade de erro. Como exemplo temos que um $Q_{\text{PHRED}} = 10$ indica uma probabilidade de identificação incorreta de uma base na ordem de 1 em 10 proporcionando uma acurácia na identificação na ordem de 90%. Já um $Q_{\text{PHRED}} = 30$ indica uma probabilidade de chamada incorreta de uma base na ordem de 1 em 1000 bases, ou seja, a acurácia na identificaçõesobe para 99,9% (Ewing e Green, 1998). Os arquivos obtidos e que foram usados nas análises subsequentes apresentam o formato .fastq (Figura 5).

Neste arquivo estão presentes informações dispostas em quatro linhas. A primeira linha representa uma linha de título iniciada pelo sinal gráfico '@'. Este é um campo de formato livre, sem limite de comprimento permitindo anotação arbitrária ou comentários a serem incluídos. A segunda linha representa a linha da sequência, onde é disposta cada base lida. Assim como no formato FASTA, não há nenhuma limitação explícita sobre os caracteres esperados. Espera-se que estejam presentes apenas os códigos de uma letra da IUPAC para DNA ou RNA, em maiúscula. Na terceira linha apresenta-se o sinal '+' para sinalizar o fim da linha de sequência e do início da sequência de qualidade. A quarta linha refere-se a qualidade de cada base lida que é representada por um único caractere pertencente ao subconjunto de caracteres ASCII, normalmente variando entre os caracteres ASCII 33 a ASCII 126, com um mapeamento de deslocamento simples. Fundamentalmente, depois de concatenação (remoção de quebras de linha), a sequência de qualidade deve possuir um número de caracteres igual ao da segunda linha (Cock et al., 2010).

```
@SRR014849.1 EIXKN4201CFU84 length=93
GGGGGGGGGGGGGGGGGCTTTTTTTGTTTGAACCGAAAGG
GTTTTGAATTTCAAACCTTTTCGGTTTCCAACCTTCCAA
AGCAATGCCAATA
+SRR014849.1 EIXKN4201CFU84 length=93
3+&$#"7F@71,'";C?,B;?6B;:EA1EA
1EA5'9B:?:#9EA0D@2EA5':>5?:%A;A8A;?9B;D@
/=<?7=9<2A8==
```

Figura 4. Padrão de informações presentes nos arquivos com formato .fastq. Figura obtida de Cock et al., 2010.

3.7 Análise dos dados obtidos pelo sequenciamento das bibliotecas

A qualidade dos dados recebidos do sequenciamento foram verificados pelo uso da ferramenta FastQC (Babraham Bioinformatics) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (Figura 6). Dele foram obtidas estatísticas básicas sobre a biblioteca tais como: nome do arquivo, tipo de arquivo, formato, total de sequências, sequências filtradas, comprimento das sequências e porcentagem de conteúdo de guanina + citosina (%GC) de cada base em cada sequência lida. Além disso, foram verificados parâmetros como o score de qualidade por sequência, qualidade da sequência por base, conteúdo da sequência por base, conteúdo GC por base, conteúdo GC por sequência, conteúdo de N por base, distribuição do comprimento das sequências, sequências duplicadas, sequências superexpressas e Kmers super-representados.

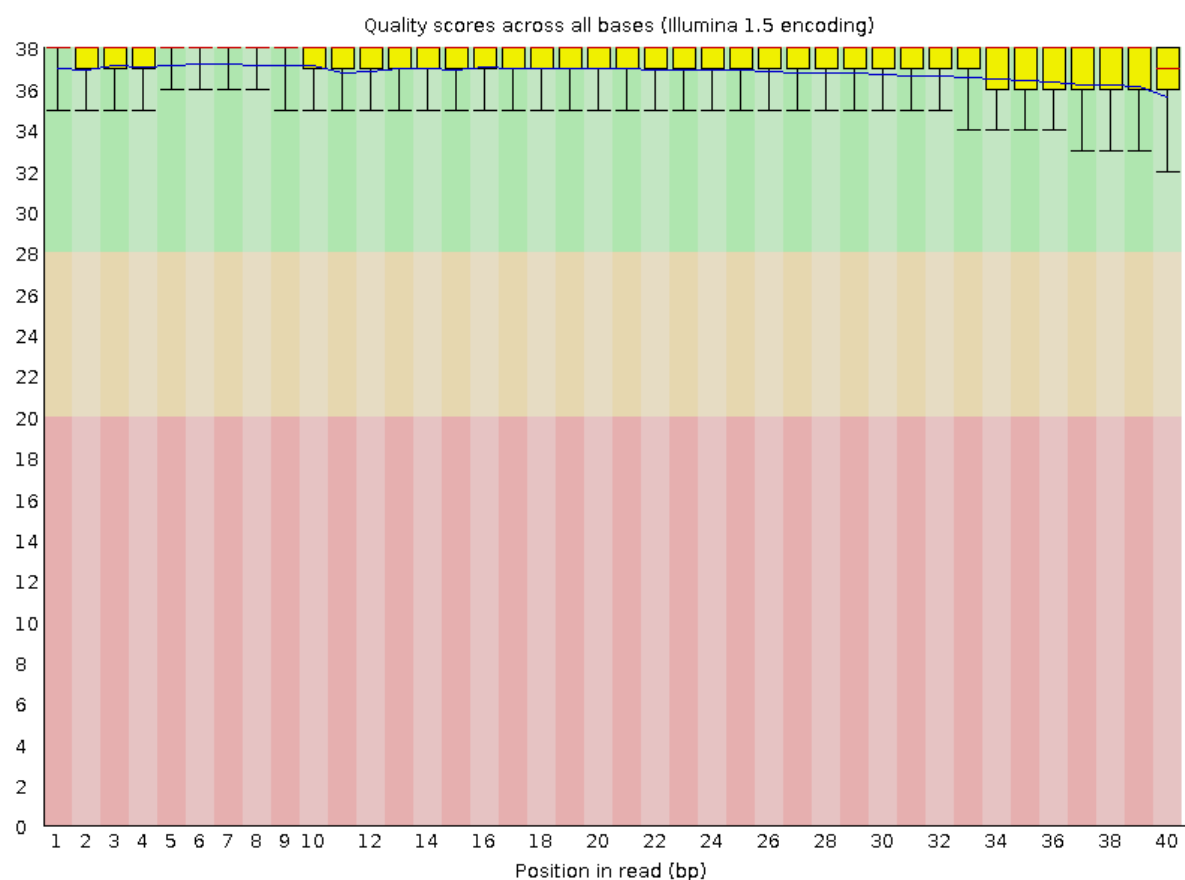


Figura 5. Representação da qualidade de *reads* obtida pelo uso da ferramenta fastQC. Figura obtida de: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Para a limpeza das sequências de cada biblioteca foi utilizada a ferramenta Trimmomatic (Lohse et al., 2012). Foram utilizados os seguintes parâmetros: ILLUMINACLIP, para o corte de adaptadores e outras sequências específicas de *reads* realizadas em plataforma Illumina; SLIDINGWINDOW, que realiza cortes sempre que a qualidade média dentro da janela cair abaixo de um limiar estabelecido; LEADING e TRAILING, para cortar bases a partir do começo ou do final de uma *read* caso a qualidade esteja abaixo de um determinado limiar, respectivamente; e MINLEN, para eliminar *reads* que após os demais procedimentos tenham ficado com um comprimento menor que o especificado. Neste experimento foram utilizados os seguintes valores: ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 e MINLEN:80.

Após a confirmação da adequação das *reads* obtidas pela metodologia RNA-Seq fez-se uso do protocolo Tuxedo (Trapnell et al., 2012). Este protocolo fornece soluções para as três etapas essenciais na análise de RNA-Seq: alinhamento de sequências, quantificação e análise de expressão diferencial de genes. Na figura 7 está representado um fluxograma dos procedimentos necessários até a obtenção dos genes diferencialmente expressos. O primeiro passo consiste no uso do programa TopHat (Trapnell et al., 2009) onde as *reads* são mapeadas no genoma de referência. Como genoma de referência, tanto para os genótipos Nipponbare e BRS Atalanta, bem como para os *bulks* das RILs que se mostraram mais próximas aos respectivos pais, foi utilizado o genoma da genótipo Nipponbare (Os-Nipponbare-Reference-IRGSP-1.0) (disponível no endereço <http://rapdb.dna.affrc.go.jp/download/irgsp1.html>).

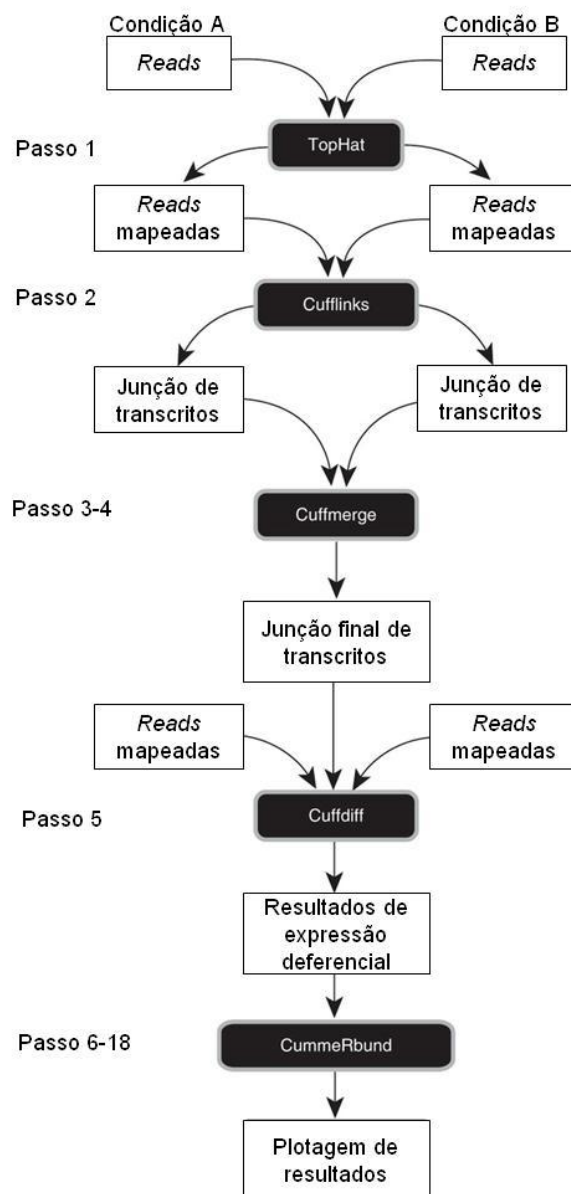


Figura 6. Em um experimento envolvendo duas condições, as reads são mapeadas contra o genoma com TopHat. As reads para cada replicata biológica são mapeados de forma independente. Estas reads mapeadas são fornecidas como entrada para o programa Cufflinks, que produz um arquivo de transfrags montadas para cada repetição. Os arquivos de montagem são mesclados, por Cuffmerge, com a anotação do transcrito de referência, em uma anotação unificada para análise posterior. Esta anotação mesclada é quantificada para cada condição por Cuffdiff, que produz dados de expressão em um conjunto de arquivos tabulares. Esses arquivos são indexados e visualizado com CummeRbund. Para facilitar a exploração de genes identificados por Cuffdiff como genes diferencialmente expressos, apresentando splincing alternativo ou sendo transcricionalmente regulados. Fonte: Trapnell et al., 2012.

O programa TopHat tem embarcado as soluções de alinhamento apresentadas pelos programas Bowtie (Langmead et al., 2009) e Bowtie2 (Langmead et al., 2012). Bowtie é um programa de alinhamento ultra-rápido e que usa eficientemente a memória da máquina para alinhar pequenas sequências de DNA a grandes genomas. Este programa faz uso das técnicas de Burrows-Wheeler,

baseadas no índice FM (*full-text minute-space index*) (Ferragina e Manzini, 2001) para armazenar a sequência do genoma de referência e permite que ele seja pesquisado rapidamente. A utilização de um algoritmo denominado algoritmo de recuo com reconhecimento de qualidade (*quality-aware backtracking algorithm*) permite a presença de *mismatches* entre o genoma de referência e as *reads* a serem alinhadas. Além disso esse algoritmo favorece alinhamentos de alta qualidade e apresenta dupla indexação, estratégia que evita o recuo excessivo (Langmead et al., 2009). No entanto, Bowtie não é apropriado para todas as tarefas de alinhamento de sequências. Ele não permite alinhamentos entre uma *read* e o genoma de referência quando ocorrerem grandes lacunas (*gaps*) e portanto, não pode alinhar *reads* que foram originadas por uma junção gerada pela exclusão de íntrons. As principais diferenças do programa Bowtie2 para o primeiro são: suporte a presença de *gaps*, onde o número e o comprimento dos *gaps* não é restrito; maior velocidade e maior sensibilidade para *reads* maiores que 50 bp; eliminação do limite superior para o comprimento das *reads* (que era fixado em 1000 bp no Bowtie); o modo de alinhamento *paired-end* é mais flexível que na versão anterior. Vários núcleos do processador podem ser usados simultaneamente para alcançar ainda maiores velocidades de alinhamento.

O programa TopHat (Trapnell et al., 2009) foi criado para resolver a principal limitação do programa Bowtie. Na análise do sequenciamento foi utilizada a versão TopHat 2.0.9, *download* em Agosto de 2013. TopHat (Figura 8) usa Bowtie como um 'motor' de alinhamento e fragmenta *reads*, que o Bowtie não conseguiu alinhar por conta própria, em pequenos pedaços chamados segmentos. Muitas vezes, estas peças, quando processados de forma independente, irão alinhar ao genoma. Quando vários segmentos de uma *read* alinham no genoma em posições distantes uma da outra (por exemplo, entre 100 pb e várias centenas de quilo bases), TopHat infere que a *read* se estende por uma junção de processamento e estima onde os locais de junção estão. Ao processar cada *read* inicialmente não mapeável, TopHat pode construir um índice de locais de *splicing* no transcrito sem um gene *a priori* ou anotações do local das junções. Esta capacidade é crucial, porque, como numerosos estudos de RNA-seq têm mostrado agora, os catálogos de eventos de *splicing* alternativo permanecem incompletos. Mesmo nos transcritomas de organismos modelo frequentemente estudadas, novos eventos de *splicing* são descobertos a cada estudo adicional utilizando a técnica de RNA-Seq (Trapnell et

al., 2009; Trapnell et al., 2012).

O programa TopHat2 (Kim et al., 2013) incorpora os avanços presentes no programa Bowtie2 e permite que muitas das configurações do Bowtie2 possam ser aplicadas diretamente usando configurações disponíveis em TopHat2. Esta nova versão propõe solucionar um problema que ocorria em TopHat que era a impossibilidade de encontrar, em algumas situações, um alinhamento adequado. Tal situação ocorre pois transcritos de um genoma alvo pode diferir substancialmente do genoma de referência, possivelmente contendo inserções, exclusões e outras variações estruturais, como mostrado por Levy et al. (2007) e Xing et al. (2009). Em TopHat2 foram implantados novos procedimentos para garantir que as *reads* sejam alinhadas com as verdadeiras inserções e deleções (indels). *Indels* que ocorrem devido a erros de sequenciação e estes indels curtos podem ser detectados eficientemente por Bowtie2, o mecanismo subjacente de mapeamento de TopHat2. Além disso, grandes deleções e inversões no mesmo cromossomo, e translocações envolvendo diferentes cromossomos são detectadas pelos algoritmos TopHat-Fusion (Kim e Salzberg, 2011) que foram incorporados ao software TopHat2. Para as análises com o programa TopHat2 foram utilizadas as configurações *default* acrescido da opção *--report-secondary-alignment*.

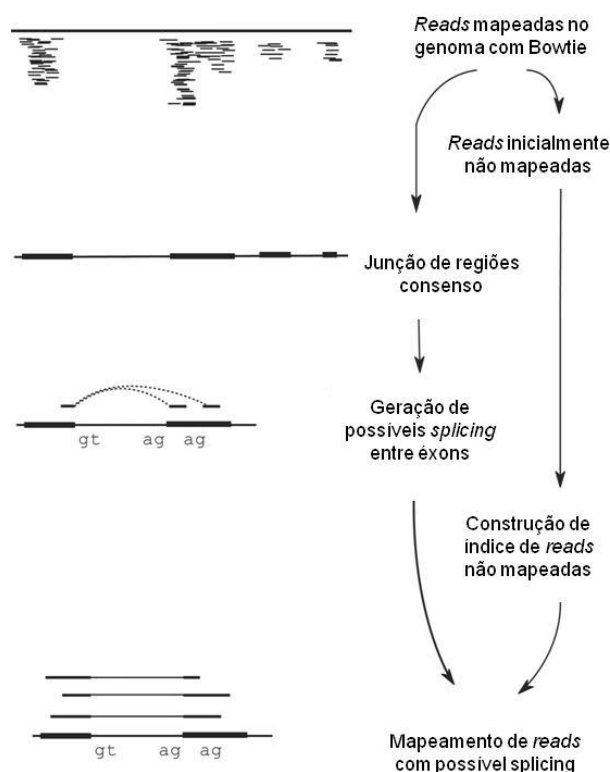


Figura 7. A pipeline de TopHat. Reads de RNA-Seq são mapeadas contra o genoma de referência, e aquelas reads que não são mapeadas são separadas. Um consenso inicial das regiões mapeadas é calculado utilizando o módulo de montagem Maq (Li et al., 2008). Sequências de flanking com potencial de doador / receptor de locais de splicing dentro de regiões vizinhas são unidas para formar potenciais junções de processamento alternativo. As reads inicialmente não mapeadas (IUM) são indexadas e alinhadas para estas sequências de junção de emendas. Fonte: Trapnell, et al., 2009.

As *reads* para cada replicata biológica são mapeadas de forma independente pelo programa. Estas *reads* mapeadas são providas como *input* para o programa Cufflinks (Trapnell et al., 2010), que produz um arquivo de *transfrags* (fragmentos transcritos) montados para cada replicata biológica. As contagens do número de transcritos em um experimento RNA-seq constitui-se de números inteiros não negativos e que, inerentemente, seguem uma distribuição discreta. Já as intensidades de cor obtidas em experimentos de *microarrays* são tratados como medições contínuas que comumente costumam seguir uma distribuição log-normal. Nos métodos desenvolvidos para a análise da expressão diferencial em experimentos de RNA-Seq estão disponíveis, atualmente, algoritmos de contagem que utilizam distribuição de Poisson, a distribuição binomial negativa (NB) e beta-binomiais (Soneson e Delorenzi, 2013).

Além do uso direto das distribuições discretas, alguns pacotes de software representam os dados de RNA-Seq por meio de quantidades transformadas tais como RPKM (*reads* por quilo base por milhão de *reads* mapeadas) descrita por

Mortazavi et al. (2008), como mostrado na Figura 9, ou a variação de RPKM denominada FPKM (fragmentos por quilo base por milhão de *reads* mapeadas) descrita por Trapnell et al. (2010) para utilização em sequenciamento com *reads* em *paired-end*. O objetivo de tais transformações é normalizar as contagens no que diz respeito aos diferentes tamanhos de biblioteca e também em relação ao comprimento dos transcritos. Isso acontece pois de um transcrito longo espera-se obter mais *reads* do que de um transcrito curto com o mesmo nível de expressão. Em experimentos que utilizam sequenciamento *paired-end* e, portanto, são geradas duas *reads* por fragmento. Isso não significa que ambas serão mapeadas ou que ambas tenham boa qualidade. Para corrigir essa possível distorção são contados fragmentos mapeados e não o número de *reads* mapeadas. De acordo com Trapnell et al. (2010) para os experimentos onde as *reads* obtidas no sequenciamento são no formato *single-end*, o valor de FPKM passa a ser o mesmo que RPKM. O valor de RPKM é calculado pela seguinte fórmula:

$$\text{RPKM} = \frac{10^9 \times \text{número de } \textit{reads} \text{ mapeadas em um mRNA}}{(\text{número total de } \textit{reads} \text{ mapeadas em 1 amostra}) \times (\text{somatório de éxons em par de bases})}$$

Estes arquivos de montagem foram mesclados com as anotações do transcriptoma de referência em uma anotação unificada para ser utilizada em análises posteriores. Como transcriptoma de referência foi utilizado o transcriptoma do genótipo Nipponbare composto por informações de CDS (sequências codificantes). Esta anotação mesclada é quantificada em cada condição por meio do programa Cuffdiff (que além das ferramentas Cuffcompare e Cuffmerge fazem parte da programa Cufflinks), que produz um conjunto de dados de expressão em arquivos tabulares. Estes arquivos são indexados e visualizados por meio do pacote estatístico CummeRbund (Goff et al., 2012) que foi desenvolvido para uso com o programa estatístico R (R Core Team, 2013). Esta ferramenta é utilizada para a exploração de genes diferencialmente expressos, *splicing* alternativo e de genes transcricionalmente regulados identificados pela ferramenta Cuffdiff.

Neste trabalho foram analisadas 12 combinações par a par visando identificar as diferenças na expressão gênica em cada uma delas. As combinações analisadas estão representadas na Figura 10. As combinações AC x AT, BAC x BAT, BNC x BNT, BNC x BNT e NC x NT visam elucidar as diferenças na expressão gênica quando os genótipos são submetidos ao estresse por frio. As combinações AC x

BAC e NC x BNC visam avaliar a similaridade dos *bulks* formados com as RILs mais próximas de cada genitor com o seu respectivo genitor na condição controle. As combinações AT x BAT e NT x BNT visam avaliar a similaridade dos *bulks* formados com as RILs mais próximas de cada genitor com o seu respectivo genitor quando submetidos ao estresse por frio. As combinações NC x AC e NT x AT foram analisadas para mensurar as diferenças entre os genitores na condição controle e na condição tratamento. Já as combinações BNC x BAC e BNT x BAT visam enumerar as diferenças entre os *bulks* na condição controle e tratamento, respectivamente.

Os genes sub e superexpressos foram designados da seguinte maneira: se o gene no segundo componente da comparação apresentar um nível de expressão significativamente maior ($FDR < 0.05$ e $\log_2\text{-fold-change} \geq |1.5|$) que o primeiro componente, estes genes foram considerados superexpressos. Já quando o segundo componente apresentar um nível de expressão menor que o primeiro componente, estes genes foram considerados subexpressos. A análise dos genes diferencialmente expressos foi realizada com a comparação entre as condições, sendo realizada a análise de 12 combinações conforme mostrado na Figura 10.

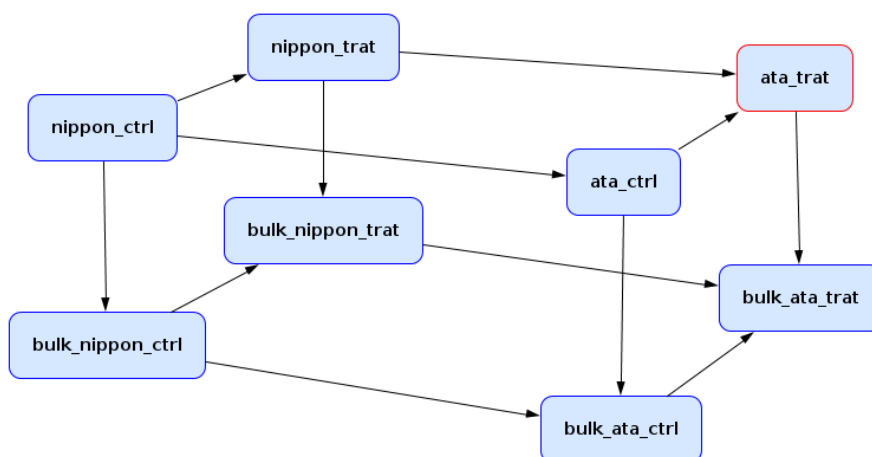


Figura 8. Esquema mostrando as combinações par a par analisadas para obtenção dos genes diferencialmente expressos entre cada combinação. A origem da seta indica a primeira condição e o destino da seta indica a segunda condição utilizada na comparação. nippon_ctrl: cultivar Nipponbare na condição controle; nippon_trat: cultivar Nipponbare na condição tratamento por frio; bulk_nippon_ctrl: *bulk* de RILs mais próximos da cultivar Nipponbare, na condição controle; bulk_nippon_trat: *bulk* de RILs mais próximos da cultivar Nipponbare, na condição tratamento; ata_ctrl: cultivar BRS Atalanta na condição controle; ata_trat: cultivar BRS Atalanta na condição tratamento por frio; bulk_ata_ctrl: *bulk* de RILs mais próximos da cultivar BRS Atalanta, na condição controle; bulk_ata_trat: *bulk* de RILs mais próximos da cultivar BRS Atalanta, na condição tratamento.

O programa Blast2GO (Conesa, et al., 2005) foi utilizado para fazer o alinhamento local (BLAST) (Altschul et al., 1990), com a opção 'blastn', ou seja, busca em banco de nucleotídeos utilizando uma sequência de nucleotídeos. Foram considerados genes superexpressos aqueles que apresentaram um $\log_2$ -fold-change $\geq 1,5$ (FRD $< 0,05$) e como subexpressos aqueles que apresentaram $\log_2$ -fold-change $\leq -1,5$ (FRD $< 0,05$). Somente genes que se enquadravam nestes critérios foram utilizados nesta análise. Os bancos de dados *Gene Ontology* (Ashburner et al., 2000) e KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa e Goto, 2000; Kanehisa et al., 2014), embarcados no programa Blast2GO, foram utilizados para a busca da ontologia gênica e das rotas metabólicas, respectivamente, para os genes diferencialmente expressos. Diagramas de Venn (Oliveros, 2007) foram utilizados para mostrar os genes diferencialmente expressos que estão presentes em uma combinação ou na outra e também dos genes que são diferencialmente expressos nas duas comparações, podendo estes apresentar diferenças significativas entre as comparações. Entre os genes diferencialmente expressos detectados para cada combinação foi feita a busca e identificação de TF. Para isso, usou-se os dados de anotação disponíveis no banco de dados PlantTFDB (*Plant Transcription Factor Database v3.0*) (Jin et al., 2014). Foram criados diversos scripts em linguagem de programação Perl para o tratamento específico dos resultados obtidos após o uso do protocolo Tuxedo. A Tabela 1 apresenta os *softwares* utilizados neste trabalho, com a respectiva versão utilizada, o autor e o *link* para o programa.

Tabela 1. Lista de *softwares*, com a versão, autor e página, utilizados nas análises realizadas neste trabalho.

Softwares utilizados			
<i>Softwares</i>	versão	Autor	<i>link</i>
Blast2Go	2.7.0	Conesa et al., 2005	http://www.blast2go.com/b2ghome
Bowtie2	2.1.0.0	Langmead et al., 2012	http://bowtie-bio.sourceforge.net/
Cufflinks	2.1.1	Trapnell et al., 2010	http://cufflinks.cbc.umd.edu/
CummeRbund	2.0.0	Goff et al., 2012	http://compbio.mit.edu/cummeRbund/
FastQC	0.10.1		http://www.bioinformatics.babraham.ac.uk/
SAM tools	0.1.18.0	Li et al., 2009	http://samtools.sourceforge.net/
TASSEL	-	Bradbury et al., 2007	http://sourceforge.net/projects/tassel/
TopHat2	2.1.0	Kim et al., 2013	http://tophat.cbc.umd.edu/
Trimmomatic	0.32	Lohse et al., 2012	http://www.usadellab.org/cms/?page=trimmomatic
VENNY	-	Oliveros, 2007	http://bioinfogp.cnb.csic.es/tools/venny/index.html

4 Resultados e Discussão

Na Figura 11 estão representadas, por meio de gráfico do tipo boxplot, as 24 bibliotecas analisadas no trabalho. Nelas foram avaliadas características qualitativas para os dados normalizados. Os valores demonstrados encontram-se em uma escala $\log_{10}$ que foram obtidos após a normalização FPKM, ou seja, fragmentos por quilo base por milhão de fragmentos mapeados. Esta normalização corresponde a RPKM (*reads* por quilo base por milhão de *reads* mapeadas) para *reads single-end*, como é o caso deste trabalho. Estes valores representam o conjunto total de genes identificados em cada biblioteca. Neste gráfico são apresentados os valores mínimos e máximos, os quartis superiores e inferiores, e a mediana obtida para cada biblioteca. Podemos visualizar que todas as bibliotecas apresentam medianas muito próximas uma da outra. O quartil superior foi, da mesma forma, muito semelhante para todas as bibliotecas analisadas. Em relação ao quartil inferior é possível visualizar que bibliotecas que compõem as repetições biológicas para a cultivar BRS Atalanta na condição tratamento e também na condição controle apresentam os menores valores (valores mais negativos). Contudo, essas diferenças são pequenas, o que pode ser confirmado pela pequena amplitude entre o quartil superior e inferior. Tais resultados indicam que o método de normalização FPKM foi eficaz na estabilização da contagem de *reads* em todas as repetições. Quando as bibliotecas obtidas apresentam pequenas diferenças em seu tamanho e pouca variabilidade inter-amstral, muitas vezes as normalizações não são necessárias. No entanto, quando existem grandes diferenças de tamanho e na variabilidade inter-amstral torna-se necessário o uso de normalização de dados (Dillies et al., 2013).

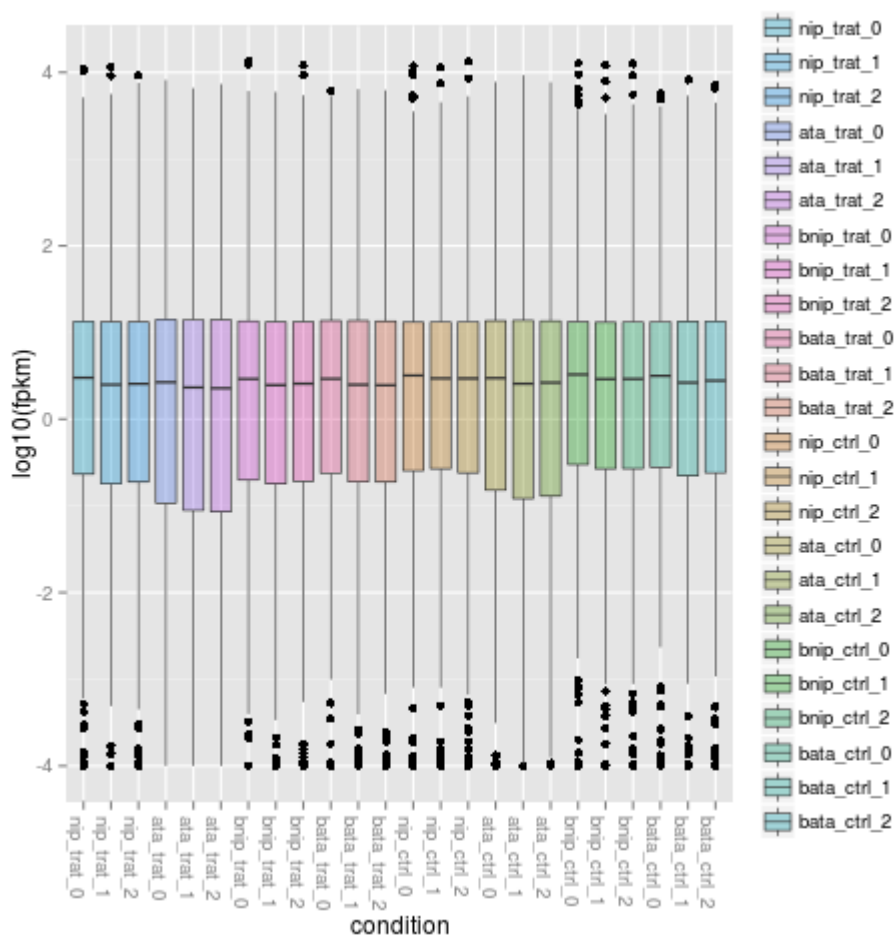


Figura 9. Gráfico do tipo boxplot onde são apresentados os valores mínimo e máximo, os quartis superiores e inferiores, e a mediana para cada uma das 24 bibliotecas analisadas. No eixo y estão representados os valores de expressão, após a normalização por meio do método FPKM, em escala $\log_{10}$. No eixo x são apresentadas cada condição com suas replicatas biológicas. nip_trat_0 equivale a repetição biológica 1. nip_trat_1 equivale a repetição biológica 2 e nip_trat_2 equivale a repetição biológica 3. nip_trat: Nipponbare tratamento; ata_trat: BRS Atalanta tratamento; bnip_trat: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição tratamento; bata_trat: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição tratamento; nip_ctrl: Nipponbare controle; ata_ctrl: BRS Atalanta controle; bnip_ctrl: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição controle; bata_ctrl: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição controle. Criado com o pacote CummeRbund.

Na Tabela 2 podemos observar que o número de *reads* varia consideravelmente. A biblioteca com maior número de *reads* é a primeira biblioteca da condição que corresponde ao *bulk* de RILs mais próximas a cultivar BRS Atalanta sob tratamento (BAT), onde foram obtidas 52 milhões de *reads*. No extremo oposto encontra-se a primeira biblioteca da condição formada pelo *bulk* de RILs mais próximos do genótipo Nipponbare sob tratamento (BNT), onde foram obtidas menos que 10 milhões de *reads*.

Tabela 2. Número de *reads* usadas pelo programa TopHat para realizar o alinhamento com o genoma e o transcrito de referência. Total de *reads* representa o input no programa, número de *reads* mapeadas e porcentagem frente ao total de *reads*, alinhamentos únicos representam *reads* que alinharam em um único local no genoma e a porcentagem frente ao total de *reads*. Alinh. multi-posições e a sua porcentagem representam *reads* que alinharam em mais de um local no genoma. Leituras não mapeadas e sua porcentagem representam *reads* que não alinharam com o genoma e o transcrito de referência.

	AC	AC	AC	Somatório
Total de <i>reads</i>	49205061	39741334	42461731	131408126
<i>Reads</i> mapeadas	42492082	34725736	36195669	113413487
%	86.4	87.4	85.2	86.3
Alinhamentos únicos	39729209	32680682	33034993	105444884
%	79.9	81.5	76.5	79.3
Alinh. multi-posições	2762873	2045054	3160676	7968603
%	6.5	5.9	8.7	7.0
Leituras não mapeadas	6712979	5015598	6266062	17994639
%	13.4	12.6	14.8	13.6
	AT	AT	AT	Somatório
Total de <i>reads</i>	35698677	28987034	31287128	95972839
Leituras mapeadas	30341395	20267963	27386226	77995584
%	85.0	69.9	87.5	80.8
Alinhamentos únicos	28746400	19048784	25989526	73784710
%	79.7	63.9	82.4	75.3
Alinh. multi-posições	1594995	1219179	1396700	4210874
%	5.3	6.0	5.1	5.5
Leituras não mapeadas	5357282	8719071	3900902	17977255
%	15.0	30.1	12.5	19.2
	NC	NC	NC	Somatório
Total de <i>reads</i>	36044731	32470379	14280561	82795671
Leituras mapeadas	31519707	27148907	13227042	71895656
%	87.4	83.6	92.6	87.9
Alinhamentos únicos	25226282	19716106	12342506	38352037
%	80	56.3	85.9	74.1
Alinh. multi-posições	6293425	7432801	884536	33543619
%	7.4	27.4	6.7	13.8
Leituras não mapeadas	4525024	5321472	1053519	10900015
%	12.6	16.4	7.4	12.1
	NT	NT	NT	Somatório
Total de <i>reads</i>	25196706	28941792	28647892	82786390
Leituras mapeadas	22354688	24103683	25586232	72044603
%	88.7	83.3	89.3	80.1
Alinhamentos únicos	20133928	20786329	23848964	64769221
%	78.8	69.5	82.5	76.9
Alinh. multi-posições	2220760	3317354	1737268	7275382
%	9.9	13.8	6.8	10.2
Leituras não mapeadas	2842018	4838109	3061660	10741787
%	11.3	16.7	10.7	
	BAC	BAC	BAC	Somatório
Total de <i>reads</i>	34433927	27856758	31291191	93581876
Leituras mapeadas	30895450	25427178	27906308	84228936
%	89.7	91.3	89.2	90.1
Alinhamentos únicos	20671918	23367700	26322593	70362211
%	56.6	83.2	83.5	74.4
Alinh. multi-posições	10223532	2059478	1583715	13866725

%	33.1	8.1	5.7	15.6
Leituras não mapeadas	3538477	2429580	3384883	9352940
%	10.3	8.7	10.8	9.9
	BAT	BAT	BAT	Somatório
Total de <i>reads</i>	52025004	27908102	21282387	101215493
Leituras mapeadas	46720703	22337282	18830518	87888503
%	89.8	80.0	88.5	86.1
Alinhamentos únicos	43572569	21159937	17518628	82251134
%	83.1	74.7	81.5	79.8
Alinh. multi-posições	3148134	1177345	1311890	5637369
%	6.7	5.3	7.0	6.3
Leituras não mapeadas	5304301	5570820	2451869	13326990
%	10.2	20.0	11.5	13.9
	BNC	BNC	BNC	Somatório
Total de <i>reads</i>	42232054	52133973	39593933	133959960
Leituras mapeadas	38275763	44616657	36548167	119440587
%	90.6	85.6	92.3	89.5
Alinhamentos únicos	35736074	34027374	24617488	94380936
%	84.0	61.9	59.7	68.5
Alinh. multi-posições	2539689	10589283	11930679	25059651
%	6.6	23.7	32.6	21.0
Leituras não mapeadas	3956291	7517316	3045766	14519373
%	9.4	14.4	7.7	10.5
	BNT	BNT	BNT	Somatório
Total de <i>reads</i>	9888133	35367082	23233078	68488293
Leituras mapeadas	9054003	32448489	21470540	62973032
%	91.6	91.7	92.4	91.9
Alinhamentos únicos	7345740	23946939	19748406	51041085
%	72.7	65.5	84.4	74.2
Alinh. multi-posições	1708263	8501550	1722134	11931947
%	18.9	26.2	8	17.7
Leituras não mapeadas	834130	2918593	1762538	5515261
%	8.4	8.3	7.6	8.1

AC: BRS Atalanta controle; AT: BRS Atalanta tratamento; NC: Nipponbare controle; NT: Nipponbare tratamento; BAC: *Bulk* de RILs BRS Atalanta controle; BAT: *Bulk* de RILs BRS Atalanta tratamento; BNC: *Bulk* de RILs BRS Nipponbare controle; BNT: *Bulk* de RILs BRS Nipponbare tratamento. A porcentagem na coluna Somatório indica a porcentagem média obtida para cada replicata biológica.

O número de *reads* obtidas no sequenciamento é um fator importante para a detecção de genes diferencialmente expressos. Na Tabela 2 são apresentados os valores obtidos para cada biblioteca por ocasião do sequenciamento. Nela podemos observar os diversos parâmetros analisados em cada biblioteca. O número de *reads* mapeadas indica o número de *reads* que foram mapeadas ao genoma de referência por meio do programa TopHat2. O total de alinhamentos únicos para cada biblioteca indica o número de *reads* que alinharam em um único local no genoma de referência. O alinhamento multi-posições indica o número de *reads* que alinharam em mais de um local no genoma. O número de *reads* não mapeadas é obtido pela

subtração do total de *reads* pelo número de *reads* mapeadas. Para a cultivar BRS Atalanta, na condição controle (BAC) temos que o número de *reads* mapeadas variou entre 49 milhões, 39 milhões e 42 milhões de *reads* para as replicatas biológicas 1, 2 e 3, respectivamente. No somatório das bibliotecas para esta condição temos, então, que o somatório das *reads* obtidas foi de 131 milhões de *reads*. O número de *reads* mapeadas foi de 42 milhões, 34 milhões e 36 milhões, o que representa porcentagem de 86,4%, 87,4% e 85,2%, respectivamente. O número de *reads* que alinharam em um único local no genoma foi de 39 milhões, 34 milhões e 36 milhões de *reads* o que corresponde a 79,9%, 81,5% e 76,5% do total de *reads* obtidas, respectivamente. No somatório das três bibliotecas para esta condição foram obtidas 105 milhões de *reads* mapeadas em um único local no genoma, indicando um percentual médio de alinhamento de 79,3%. As *reads* que alinharam a mais de um local no genoma representam 2,7 milhões, 2,0 milhões e 3,1 milhões de *reads* representando 6,5%, 5,9% e 8,7% do total de *reads* obtido em cada biblioteca, respectivamente. No somatório foram obtidas 7,9 milhões de *reads* que se alinharam a mais de um local no genoma representando, em média, 7,0% do total de *reads*. As *reads* não mapeadas representam 6,7 milhões, 5,0 milhões e 6,2 milhões de *reads* totalizando 17,9 milhões de *reads* não mapeadas. Este montante corresponde a 13,6% do total de *reads* obtidas para esta condição por meio das três replicatas biológicas utilizadas. Na condição que constitui o *bulk* das RILs mais próximas ao genótipo Nipponbare na condição controle (BNC) foi obtido o maior número de *reads* no sequenciamento, chegando a 133 milhões de *reads*. Já no somatório do número de alinhamentos únicos quem se destacou foi a condição BRS Atalanta controle, com um montante de 105 milhões de *reads* alinhadas em um único local no genoma.

Neste trabalho, para cada condição, foram utilizadas três replicatas biológicas, pois é conhecido que o material biológico, mesmo sendo proveniente de indivíduos homozigotos, pode apresentar variações entre si. Quanto à questão de utilizar um maior número de repetições ou uma maior profundidade de sequenciamento, preferiu-se utilizar um maior número de repetições biológicas. Liu et al. (2013), mostraram a existência de um dilema entre ter maior número de repetições biológicas ou uma maior profundidade de sequenciamento visando o aumento do poder para detectar genes diferencialmente expressos. Tendo como base o estudo da linha celular humana MCF7, foi verificado que o aumento da

profundidade de sequenciamento além de 10 milhões de *reads* apresenta retornos decrescentes sobre o poder de detectar genes diferencialmente expressos, enquanto que a adição de repetições biológicas melhora o poder de detecção de forma significativa, independentemente da profundidade de sequenciamento. O aumento da profundidade de sequenciamento só teria justificativa se o objetivo do trabalho estiver relacionado a expressão diferencial de éxons e a transcrição de transcritos específicos.

Assim como os demais experimentos conduzidos, há a necessidade da condução do experimento, coleta de dados e análise dos dados obtidos levando em conta preceitos estatísticos como amostragem, casualização, repetições e a condução em modelo estatístico adequado (Auer e Doerge, 2010). Marioni et al. (2008) demonstrou a ausência da necessidade do uso de replicatas técnicas de RNA-Seq, realizados em sequenciadores Illumina, uma vez que os resultados se apresentam altamente replicáveis e com relativamente pouca variação técnico. Outro ponto de grande importância se refere ao uso de qRT-PCR juntamente com os dados obtidos por RNA-Seq. A validação usando qRT-PCR nas mesmas amostras de RNA testados na análise de RNA-seq só valida a tecnologia na qual o sequenciamento foi realizado. Ele não valida a conclusão sobre os tratamentos/condições (Fang e Cui, 2011). As conclusões biológicas a partir de experimentos de RNA-seq somente podem ser validadas por meio da utilização de diferentes repetições biológicas.

Na Figura 12 é apresentado um dendrograma com distâncias de Jensen-Shannon que mostra a dissimilaridade entre as 24 bibliotecas analisadas. No dendrograma são formados dois ramos com uma altura de 0.1375. Percebe-se que ocorre uma clara divisão entre os genótipos, onde as bibliotecas correspondentes a genótipo Nipponbare e as RILs mais próximas a ela estão em um dos ramos e no outro ramo está a cultivar BRS Atalanta e as RILs correlatas. Estes resultados demonstram a eficiência na seleção de RILs mais próximos a cada um dos genitores (BRS Atalanta e Nipponbare) por meio da ferramenta *GBS analysis*, uma extensão do programa TASSEL (Bradbury et al., 2007). A divergência de Jensen-Shannon é um método para mensurar a dissimilaridade entre duas distribuições de probabilidade, sendo também conhecida como divergência total da média. Ela é baseada na divergência de Kullback-Leibler, tendo como principais diferenças a sua condição simétrica e por sempre constituir um valor finito (Lin, 1991). Seus valores

são limitados entre 0 e 1. A raiz quadrada da divergência de Jensen-Shannon é uma medida muitas vezes referida como distância de Jensen-Shannon.

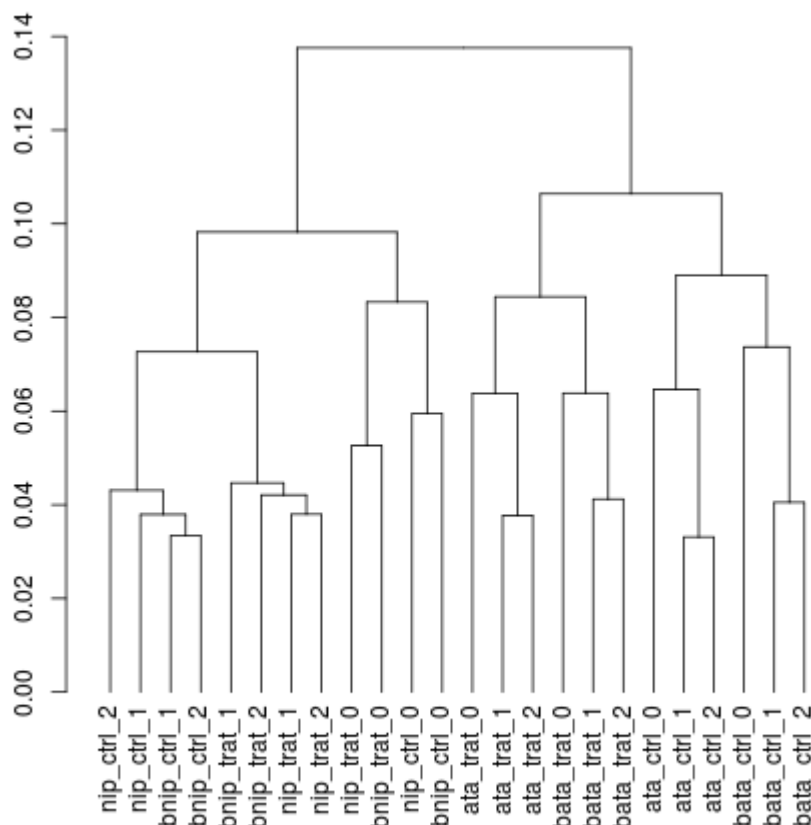


Figura 10. Dendrograma de distâncias de Jensen-Shannon mostrando as distâncias entre as condições em estudo. nip_trat: Nipponbare tratamento; ata_trat: BRS Atalanta tratamento; bnip_trat: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição tratamento; bata_trat: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição tratamento; nip_ctrl: Nipponbare controle; ata_ctrl: BRS Atalanta controle; bnip_ctrl: *Bulk* de RILs mais próximos a cultivar Nipponbare, condição controle; bata_ctrl: *Bulk* de RILs mais próximos a cultivar BRS Atalanta, condição controle. Criado com o pacote CummeRbund.

Na Tabela 3 podemos observar o número total de genes diferencialmente expressos. Este número condiz com o número de genes para o qual foi rejeitada a Hipótese de nulidade (H_0), com uma taxa de falsas descobertas (FDR) fixado em 5% e um $\log_2$ -fold-change ≥ 1.5 ou ≤ -1.5 . Foram analisadas 12 combinações, duas a duas, para identificar o número de genes diferencialmente expressos em cada uma delas. A combinação entre a cultivar BRS Atalanta na condição controle e a cultivar BRS Atalanta, submetida ao frio de 13°C por 24h, apresentou o maior número de genes diferencialmente expressos, com 15.673 genes. As combinações BAC x BAT, BNC x BNT, NC X NT e NT x AT também

apresentaram um elevado número de genes diferencialmente expressos, situando-se entre 12 mil e 13 mil genes. O menor número de genes diferencialmente expressos foi obtido nas combinações NC x BNC e NT x BNT que apresentaram apenas 142 e 321 GDE, respectivamente. Este número indica que o *bulk* formado com as RILs mais próximas ao genitor Nipponbare apresentam pequenas diferenças em relação ao genitor. Os indivíduos que constituem os *bulks* apresentam respostas muito parecidas ao genitor tanto na condição controle bem como quando expostos ao estresse por frio. As combinações AC x BAC e AT x BAT apresentaram 3318 e 5381 genes diferencialmente expressos. Estas duas combinações, juntamente com a combinação NT x BNT, apresentaram um número de genes com log2-fold-change ≥ 1.5 maior que o número de genes com log2-fold-change ≤ -1.5 . Nas demais combinações o comportamento foi o inverso, com um maior número de genes subexpressos (com log2-fold-change ≤ -1.5) em relação ao número de genes superexpressos.

Tabela 3. Genes diferencialmente expressos (GDE) entre as diversas combinações. Número total de genes diferencialmente expressos obtidos com o uso do protocolo Tuxedo (Trapnell et al., 2012). Número de genes que apresentaram uma expressão diferencial com log2-fold-change maior ou igual a 1,5 e número de genes que apresentam um nível de expressão diferencial com log2-fold-change menor ou igual a -1,5. Número de genes que estão anotados no banco de dados de arroz RAP-DB.

Combinação	Genes diferencialmente expressos				
	Total	Log2FC ≥ 1.5	Anotados	Log2FC ≤ -1.5	Anotados
AC x AT	15673	848	778	1464	1364
AC x BAC	3318	454	413	62	53
AT x BAT	5381	401	380	50	36
BAC x BAT	13199	770	708	1226	1091
BNC x BAC	7557	537	460	1481	1265
BNC x BNT	13583	446	404	1226	1101
BNT x BAT	8439	647	597	1312	1151
NC x AC	9966	595	557	1327	1242
NC x BNC	142	5	5	45	20
NC x NT	13350	402	374	1200	1084
NT x AT	12878	1017	927	1345	1255
NT x BNT	321	31	18	6	6

AC: BRS Atalanta controle; AT: BRS Atalanta tratamento; BAC: *Bulk* de BRS Atalanta controle; BAT: *Bulk* de BRS Atalanta tratamento; BNC: *Bulk* Nipponbare controle; BNT: *Bulk* Nipponbare tratamento; NC: Nipponbare controle; NT: Nipponbare tratamento.

Para os genes que apresentaram um log2-fold-change ≥ 1.5 ou ≤ -1.5 foi feita busca da anotação gênica no banco de dados RAP-DB (RAP-DB, 2013). Os resultados obtidos estão dispostos na Tabela 3 e os genes para qual foi encontrada a anotação estão representados na coluna “com ID”, tanto para os genes

superexpressos quanto para aqueles que foram subexpressos nas 12 combinações analisadas. Pode ser observada uma considerável diferença entre o número de genes diferencialmente expressos e o número de genes para os quais foi encontrada uma ID. Isso indica que foram detectados genes diferencialmente expressos que ainda não estão anotados no banco de dados RAP-DB. Desta forma, este trabalho corrobora com Trapnell et al. (2012) que afirma que, mesmo para os organismos bem estudados e anotados, a maioria dos experimentos de RNA-Seq revela novos genes e transcritos. Este autor cita um trabalho de Cabili et al. (2011) onde foram realizadas análises de amostras de RNA-Seq, com grande profundidade de sequenciamento, de 24 tecidos humanos e linhas celulares, sendo obtidos mais de 8.000 novos RNAs não-codantes, juntamente com numerosos potenciais genes codificadores de proteínas.

Na Figura 13 podemos observar, por meio de um diagrama de Venn, os genes subexpressos para quatro comparações. Estas se referem às combinações: AC x AT, BAC x BAT, BNC x BNT e NC x NT. A combinação AC x AT apresentou 1380 genes que estão subexpressos exclusivamente nesta comparação. Já para BAC x BAT o número de genes exclusivos foi de 504 genes subexpressos. Para as combinações BNC x BNT e NC x NT este número foi de 323 e 282 genes subexpressos, respectivamente. Quanto aos genes que foram subexpressos simultaneamente em mais de uma combinação podemos citar que 424 genes se mostraram subexpressos nas combinações BAC x BAT, BNC x BNT e NC x NT. Foram obtidos 128 genes subexpressos em comum para as combinações BAC x BAT e NC x NT, enquanto que 120 genes em comum foram subexpressos entre BAC x BAT e BNC x BNT. Um total de 20 genes foram subexpressos nas quatro comparações.

fecundação e embriões com sete dias após a fecundação. Neste mesmo trabalho, na comparação de embriões aos sete dias e aos 14 dias após a fecundação foram obtidos 376 genes subexpressos e 128 genes superexpressos.

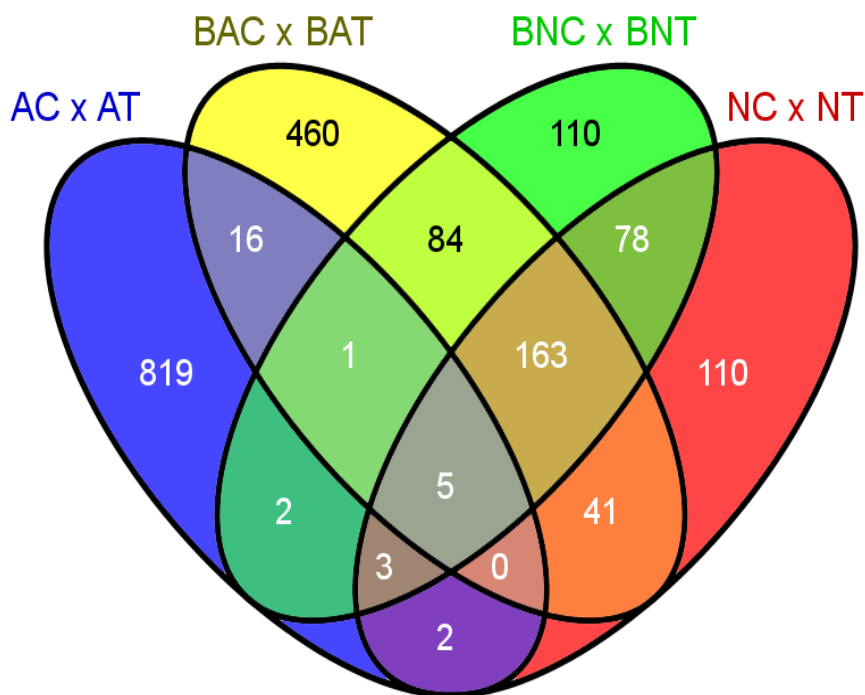


Figura 12. Diagrama de Venn apresentando genes superexpressos para a comparação entre a condição controle e a condição tratamento para os genótipos BRS Atalanta, Nipponbare e para os *bulks* de RILs que se mostraram mais próximos de cada genitor. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento.

Na Figura 15 estão representados os genes que codificam fatores de transcrição (TF), de acordo com o banco de dados de fatores de transcrição em plantas PlantTFDB (Jin et al., 2014), que se mostraram superexpressos nas combinações AC x AT, BAC x BAT, BNC x BNT e NC x NT. Nas combinações AC x AT e BAC x BAT, 24 e 32 genes estão superexpressos em cada condição, respectivamente. Na combinação BNC x BNT quatro TFs foram superexpressos exclusivamente nesta combinação: Os05g0415400 e Os09g0552900 (NAC), Os02g0177300 (SBP), Os07g0549600 (bHLH). Oito genes estão superexpressos em três combinações, que são: BAC x BAT, BNC x BNT e NC x NT, sendo eles: Os02g0232000 e Os06g0553100 (HSF), Os12g0578600 (C2H2), Os01g0911700 (B3), Os04g0583900 (MYB-related), Os04g0460600 e Os02g0594800 (NAC),

Os09g0309700 (ERF). Os TFs que se mostram superexpressos nessas três comparações são muito importantes, uma vez que foram transferidos do genitor tolerante Nipponbare para as RILs mais próximas ao genitor sensível BRS Atalanta e assim podem ser muito relevantes na obtenção de novas cultivares que mantenham as características agrônômicas idênticas ao genitor BRS Atalanta, mas com um incremento no *background* para tolerância ao frio. A mesma análise pode ser feita para os quatro genes que foram superexpressos na combinação BAC x BAT e em NC x NT: Os01g0658900 e Os03g0809200 (bZIP), Os06g0184000 (bHLH) e Os01g0733200 (HSF). Nenhum gene está superexpresso em todas as combinações.

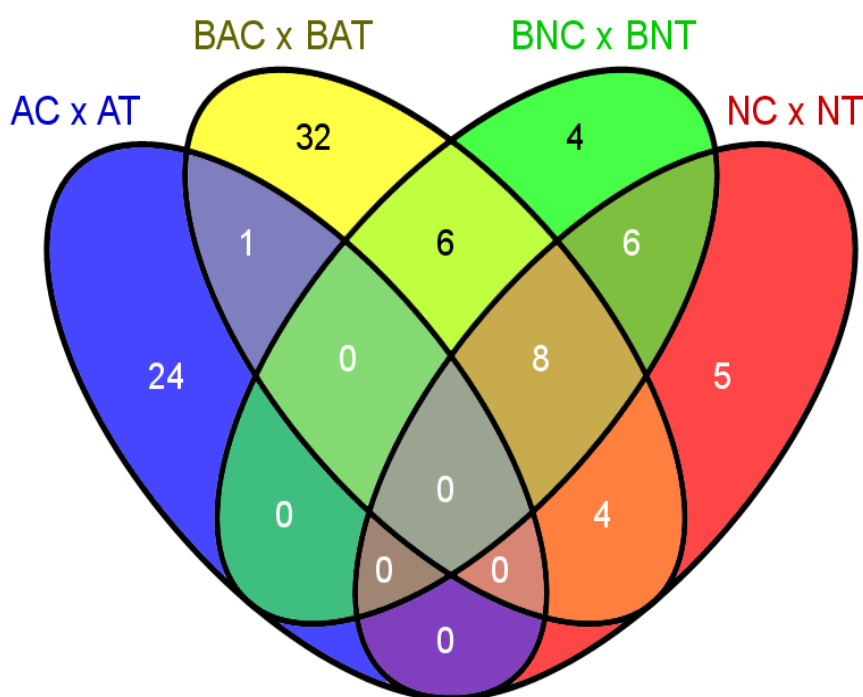


Figura 13. Fatores de transcrição superexpressos envolvendo a condição controle e a condição tratamento para as combinações: AC x AT, BAC x BAT, BNC x BNT e NC x NT. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento.

Na Figura 16 são apresentados os fatores de transcrição que se mostram subexpressos nas quatro combinações avaliadas. Na combinação AC x AT, 65 TF são exclusivos desta comparação. Além destes, outros dois fatores de transcrição que fazem parte desta combinação se apresentam subexpressos nas quatro

combinações, sendo eles: Os04g0304400, que pertence à família MIKC e o gene Os08g0196700, que pertence à família NF-YA de TFs. As comparações BAC x BAT e BNC x BNT apresentam 16 e 12 fatores de transcrição unicamente subexpressos naquela comparação, respectivamente. Um total de 15 genes são subexpressados em três combinações, exceto em AC x AT. Três genes são subexpressos tanto em BAC x BAT quanto em NC x NT, sendo que todos pertencem a família bHLH de TF (Os01g0108400, Os03g0188400 e Os06g0226500). Genes desta família de TFs estão descritos como apresentando funções importantes na tolerância ao estresse por frio pela remoção de ERO pela regulação positiva da enzima peroxidase (Huang et al., 2013). Quatro genes são subexpressos entre os *bulks*, sendo eles: Os01g0140700 (RAV), Os09g0401000 e Os06g0519400 (MYB) e Os04g0526000 (TCP).

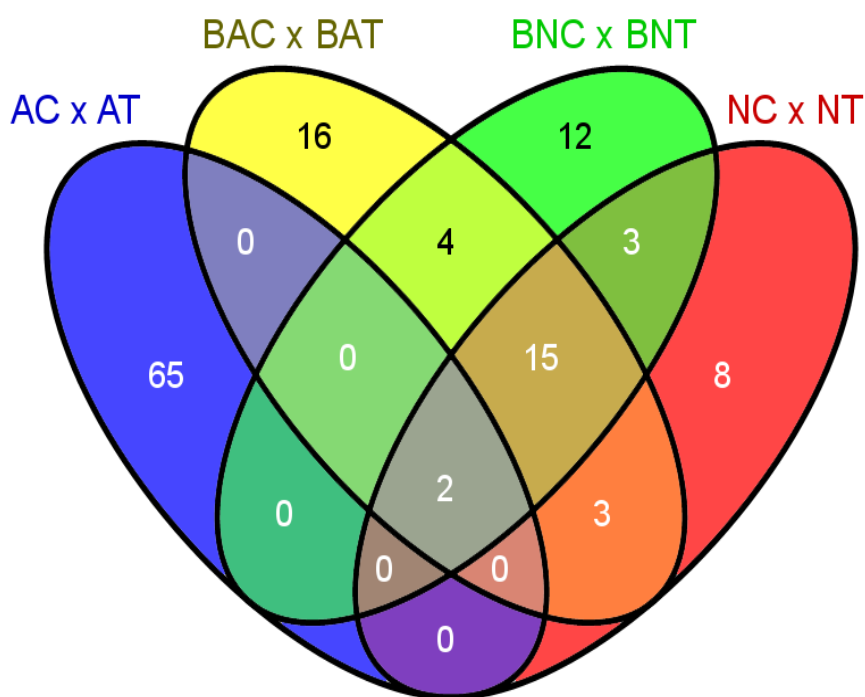


Figura 14. Fatores de transcrição subexpressos envolvendo a condição controle e a condição tratamento para as combinações: AC x AT, BAC x BAT, BNC x BNT e NC x NT. AC: BRS Atalanta controle, AT: BRS Atalanta tratamento, BAC: *bulk* de RILs de BRS Atalanta controle, BAT: *bulk* de RILs de BRS Atalanta tratamento, BNC: *bulk* de RILs de Nipponbare controle, BNT: *bulk* de RILs de Nipponbare tratamento, NC: Nipponbare controle, NT: Nipponbare tratamento.

Nos Apêndices A e B são apresentados os fatores de transcrição, identificados a partir do banco de dados PlantTFDB (Jin et al., 2014), que se mostraram diferencialmente expressos (subexpressos e superexpressos,

respectivamente) nas combinações avaliadas neste trabalho. As redes de regulação transcricional são importantes na resposta ao estresse abiótico em *Arabidopsis* e gramíneas (Chinnusamy et al., 2004; Sunkar et al., 2007). Os fatores de transcrição são reguladores-chave devido à capacidade de um único TF controlar a expressão de muitos genes-alvo a jusante (Mittal et al., 2012). Diversas famílias de TFs são descritas na literatura como responsivas ao estresse por frio em arroz.

Em condições de frio, as plantas crescem mais lentamente e algumas mostram até mesmo defeitos ou danos no crescimento. Algumas destas alterações de crescimento induzidas pelo frio podem ser atribuídas à diminuição da fotossíntese e as atividades metabólicas geralmente mais baixas no frio (Kubien et al., 2003). A alteração no crescimento e desenvolvimento das plantas em baixas temperaturas pode estar relacionado a inibição dos genes de expansinas, que atuam na expansão celular. Além disso, a alteração da homeostase da auxina e, possivelmente, outros fitormônios, tais como o etileno, GA, e brassinosteroides também pode perturbar o desenvolvimento das plantas em baixas temperaturas. Fatores de transcrição relevantes para o desenvolvimento (ARF, GRAS, homeobox, MADS e NAC) também são responsivos ao frio sugerindo seu envolvimento potencial na reprogramação do desenvolvimento da planta sob estresse frio (Lee et al., 2005).

A família ERF apresenta um expressivo número de genes diferencialmente expressos nas diferentes combinações e a grande maioria deles se mostrou subexpressos nas combinações analisadas. Na combinação BNC x BNT, oito genes desta família estão subexpressos. A família ERF é uma grande família de genes de fatores de transcrição e faz parte da superfamília AP2/ERF, que também contém as famílias AP2 e famílias RAV. A família ERF é por vezes dividida em duas subfamílias principais, a subfamília ERF e a subfamília CBF/DREB. Muitas proteínas na família ERF foram identificadas e relacionadas a diversas funções nos processos celulares, tais como a transdução de sinal hormonal, respostas a estresses bióticos e abióticos, regulação do metabolismo, e em processos de desenvolvimento em diversas espécies de plantas (Nakano et al., 2006). Estes autores reportaram, também, que três membros da família CBF/DREB1, CBF1, CBF2, e CBF3 (ou DREB1b, DREB1c, e DREB1a, respectivamente), são induzidos após 15min de exposição das plantas à temperaturas baixas, e após 2h de exposição ocorre a expressão do *regulon* CBF em genes alvo, cujos promotores contém o elemento regulador CRT/DRE. O *regulon* CBF inclui genes que atuam em conjunto para

melhorar a tolerância ao frio. A sobre-expressão de fatores de transcrição nas CBF/DREB1 transgênicas de *Arabidopsis* resulta na acumulação de solutos compatíveis que têm actividades de crioprotetores, que incluem prolina, sacarose, e rafinose (Gilmour et al., 2000). Os fatores de transcrição da família RAV são rapidamente induzidos em resposta a exposição de plantas a baixas temperaturas, sendo acumulados após 1h de transferência das plantas para baixas temperaturas (Fowler e Thomashow, 2002). Estes autores demonstraram que TFs se expressam em paralelo com genes da família CBF (*cold binding factors*) e codificam um elemento de ligação ao DNA do tipo AP2. No banco de TFs PlantTFDB são descritos, atualmente, 5 genes desta família em *Oryza sativa* subsp. *indica* e 4 genes em *Oryza sativa* subsp. *japonica*. Genes desta família se mostraram subexpressos nas combinações BAC x BAT e BNC x BNT, e superexpressos na combinação BNC x BAC.

Os fatores de transcrição HSF (*heat shock factors*) são os principais reguladores da resposta de plantas a estresse por calor. Os TFs desta família se mostraram na grande maioria superexpressos neste trabalho e estão presentes, principalmente nas comparações que envolvem os genótipos e os *bulks* em condições controle x tratamento. A superexpressão desta família de TFs nas condições tratadas com frio está ligada a um fenômeno denominado tolerância a baixas temperaturas induzida por choque térmico (HSICT- *heat-shock-induced chilling tolerance*). A injúria por frio manifesta uma variedade de sintomas visíveis, muitas vezes utilizados como indicadores de sua gravidade, incluindo a redução do vigor do crescimento, amadurecimento anormal, estimulação da respiração e a produção de etileno, comprometimento da membrana celular e susceptibilidade à doença (Saltveit e Morris, 1990). Estes sintomas podem ser reduzidos pelo tratamento de choque térmico antes da exposição a baixas temperaturas (Sato et al., 2001). Li et al. (2003) sugere que os genes Hsfs tem papel essencial na tolerância a baixas temperaturas induzida por choque térmico e a expressão constitutiva na regulação da transcrição destes genes em espécies sensíveis ao frio pode ser útil na em aumentar a tolerância ao frio nestas culturas. As proteínas Hsps funcionam como chaperonas moleculares e formam complexos temporários com polipeptídeos desdobrados ou ou mal dobrados. Em sua interação com outros Hsps, estas proteínas protegem o citoesqueleto, a estrutura e função celular. Assim, proteínas Hsp podem travar ou impedir a desnaturação irreversível de proteínas induzida por

vários stresses bióticos e abióticos, incluindo o stresse térmico (Feder e Hofmann, 1999). Em algumas espécies, Hsps podem ser induzidas por exposição a frio, tanto a curto e longo prazo e têm um impacto importante sobre a tolerância ao frio (Waagner et al., 2013).

Os TFs MYB são fatores-chave para redes de regulação que controlam o desenvolvimento, metabolismo e resposta a estresses bióticos e abióticos. Os fatores de transcrição MYB e MYB-related foram encontrados neste trabalho sem apresentar um padrão de comportamento entre as combinações e entre as condições controle e tratamento. A expressão do gene *Osmyb4*, detectado com baixo nível de expressão em coleóptilos arroz - cultivadas durante 3 dias a 29°C, porém é fortemente induzida por tratamentos a 4°C. As temperaturas subletais de 10 e 15°C, a sua expressão em mudas de arroz já é evidente, mas este efeito não pode ser verificado em outros estresses e tratamento com ABA (Vannini et al., 2004). Outro gene, *MYB3R-2*, que tem expressão induzida por frio, seca e estresse salino aumenta a tolerância a estes estresses por meio da regulação de genes responsivos a estresses envolvidos nas vias CBF-dependentes ou CBF-independentes em *Arabidopsis* (Dai et al., 2007). O TF *MYBS3* é relatado na literatura como induzido por frio. A quantidade de mRNA de *MYBS3* foi detectada a 28°C e aumentou em 5x após exposição das plantas a 4°C por 72h (Su et al., 2010). A via de sinalização DREB1-dependente de resposta ao frio está entre aquelas reprimidas por *MYBS3* em arroz. Além disso, os autores sugerem que as vias *MYBS3* e DREB1-dependentes podem complementar-se e agir em sequência para se adaptar ao estresse pelo frio imediato e persistente em arroz.

Fatores de transcrição NAC estão envolvidos em resposta a estresses bióticos e abióticos. Esses fatores de transcrição podem ser efetores da sinalização por frio e atuar na reprogramação do desenvolvimento das plantas para lidar com o estresse por frio (Tran et al, 2004). Nakashima et al. (2007) descreveram o gene *OsNAC6*, sendo que sua expressão é induzida por estresses abióticos, incluindo frio, seca e salinidade. Expressão gênica *OsNAC6* também é induzida por fermentos e brusone.

Os genes WRKY têm sido sugeridos como uma família de TFs que desempenham um papel importante na regulação da reprogramação transcricional associada a respostas por stress em plantas. Modificações no padrão de expressão de genes e/ou alterações na sua atividade contribuem para a elaboração de várias

vias de sinalização e redes regulatórias. Além disso, um único gene WRKY muitas vezes responde a diversos fatores de estresse e, em seguida, suas proteínas podem participar na regulação de vários processos aparentemente díspares como reguladores negativos ou positivos. Um exemplo de gene da família WRKY que aumenta a tolerância ao estresse por frio em plantas transgênicas de *Arabidopsis* é o gene *GmWRKY21* originário da soja (*Glycine max* (L.) Merrill) (Zhou et al., 2008).

Além dos fatores de transcrição, inúmeras outras famílias gênicas estão envolvidas na indução da tolerância ao frio em arroz. Cheng et al. (2007) mostrou um módulo regulatório mediado por ERO (espécies reativas de oxigênio) que atuam como um componente inicial na via de resposta ao estresse por frio em arroz. Estes autores fizeram a análise do transcriptoma de uma cultivar de arroz *japonica* tolerante ao frio identificando um subconjunto de 121 genes de resposta precoce a este estresse e que foram sobrerregulados durante as primeiras 24h de exposição a uma temperatura de 10°C. Nesse grupo estavam quatro fatores de transcrição, incluindo ERO-bZIP1 e outro subgrupo maior, com a característica comum de ter os elementos as1/ocs-like em seus promotores. A indução por frio de ERO-bZIP1 precedeu a indução de genes contendo elementos as1/ocs-like e eles também foram induzidos por H₂O₂ exógeno à temperatura ambiente. Padrões de expressão coordenada e arquitetura semelhante do promotor entre os genes de resposta rápida sugerem que eles pertencem a um *regulon* potencial (módulo regulatório ERO-bZIP - as1/ocs) que responde a níveis elevados de ERO durante o estresse por frio. Padrões de expressão cultivar-específico dos genes candidatos indicam uma correlação positiva entre a atividade deste *regulon* putativo e variações genotípicas na tolerância ao frio em arroz.

Diversos genes que atuam na detoxificação de ERO foram encontrados nas combinações analisadas. De acordo com a anotação do banco de dados RAP-DB, na combinação AC x AT foram encontrados genes subexpressos com descritos como Glutathione S-transferase (Os01t0369700-01, Os07t0468100-01), Glutathione peroxidase (Os06t0185966-00), Peroxidase (Os07t0677600-01, Os01t0270300-01, Os01t0327400-01). Nos genes superexpressos nesta combinação foram encontrados os genes para Glutathione transferase (Os01t0370600-00, Os03t0134900-01) e Glutathione reductase (Os02t0813500-01). Na combinação NC x NT foram encontrados subexpressados os seguintes genes ligados a detoxificação de ERO: Glutathione S-transferase (Os01t0950000-01), Peroxidase (Os01t0270300-

01, Os07t0626700-01, Os05t0162000-01, Os10t0536600-00, Os01t0963000-01, Os07t0104500-01, Os10t0536700-01, Os04t0651000-01, Os05t0499300-01, Os07t0115300-01, Os06t0522100-00, Os09t0507500-01, Os02t0240300-01, Os03t0434800-00, Os02t0240100-01, Os06t0546500-01, Os04t0688500-01), Ascorbate peroxidase (Os07t0694700-01), Glutathione S-transferase (Os01t0950000-01). Nos genes superexpressos ligados a EROs encontram-se: Glutathione S-transferase (Os01t0558100-01, Os10t0530900-01), Glutathione reductase (Os10t0415300-01). Na combinação BAC x BAT os genes subexpressos ligados a ERO são: Glutathione S-transferase (Os11t0245100-01), Peroxidase (Os02t0240300-01, Os11t0112200-01, Os09t0507500-01, Os05t0499300-01, Os03t0434800-00, Os07t0677100-01, Os04t0656800-01, Os06t0522100-00, Os06t0521900-00, Os01t0963000-01, Os10t0536600-00, entre outros) e os genes superexpressos consistem de: Glutathione S-transferase (Os10t0525800-00, Os10t0481300-01, Os10t0528400-01, Os01t0369700-03, Os10t0415300-01, Os01t0558100-01, Os10t0529500-01, Os10t0530900-01). Já na combinação BNC x BNT estão: Glutathione S-transferase (Os01t0692000-01), Peroxidase (Os07t0104500-01, Os05t0499300-01, Os06t0522100-00, Os06t0546500-01, Os10t0536700-01, Os01t0270300-01, Os07t0626700-01, Os07t0694700-01, Os09t0507500-01, Os02t0240100-01, Os07t0115300-01, Os02t0240300-01, Os03t0434800-00, Os04t0651000-01). Os genes relacionados à ERO superexpressos são: Glutathione S-transferase (Os01t0558100-01, Os10t0415300-01, Os10t0529500-01, Os10t0530900-01), Peroxidase (Os07t0677200-01). Para a combinação AT x BAT, que apresentou um número maior de genes superexpressos em relação a genes subexpressos, não foram encontrados genes subexpressos ligados a detoxificação de ERO. Para os genes superexpressos foram encontrados Glutathione S-transferase (Os11t0245100-01), Peroxidase (Os05t0134800-00, Os07t0499500-01, Os07t0104500-01), Ascorbate peroxidase (Os12t0178200-01). Desta forma, a presença de um maior número de genes subexpressos relacionados à detoxificação de ERO vai contra a descrição na literatura de que ocorre indução destes genes quando a planta está submetida a estresse por frio, onde o estresse aumenta a transcrição, o nível de proteína e a atividade de diferentes enzimas detoxificadoras de ERO (Sato et al., 2001). Contudo, Mittler et al. (2004) sugere que o delicado equilíbrio entre a produção de ERO e a sua detoxificação é orquestrada por uma grande rede de genes que em *Arabidopsis* inclui mais de 150 genes. O

mesmo fazendo um apanhado dos dados de expressão destes genes mostrou que há diferenças no comportamento dentro de uma mesma família gênica para a resposta ao estresse por frio. No atual experimento, um grande número de genes ligados a detoxificação de ERO se mostrou diferencialmente expresso (sub e superexpresso), porém com log2-fold-change menor que o ponto de corte adotado neste trabalho.

Os resultados obtidos por meio da anotação gênica no banco de dado RAP-DB apresentam grandes diferenças em relação aos dados obtidos quando realizada a ontologia dos genes diferencialmente expressos por meio do software Blast2GO. Um exemplo disso pode ser encontrado na ontologia para função molecular na combinação AC x AT. No nível 2, a GO:0016209 contém três genes para atividade antioxidante (Os07t0677600-01, Os01t0270300-01, Os01t0327400-01), ou seja, os mesmos obtidos pela descrição gênica do RAP-DB. Contudo, dentro da GO:0003824, que engloba genes com atividade catalítica, são encontrados 24 genes para atividade de oxidoreductase onde poder estar genes referentes a Catalase, uma importante enzima participante do processo de detoxificação de ERO.

A dessaturação de ácidos graxos dos lipídeos da membrana é uma estratégia para que as plantas sobrevivam a temperaturas subótimas ou ao congelamento. Plantas sensíveis ao frio normalmente têm uma proporção mais elevada de ácidos graxos saturados e, por conseguinte, uma superior temperatura de transição, que corresponde à mudança da membrana de um estado semifluido para um estado semicristalino. Genótipos resistentes ao frio, por outro lado são marcados pela maior proporção de ácidos graxos insaturados e, correspondentemente, uma menor temperatura de transição (Mahajan e Tuteja, 2005). Murata e Los (1997) demonstraram que as baixas temperaturas causam uma diminuição na fluidez das membranas biológicas e que a capacidade das células para se aclimatar a frio é determinada pela sua capacidade para sintetizar os ácidos graxos insaturados para tornar a bicamada lipídica mais fluida e evitar a mudança de fase dos lipídios induzida pelo frio. Dessaturases de ácidos graxos são as enzimas que introduzem ligações duplas em ácidos graxos o que permite uma maior mobilidade das membranas a baixas temperaturas. Em plantas superiores, existem dois tipos principais de dessaturases de ácidos graxos: acil-ACP dessaturases solúveis que introduzem ligações duplas nas cadeias de acil ligados a ACP, e acil-lipídeo dessaturases ligados a membranas que apresentam ligações duplas ligados às

cadeias de acil esterificados com estruturas de glicerol (Los e Murata, 1998). Alguns genes com função de dessaturase foram encontrados na anotação por meio do RAP-DB. Na combinação AC x AT, o gene Os02t0504800-01 se mostrou subexpresso. Nas combinações BAC x BAT, BNC x BNT e NC x NT os genes Os07t0417200-01 e Os01t0880800-01 se mostraram subexpressos. Na literatura encontramos que os genes de desaturases possuem importante função na tolerância a estresses por frio e altas temperaturas (Routaboul et al., 2000; Wang et al., 2006; Prakash et al., 2009). Na análise do perfil de expressão de todo o genoma do arroz, Zhang et al., (2012) encontram genes de dessaturase superexpressos em resposta a frio em plântulas de arroz.

A Ontologia Gênica (GO) para os genes diferencialmente expressos para as diversas combinações analisadas neste trabalho são descritas a seguir. Três ontologias independentes são analisadas para cada combinação: função molecular, componente celular e processo biológico. Para a ontologia relacionada ao processo biológico na combinação AC x AT (Figura 17) foram encontradas GOs para 315 genes subexpressos e 199 GOs para genes superexpressos. Os principais processos biológicos afetados e que aparentemente estão mais ligados à resposta da planta ao estresse ao qual foi submetida são: sinalização (70% reprimidos e 30% induzidos), regulação biológica (65% reprimidos e 35% induzidos), processo do sistema imune (55% reprimidos e 45% induzidos), resposta a estímulos (60% reprimidos e 40% induzidos). Para o componente celular foram encontrados 734 genes subexpressos e 432 superexpressos que apresentaram GO. Os principais componentes celulares encontrados foram: célula (60% reprimidos e 40% induzidos), membrana (55% reprimidos e 45% induzidos), região extracelular (75% reprimidos e 25% induzidos), organela (60% reprimidos e 40% induzidos). Quanto à função molecular foram encontradas GOs para 266 genes subexpressos e para 162 genes superexpressos. As principais funções moleculares afetadas pelo estresse foram: atividade catalítica (60% reprimidos e 40% induzidos), *binding* (65% reprimidos e 35% induzidos), atividade de transporte (50% reprimidos e 50% induzidos), fator de transcrição com atividade de ligação ao ácido nucleico (80% reprimidos e 20% induzidos), atividade antioxidante (60% reprimidos e 40% induzidos), atividade reguladora de enzima (75% reprimidos e 25% induzidos), atividade de carreador de elétrons (100% reprimidos).

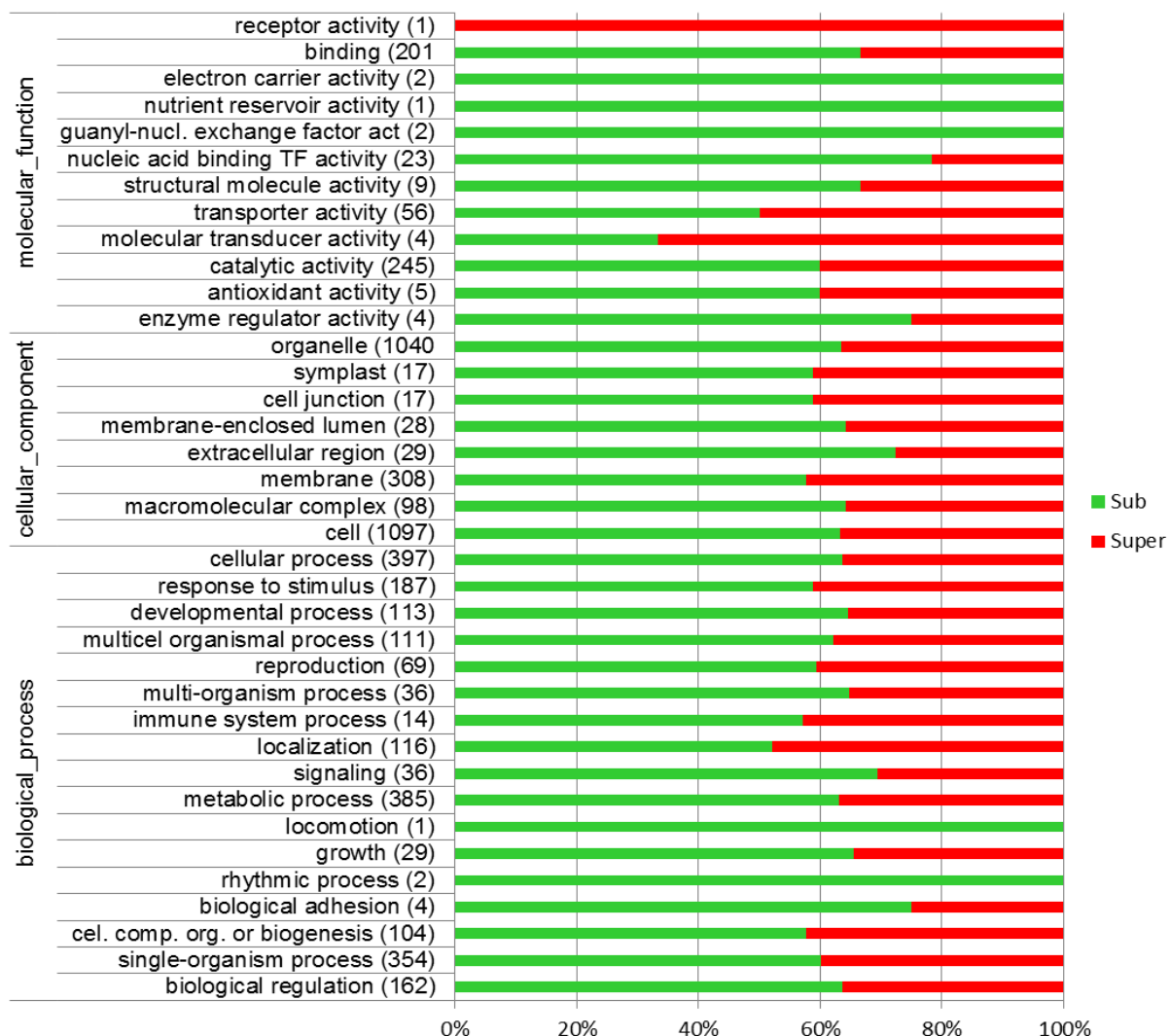


Figura 15. Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação AC x AT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação AC x AT.

Na Figura 18 estão representados os genes que se expressaram diferencialmente na combinação NC x NT. Para processo biológico foi encontrada ontologia para 219 genes subexpressos e para 109 genes superexpressos. Os principais processos metabólicos afetados e que aparentemente estão envolvidos com a resposta ao estresse por frio são: regulação biológica (65% reprimidos e 35% induzidos), sinalização (70% reprimidos e 30% induzidos), processos do sistema imune (55% reprimidos e 45% induzidos), resposta a estímulos (60% reprimidos e 40% induzidos). Para componente celular foram encontradas GOs para 616 genes subexpressos e para 210 genes superexpressos. Estes genes fazem parte dos

seguintes componentes celulares: organela (75% reprimidos e 25% induzidos), membrana (70% reprimidos e 30% induzidos), complexo macromolecular (60% reprimidos e 40% induzidos), região extracelular (65% reprimidos e 35% induzidos). Foram encontradas GOs para 189 genes subexpressos e para 75 genes superexpressos. As principais funções moleculares que foram afetadas pelo estresse e para os quais foram obtidas informações sobre sua ontologia foram: atividade catalítica (70% reprimidos e 30% induzidos), binding (75% reprimidos e 25% induzidos), atividade de transporte (70% reprimidos e 30% induzidos), atividade antioxidante (90% reprimidos e 10% induzidos), fatores de transcrição com atividade de ligação a ácido nucleico (65% reprimidos e 35% induzidos), atividade de carreador de elétrons (100% reprimidos) e atividade de regulação de enzimas (65% reprimidos e 35% induzidos).



Figura 16. Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação NC x NT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação NC x NT.

A anotação da ontologia gênica realizada para a combinação envolvendo o *Bulk* de RILs mais próximas genotipicamente ao genitor Nipponbare na condição controle versus a condição tratamento (BNC x BNT) (Figura 19) teve como resultado a identificação de GOs para a ontologia referente a processos biológicos, sendo 190 genes subexpressos e 117 para genes superexpressos. Os processos biológicos mais afetados pelo estresse foram: processos metabólicos (60% reprimidos e 40% induzidos), resposta a estímulos (65% reprimidos e 35% induzidos), regulação biológica (65% reprimidos e 35% induzidos), processo de desenvolvimento (45% reprimidos e 55% induzidos), sinalização (80% reprimidos e 20% induzidos),

processos do sistema imune (90% reprimidos e 10% induzidos). Em relação ao componente celular foi encontrada a GO para 613 genes subexpressos e para 215 genes superexpressos.

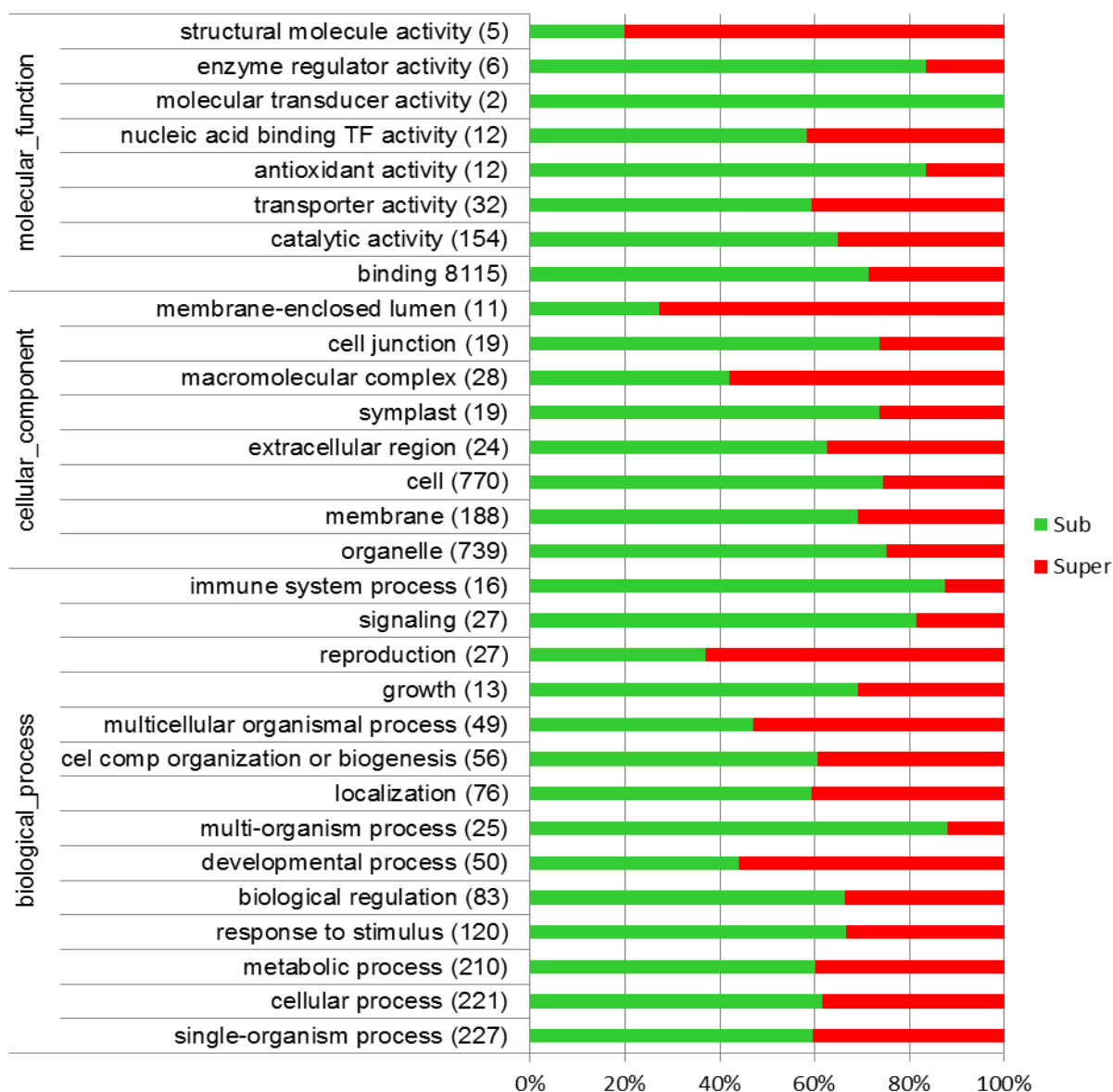


Figura 17. Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação BNC x BNT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação BNC x BNT.

Estes genes estão ligados principalmente aos seguintes componentes celulares: organela (65% reprimidos e 35% induzidos), região extracelular (55% reprimidos e 45% induzidos), simplasto (70% reprimidos e 30% induzidos). Os genes para os quais foram identificadas GOs referentes a sua função molecular perfazem

um total de 176 genes subexpressos e 86 para genes superexpressos. As principais funções moleculares as quais estes genes estão reportados são: *binding* (70% reprimidos e 30% induzidos), atividade catalítica (65% reprimidos e 35% induzidos), atividade de transporte (60% reprimidos e 40% induzidos), atividade antioxidante (85% reprimidos e 15% induzidos), atividade de fatores de transcrição que se ligam ao ácido nucleico (60% reprimidos e 40% induzidos), atividade reguladora de enzimas (85% reprimidos e 15% induzidos).

A busca pela ontologia gênica para a combinação BAC x BAT (Figura 20), formada pelo *bulk* de RILs mais próximas da cultivar BRS Atalanta, nas condições controle e tratamento por frio resultou na obtenção da ontologia referente ao processo biológico para 240 genes subexpressos e para 164 genes superexpressos. Os processos biológicos mais afetados pelo estresse foram: resposta a estímulos (60% reprimidos e 40% induzidos), regulação biológica (55% reprimidos e 45% induzidos), sinalização (65% reprimidos e 35% induzidos), processos do sistema imune (55% reprimidos e 45% induzidos). Os genes que se mostraram diferencialmente expressos nesta combinação, sendo que para 632 genes subexpressos e 336 genes superexpressos foram encontradas GOs, estão relacionados aos seguintes componentes celulares: organela (70% reprimidos e 30% induzidos), membrana (65% reprimidos e 35% induzidos), região extracelular (55% reprimidos e 45% induzidos). A busca da ontologia dos genes diferencialmente expressos para a função molecular desses genes indica 195 genes subexpressos e 127 genes superexpressos. Estes genes apresentam as seguintes funções moleculares: *binding* (60% reprimidos e 40% induzidos), atividade catalítica (65% reprimidos e 35% induzidos), atividade antioxidante (90% reprimidos e 10% induzidos), atividade de transporte (60% reprimidos e 40% induzidos), atividade de fatores de transcrição ligantes a ácido nucleico (40% reprimidos e 60% induzidos).

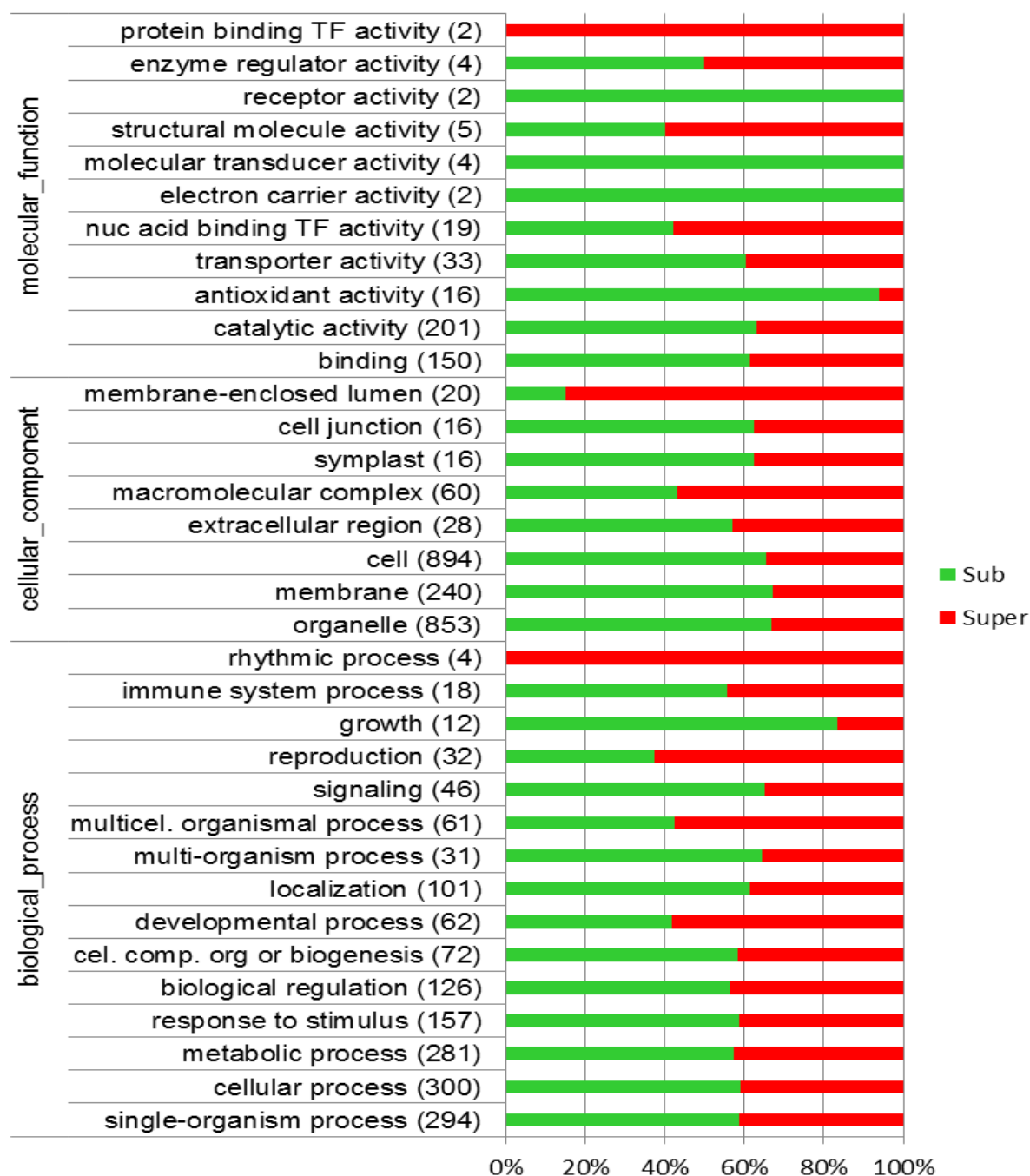


Figura 18. Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação BAC x BAT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação BAC x BAT.

A anotação da ontologia gênica para a combinação AT x BAT (Figura 21) mostrou, para a categoria processo biológico, 6 genes subexpressos e 62 genes superexpressos. Os principais processos biológicos dos quais estes genes fazem parte são: resposta a estímulos (15% reprimidos e 85% induzidos), processo de

desenvolvimento (100% induzidos), regulação biológica (100% induzidos), sinalização (100% induzidos). Em relação ao componente celular foram obtidos, com GO, 21 genes subexpressos e 211 genes superexpressos. Estes genes tem relação com os seguintes componentes celulares: organela (10% reprimidos e 90% induzidos), membrana (5% reprimidos e 95% induzidos), complexo macromolecular (15% reprimidos e 85% induzidos). Em relação à função molecular foram encontrados 4 genes subexpressos e 53 genes superexpressos que apresentam GO. Estes estão envolvidos principalmente com atividade catalítica (10% reprimidos e 90% induzidos), *binding* (10% reprimidos e 90% induzidos), atividade de transporte (100% induzidos), atividade de fatores de transcrição ligantes a ácido nucleico (100% induzidos).

Nas comparações apresentadas, entre 60% e 70% dos genes diferencialmente expressos se mostraram subexpressos quando comparados os genótipos nas condições controle versus a condição tratamento, ou seja, o nível de expressão na condição tratamento é menor do que na condição controle. Chawade et al. (2013) verificando o perfil global de expressão de genes induzidos por baixas temperaturas em uma cultivar tolerante ao frio durante curtos períodos de exposição ao frio (entre 0,5h e 24h) encontrou um maior número de genes subexpressos em comparação com os genes superexpressos, o que vem a corroborar com os resultados obtidos neste trabalho. Além disso, Yun et al. (2010) também encontraram um número maior de genes subexpressos em relação aos genes superexpressos após a exposição das plantas a períodos de frio (10°C) que variaram de 0.5 a 96h de exposição. Wang et al. (2011) trabalhando com resposta espaço-temporal do genoma do arroz a estresse por seca, em diversos estádios de desenvolvimento da cultura, também encontraram um número maior de genes subexpressos em relação aos genes superexpressos.

A combinação AT x BAT apresenta um padrão de comportamento distinto da maioria das demais combinações. Ela apresenta um número maior de genes superexpressos em relação aos genes subexpressos. Diversos TFs estão superexpressos nesta combinação, entre os quais podemos citar 6 ERFs: Os02g0676800, Os09g0457900 e Os09g0522100 (AP2), e Os09g0522000, Os06g0127100, e Os09g0522200 (DREB). Outros três genes são descritos como bHLH: Os04g0381700, Os01g0952800 e Os04g0301500. Dois TFs da família WRKY também se mostraram superexpressos: Os08g0198000 e Os05g0478700. Todas

estas famílias gênicas já foram descritas como responsivas a estresse por frio (Lata e Prasad, 2011; Mao e Chen, 2012; Chen et al., 2011), e podem ser muito importantes no aumento da tolerância ao frio nos *bulks* de RILs analisados neste experimento.

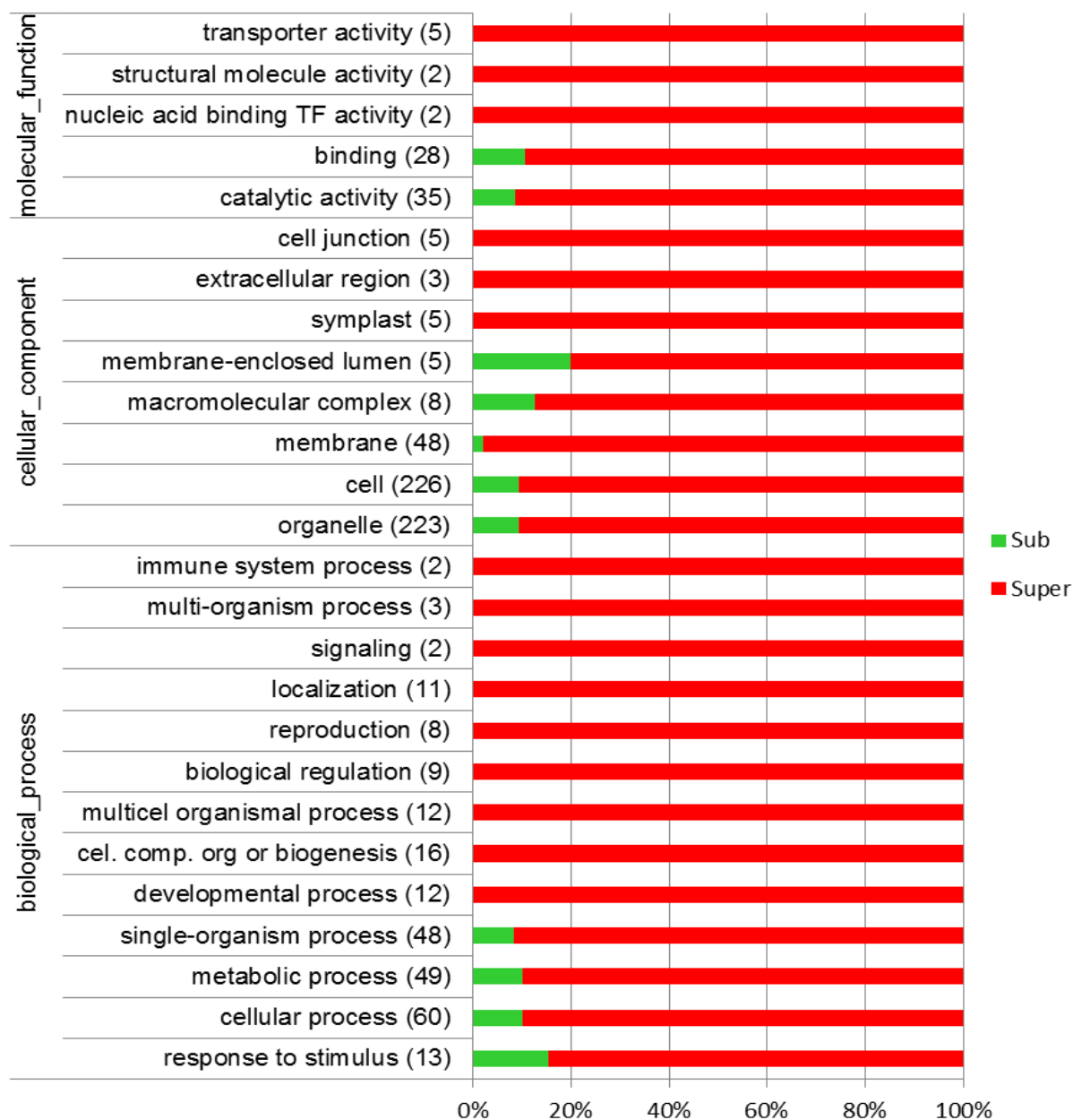


Figura 19. Anotação da Ontologia gênica para os genes diferencialmente expressos na combinação AT x BAT, onde são mostradas as três principais divisões: função molecular, componente celular e processo biológico. Além disso, são mostradas as GOs dentro de cada divisão principal e o somatório de genes subexpressos e superexpressos é mostrado ao lado de cada GO e as barras horizontais mostram a porcentagem relativa entre os genes sub e superexpressos. Ontologia gênica nível 2. Sub: genes subexpressos. Super: genes superexpressos na combinação AT x BAT.

5 Conclusões

As conclusões deste trabalho indicam que o número de genes diferencialmente expressos ($\log_2\text{-fold-change} \geq |1.5|$) varia grandemente entre as combinações, apresentando como limite superior 1481 genes subexpressos na combinação BNC x BAC e 1017 genes superexpressos em NT x AT, e 6 genes subexpressos na combinação NT x BNT e 5 genes superexpressos na combinação NC x BNC. Nas diversas combinações analisadas, os fatores de transcrição de diversas famílias mostram um padrão difuso de expressão.

6 Referências bibliográficas

AGARWAL, P.K.; JHA, B. Transcription factors in plants and ABA dependent and independent abiotic stress signaling. **Biologia Plantarum**, v. 54, p. 201-212, 2010.

ALBUQUERQUE, M.C.F.; CARVALHO, N.M. Effects of the type of environmental stress on the emergence of sunflower (*Helianthus annuus* L.), soybean (*Glycine max* (L.) Merrill) and maize (*Zea mays* L.) seeds with different levels of vigor. **Seed Science and Technology**, v. 31, p. 465-479, 2003.

ALVARADO, V.; BRADFORD, K.J. A hydrothermal time model explains the cardinal temperatures for seed germination. **Plant, Cell & Environment**, v. 25, p. 1061-1069, 2002.

AUER, P.L.; DOERGE, R.W. Statistical design and analysis of RNA sequencing data. **Genetics**, v. 185, p. 405-416, 2010.

BETHKE, P.C.; SCHUURINK, R.; JONES, R.L. Hormonal signaling in cereal aleurone. **Journal of Experimental Botany**, v. 48, p. 1337–1356. 1997. doi: 10.1093/jxb/48.7.1337

BEWLEY, J.D. Seed Germination and Dormancy. **Plant Cell**, v. 9, n. 7, p. 1055-1066, 1997.

BIERLEN, R.; WAILES, E.J.; CRAMER, G.L. The Mercosur rice economy. **Arkansas Agricultural Experiment Station Bulletin**, v. 954, p. 1–58, 1997.

BOSETTI, F.; MONTEBELLI, C.; NOVEMBRE, A.D.L.C.; CHAMMA, H.P.; PINHEIRO, J.B. Genetic variation of germination cold tolerance in Japanese rice germplasm. **Breeding Science**, v. 62, n. 3, p. 209–215, 2012. doi: 10.1270/jsbbs.62.209

BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVENS, T.M.; RAMDOSS, Y.; BUCKRIL, E.S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. **Bioinformatics**, v. 23, n. 19, p. 2633–2635, 2007.

BRAR, D.S.; SINGH, K. *Oryza*. in C. Kole, ed. Wild crop relatives: genomic breeding and resources: cereals. Springer-Verlag, Berlin, Heidelberg. 2011.

BROWN, T.A. Genomes. 2nd edition. Oxford: Wiley-Liss; 2002. <http://www.ncbi.nlm.nih.gov/books/NBK21121/>.

CABILI, M.N.; TRAPNELL, C.; GOFF, L.; KOZIOL, M.; TAZON-VEGA, B.; REGEV, A.; RINN, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. **Genes & Development**, v. 25, n. 18, p. 1915-1927, 2011.

CHANDRARATNA, M. F. Genetics and breeding of rice. London: Longmans, 1964, 389p.

CHANG, T.T. Rice. In: Simmonds, N. W. (Ed) Evolution of crop plants. New York: Longman, 1996. p. 147-155.

CHANG, T.T. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. **Euphytica**, v. 25, p. 435–441, 1976.

CHAWADE, A.; LINDLÖF, A.; OLSSON, B.; OLSSON, O. Global expression profiling of low temperature induced genes in the chilling tolerant *japonica* rice Jumli Marshi. **PloS One**, 8(12), 2013. DOI: 10.1371/journal.pone.0081729

CHEN, L.; SONG, Y.; LI, S.; ZHANG, L.; ZOU, C.; YU, D. The role of WRKY transcription factors in plant abiotic stresses. **Biochimica et Biophysica Acta**, v. 1819, n. 2, p. 120-8, 2011. doi:10.1016/j.bbagr.2011.09.002

CHEN, J.Q.; DONG, Y.; WANG, Y.J.; LIU, Q.; ZHANG, J.S.; CHEN, S.Y. An AP2/EREBP-type transcription-factor gene from rice is cold-inducible and encodes a nuclear-localized protein. **Theoretical and Applied Genetics**, v. 107, p. 972 - 979, 2003.

CHENG, C.; YUN, K.Y.; RESSOM, H.W.; MOHANTY, B.; BAJIC, V.B.; JIA, Y.; YUN, S.J.; DE LOS REYES, B.G. An early response regulatory cluster induced by low temperature and hydrogen peroxide in seedlings of chilling-tolerant *japonica* rice. **BMC Genomics**, v. 18, n. 8, p. 175, 2007.

CHINNUSAMY, V.; SCHUMAKER, K.; ZHU, J.K. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. **Journal of Experimental Botany**, v. 55, p. 225–236, 2004.

CLAYTON, S.; NEVES, P.C. Country snapshot: Brazil. **Rice Today**, v. 10, p. 16 –17, 2011.

COCK, P.J.; FIELDS, C.J.; GOTO, N.; HEUER, M.L.; RICE, P.M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. **Nucleic Acids Research**, v. 38, n. 6, p. 1767–1771, 2010. doi: 10.1093/nar/gkp1137

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de grãos, v. 1 - Safra 2013/14, n. 4 - Quarto Levantamento, Brasília, p. 1-67, jan. 2014 <http://www.conab.gov.br/>. Acessado em janeiro de 2014.

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; TALÓN, M.; ROBLES, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, p. 3674-3676, 2005.

COWLING, V.H. Regulation of mRNA cap methylation. **Biochemical Journal**, v. 425, n. 2, p. 295–302, 2010. doi: 10.1042/BJ20091352

da CRUZ, R.P.; SPEROTTO, R.A.; CARGNELUTTI, D.; ADAMSKI, J.M.; FREITASTERRA, T.; FETT, J.P. Avoiding damage and achieving cold tolerance in rice plants. **Food and Energy Security**, v. 2, n. 2, p. 96–119, 2013. doi:10.1002/fes3.25

da CRUZ, R.P.; MILACH, S.C.K.; FEDERIZZI, L.C. Inheritance of rice cold tolerance at the germination stage. **Genetics and Molecular Biology**, v. 29, p. 314 –320, 2006a.

da CRUZ, R.P.; MILACH, S.C.K.; FEDERIZZI, L.C. Rice cold tolerance at the reproductive stage in a controlled environment. **Scientia Agricola**, v. 63, p. 255–261, 2006b.

da CRUZ, R.P.; MILACH, S.C.K. Cold tolerance at the germination stage of rice: methods of evaluation and characterization of genotypes. **Scientia Agricola**, v. 61, p. 1–8, 2004.

da CRUZ, R.P.; MILACH, S.C.K. Breeding for cold tolerance in irrigated rice. **Ciencia Rural**, v. 30, p. 909–917, 2000.

DAI, X.; XU, Y.; MA, Q.; XU, W.; WANG, T.; XUE, Y.; CHONG, K. Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis. **Plant Physiology**, v. 143, n. 4, p. 1739-1751, 2007.

DILLIES, M.A.; RAU, A.; AUBERT, J.; HENNEQUET-ANTIER, C.; JEANMOUGIN, M.; SERVANT, N.; KEIME, C.; MAROT, G.; CASTEL, D.; ESTELLE, J.; GUERNEC, G.; JAGLA, B.; JOUNEAU, L.; LALOË, D.; LE GALL, C.; SCHAËFFER, B.; LE CROM, S.; GUEDJ, M.; JAFFRÉZIC, F.; FRENCH STATOMIQUE CONSORTIUM. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. **Briefings in Bioinformatics**, v. 14, n. 6, p. 671-683, 2013. doi: 10.1093/bib/bbs046.

DOEBLEY, J.F. The genetics of maize evolution. **Annual Review of Genetics**, v. 38, p. 37–59, 2004.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKRIL, E.S.; MITCHELL, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One**, v. 4, p. 6(5):e19379, 2011. doi: 10.1371/journal.pone.0019379.

EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária). BRS Atalanta: Alternativa para o uso racional de insumos nas lavouras de arroz irrigado do RS. 2008. Disponível em: <http://www.cpact.embrapa.br/publicacoes/download/folder/ATALANTA.pdf>. Acessado em outubro de 2013.

ENGSTRÖM, P.G.; STEIJGER, T.; SIPOS, B.; GRANT, G.R.; KAHLES, A. THE RGASP CONSORTIUM, RÄTSCH, G.; GOLDMAN, N.; HUBBARD, T.J.; HARROW, J.; GUIGÓ, R.; BERTONE, P. Systematic evaluation of spliced alignment programs for RNA-seq data. **Nature Methods**, v. 10, p. 1185–1191, 2013. doi:10.1038/nmeth.2722

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred.

II. Error Probabilities. **Genome Research**, v. 8, p. 186-194, 1998.

FANG Z, CUI X. Design and validation issues in RNA-seq experiments. **Brief Bioinformatics**, v. 12, n. 3, p. 280-287, 2011. doi: 10.1093/bib/bbr004.

FANG, J.; CHU, C. Absciscic acid and the pre-harvest sprouting in cereals. **Plant Signaling & Behavior**, v. 3, n. 12, p. 1046-1048, 2008.

FAOSTAT – FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS STATISTICS. <http://faostat.fao.org/>. Acessado em dezembro de 2013.

FEDER, M.E.; HOFMANN, G.E. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. **Annual Review of Physiology**, v. 61, p. 243-282, 1999.

FERRAGINA, P.; MANZINI, G. An experimental study of a compressed index. **Information Sciences**, v. 135, p. 13–28, 2001.

FOWRIL, S.; THOMASHOW, M.F. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. **Plant Cell**, v. 14, n. 8, p. 1675-90, 2002.

GAO, L.; INNAN, H. Nonindependent Domestication of the Two Rice Subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, Demonstrated by Multilocus Microsatellites. **Genetics**, v. 179; n. 2, p. 965–976, 2008. doi: 10.1534/genetics.106.068072

GE, X.; CHEN, W.; SONG, S.; WANG, W.; HU, S.; YU, J. Transcriptomic profiling of mature embryo from an elite super-hybrid rice LYP9 and its parental lines. **BMC Plant Biology**, 8():114, 2008.

GILMOUR, S.J.; SEBOLT, A.M.; SALAZAR, M.P.; EVERARD, J.D.; THOMASHOW, M.F. Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. **Plant Physiology**, v. 124, p. 1854–1865, 2000.

GOFF, L.; TRAPNELL, C.; KELLEY, D. CummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. 2012. R package version 2.2.0.

GOFF, S.A.; RICKE, D.; LAN, T.H.; PRESTING, G.; WANG, R.; DUNN, M.; GLAZEBROOK, J.; SESSIONS, A.; OELRIL, P.; VARMA, H.; HADLEY, D.; HUTCHISON, D.; MARTIN, C.; KATAGIRI, F.; et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). **Science**, v. 296, n. 5565, p. 92-100, 2002.

HANNAH, M.A.; HEYER, A.G.; HINCHA, D.K. A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. **PLoS Genetics**, 1:e26, 2005.

HE, D.; YANG, P. Proteomics of rice seed germination. **Frontiers in Plant Science**, v. 4, p. 246, 2013. doi: 10.3389/fpls.2013.00246

HOLDSWORTH, M.J.; FINCH-SAVAGE, W.E.; GRAPPIN, P.; JOB, D. Review Post-genomics dissection of seed dormancy and germination. **Trends in Plant Science**, v. 13, n. 1, p. 7-13, 2008.

HOWELL, K.A.; NARSAI, R.; CARROLL, A.; IVANOVA, A.; LOHSE, M.; USADEL, B.; MILLAR, A.H.; WHELAN, J. Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. **Plant Physiology**, v. 149, n. 2, p. 961-980, 2009.

HUANG, X.S.; WANG, W.; ZHANG, Q.; LIU, J.H. A basic helix-loop-helix transcription factor, PtrbHLH, of *Poncirus trifoliata* confers cold tolerance and modulates peroxidase-mediated scavenging of hydrogen peroxide. **Plant Physiology**, v. 162, n. 2, p. 1178-94, 2013. doi: 10.1104/pp.112.210740

IBA, K. Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. **Annual Review of Plant Biology**, v. 53, p. 225 –245, 2002.

IRGSP (International Rice Genome Sequencing Project). The map-based sequence of the rice genome. **Nature**, v. 436, p. 793-800, 2005.

IRRI, AfricaRice & CIAT. Global Rice Science Partnership (GRiSP). November 2010.

JACOBSEN, J.V.; GUBRIL, F.; CHANDRIL, P.M. Gibberellin and abscisic acid in germinating cereals, in *Plant Hormones: Physiology, Biochemistry and Molecular Biology*, ed Davies P. J., editor. (Boston, MA: Kluwer), p. 246–271, 1995.

JENNINGS, P.R.; COFFMAN, W.R.; KAUFFMAN, H.E. Rice improvement. International Rice Research Institute, Los Banos. p. 83 – 186, 1979.

JIN, J.P.; ZHANG, H.; KONG, L.; GAO, G.; LUO, J.C. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. **Nucleic Acids Research**, v. 42, p. D1182-D1187, 2014.

JULIANO, B.O. Rice in human nutrition. International Rice Research Institute and Food and Agriculture Organization of the United Nations, Los Banos, p. 1 – 15, 1993.

KIM, D.; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S.L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, v. 14:R36, 2013. doi:10.1186/gb-2013-14-4-r36

KIM, D.; SALZBERG, S.L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. **Genome Biology**, 12:R72, 2011.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, p. 357-359, 2012.

LANGMEAD, B.; TRAPNELL, C.; POP, M.; SALZBERG, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v. 10:R25, 2009. doi:10.1186/gb-2009-10-3-r25

LATA, C.; PRASAD, M. Role of DREBs in regulation of abiotic stress responses in plants. **Journal of Experimental Botany**, v. 62, n. 14, p. 4731-4748, 2011. doi:10.1093/jxb/err210

LEE, B.H.; HENDERSON, D.A.; ZHU, J.K. The Arabidopsis cold-responsive transcriptome and its regulation by ICE1. **The Plant Cell**, v.17, p.3155–3175, 2005.

LEVY, S.; SUTTON, G.; NG, P.C.; FEUK, L.; HALPERN, A.L.; WALENZ, B.P.; AXELROD, N.; HUANG, J.; KIRKNESS, E.F.; DENISOV, G.; LIN, Y.; MACDONALD, J.R.; PANG, A.W.; et al. The diploid genome sequence of an individual human. **PLoS Biology**, 5:e254, 2007. DOI: 10.1371/journal.pbio.0050254

LI, C. B.; ZHOU, A. L.; SANG, T. Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. **New Phytologist**, v. 170, p. 185–194, 2006.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R.; 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. The Sequence alignment/map (SAM) format and SAMtools. **Bioinformatics**, v. 25, p. 2078-2079, 2009.

LI, H.; RUAN, J.; DURBIN, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. **Genome Research**, v. 18, n. 11, p. 1851-1858, 2008. doi: 10.1101/gr.078212.108.

LI, H-Y.; CHANG, C-S.; LU, L-S.; LIU, C-A.; CHAN, M-T.; CHARNG, Y-Y. Over-expression of *Arabidopsis thaliana* heat shock factor gene (AtHsfA1b) enhances chilling tolerance in transgenic tomato. **Botanical Bulletin of Academia Sinica**, v. 44, p. 129-140, 2003.

LI, T.G.; VISPERAS, R.M.; VERGARA, B.S. Correlation of cold tolerance at different growth stages in rice. **Acta Botanica Sinica**, v. 23, p. 203 – 207, 1981.

LI, W.J.; ZHANG, B.; HUANG, G.W.; KANG, G.P.; LIANG, M.Z.; CHEN, L.B. Chloroplast DNA polymorphism and evolutionary relationships between Asian cultivated rice (*Oryza sativa*) and its wild relatives (*O. rufipogon*). **Genetics and Molecular Research**, v. 11, p. 4418-4431, 2012.

LIMA, M.G.S.; LOPES, N.F.; ZIMMER, P.D.; MENEGHELLO, G.E.; MENDES, C.R.; AMARANTE, L. Enzyme expression in *indica* and *japonica* rice cultivars under saline stress. **Acta Scientiarum. Biological Sciences**, v. 34, p. 473 –481, 2012.

LIN, J. Divergence measures based on the Shannon entropy. **IEEE Transactions on Information Theory**, v. 37, n. 1, p. 145–151, 1991. doi:10.1109/18.61115.

LIU, H.; OUYANG, B.; ZHANG, J.; WANG, T.; LI, H. Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. **PLoS ONE**, v. 7, p. e50785, 2012.

LIU, L.; LI, Y.; LI, S.; HU, N.; HE, Y.; PONG, R.; LIN, D.; LU, L.; LAW, M. Comparison of Next-Generation Sequencing Systems. **Journal of Biomedicine and Biotechnology**, v. 2012, p. 1-11, 2012. doi: <http://dx.doi.org/10.1155/2012/251364>

LIU, Y.; ZHOU, J.; WHITE, K.P. RNA-seq differential expression studies: more sequence or more replication? **Bioinformatics**, v. 30, n. 3, p. 301-304, 2014. doi: 10.1093/bioinformatics/btt688

LOHSE, M.; BOLGER, A.M.; NAGEL, A.; FERNIE, A.R.; LUNN, J.E.; STITT, M.; USADEL, B. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. **Nucleic Acids Research**, v. 40, p. W622-7, 2012.

LOS, D.A.; MURATA, N. Structure and expression of fatty acid desaturases. **Biochimica et Biophysica Acta**, v. 1394, p. 3–15, 1998.

LU, B.R.; ZHENG, K.L.; QIAN, H.R.; ZHUANG, J.Y. Genetic differentiation of wild relatives of rice as referred by the RFLP analysis. **Theoretical and Applied Genetics**, v. 106, p. 101–106, 2002.

LUTZ, C.S.; MOREIRA, A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. **Wiley Interdisciplinary Reviews: RNA**, v. 2, n. 1, p. 23–31, 2011. doi: 10.1002/wrna.47

MACKILL, D.J.; LEI, X. Genetic variation for traits related to temperate adaptation of rice cultivars. **Crop Science**, v. 37, p. 1340 – 1346, 1997.

MAHAJAN, S.; TUTEJA, N. Cold, salinity and drought stresses: An overview. **Archives of Biochemistry and Biophysics**, v. 444, p. 139–158, 2005.

MAO, D.; CHEN, C. Colinearity and Similar Expression Pattern of Rice DREB1s Reveal Their Functional Conservation in the Cold-Responsive Pathway. **PLoS ONE**, v. 7, n. 10, e47275, 2012. doi:10.1371/journal.pone.0047275

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends in Genetics**, v. 24, n. 3, p. 133–141, 2008.

MARIONI, J.C.; MASON, C.E.; MANE, S.M.; STEPHENS, M.; GILAD, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. **Genome Research**, v. 18, p. 1509-1517, 2008.

MERTZ, L.M.; HENNING, F.A.; SOARES, R.C.; BALDIGA, R.F.; PESKE, F.B.; MORAES, D.M. Physiological changes in rice seeds exposed to cold in the germination phase. **Revista Brasileira de Sementes**, v. 31, p. 254 – 262, 2009.

MITTRIL, R.; VANDERAUWERA, S.; GOLRILY, M.; VAN BREUSEGEM, F. Reactive oxygen gene network of plants. **Trends of Plant Science**, v. 9, n. 10, p. 490-498,

2004.

MITTAL, D.; MADHYASTHA, D.A.; GROVER, A. Genome-Wide Transcriptional Profiles during Temperature and Oxidative Stress Reveal Coordinated Expression Patterns and Overlapping Regulons in Rice. **PLoS ONE**, v. 7, n. 7, e40899, 2012. doi:10.1371/journal.pone.0040899

MIURA, K.; LIN, S.Y.; YANO, M.; NAGAMINE, T. Mapping quantitative trait loci controlling low-temperature germinability in rice (*Oryza sativa* L.). **Breeding Science**, v. 51, p. 293–299, 2001.

MOHANTY, S.; WASSMANN, R.; NELSON, A.; MOYA, P.; JAGADISH, S.V.K. Rice and climate change: significance for food security and vulnerability. IRRI Discussion Paper Series No. 49. Los Baños (Philippines): International Rice Research Institute. 14 p. 2013.

MORTAZAVI, A.; WILLIAMS, B.A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, v. 5, n. 7, p. 621-628, 2008.

MURATA, N.; LOS, D.A. Membrane fluidity and temperature perception. **Plant Physiology**, v. 115, p. 875–879, 1997.

NAGALAKSHMI, U.; WAERN, K.; SNYDER, M. RNA-Seq: a method for comprehensive transcriptome analysis. **Current Protocols in Molecular Biology**, Chapter 4(Unit 4.11), p. 1 – 13, 2010.

NAGALAKSHMI, U.; WANG, Z.; WAERN, K.; SHOU, C.; RAHA, D.; GERSTEIN, M.; SNYDER, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. **Science**, v. 320, n. 5881, p. 1344-1349, 2008.

NAKANO, T.; SUZUKI, K.; FUJIMURA, T.; SHINSHI, H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. **Plant Physiology**, v. 140, p. 411–432, 2006.

NAKASHIMA, K.; TRAN, L.S.; VAN NGUYEN, D.; FUJITA, M.; MARUYAMA, K.; TODAKA, D.; ITO, Y.; HAYASHI, N.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. **The Plant Journal**, v. 51, n. 4, p. 617-630, 2007.

NOOKAEW, I.; PAPINI, M.; PORNPUTTAPONG, N.; SCALCINATI, G.; FAGERBERG, L.; UHLÉN, M.; NIELSEN, J. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. **Nucleic Acids Research**, v. 40, n. 20, p. 10084-97, 2012. doi: 10.1093/nar/gks804.

ORVAR, B.L.; SANGWAN, V.; OMANN, F.; DHINDSA, R.S. Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. **The Plant Journal**, v. 23, p. 785–794, 2000.

PENFIELD, S.; JOSSE, E.M.; KANNANGARA, R.; GILDAY, A.D.; HALLIDAY, K.J.; GRAHAM, I.A. Cold and light control seed germination through the bHLH transcription factor SPATULA. **Current Biology**, v. 15, n. 22, p. 1998-2006, 2005.

PRAKASH, M.; NAIR, G.; KANG, I.S.; MOON, B.Y.; LEE, C.H. Effects of low temperature stress on rice (*Oryza sativa* L.) plastid ω -3 desaturase gene, OsFAD8 and its functional analysis using T-DNA mutants. **Plant Cell, Tissue and Organ Culture**, v. 98, n. 1, p. 87-96, 2009.

QUINN, E.M.; CORMICAN, P.; KENNY, E.M.; HILL, M.; ANNEY, R.; GILL, M.; CORVIN, A.P.; MORRIS, D.W. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. **PLoS One**, v. 8, n. 3, p. e58815, 2013. doi: 10.1371/journal.pone.0058815.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RAPAPORT, F.; KHANIN, R.; LIANG, Y.; PIRUN M.; KREK, A.; ZUMBO, P.; MASON, C.E.; SOCCI, N.D.; BETEL, D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. **Genome Biology**, v. 14, p. R95 2013. doi:10.1186/gb-2013-14-9-r95

ROUTABOUL, J.M.; FISCHER, S.F.; BROWSE, J. Trienoic fatty acids are required to maintain chloroplast function at low temperatures. **Plant Physiology**, v.124, p. 1697–1705, 2000.

SAITO, K.; SAITO, Y.H.; FUNATSUKI, W.M.; SATO, Y.; KATO, A. Physical mapping and putative candidate gene identification of a quantitative trait locus Ctb1 for cold tolerance at the booting stage of rice. **Theoretical and Applied Genetics**, v. 109, p. 515 –522, 2004.

SAKAI, H.; LEE, S.S.; TANAKA, T.; NUMA, H.; KIM, J.; KAWAHARA, Y.; WAKIMOTO, H.; YANG, C.C.; IWAMOTO, M.; ABE, T.; YAMADA, Y.; MUTO, A.; INOKUCHI, H.; IKEMURA, T.; MATSUMOTO, T.; SASAKI, T.; ITOH, T. Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. **Plant & Cell Physiology**, v. 54, n. 2, p. E6, 2013.

SALTVEIT, M.E.; MORRIS, L.L. Overview of chilling injury of horticultural crops, p. 1-14. In: C.Y. Wang (ed.) Chilling injury of horticultural crops. CRC Press, Boca Raton, Fla, 1990.

SANGHERA, G.S.; WANI, S.H.; HUSSAIN, W.; SINGH, N.B. Engineering Cold Stress Tolerance in Crop Plants. **Current Genomics**, v. 12, n. 1, p. 30–43, 2011. doi: 10.2174/138920211794520178

SANGHERA, G.S.; WANI, S.H.; HUSSAIN, W.; SINGH, N.B. Engineering cold stress tolerance in crop plants. **Current Genomics**, v. 12, p. 30 –43, 2011.

SATO, Y.; MURAKAMI, T.; FUNATSUKI, H.; MATSUBA, S.; SARUYAMA, H.; TANIDA, M. Heat shock-mediated APX gene expression and protection against chilling injury in rice seedlings. **Journal of Experimental Botany**, v. 52, p. 145-151, 2001.

SECK, P. A.; DIAGNE, A.; MOHANTY, S.; WOPEREIS, M.C.S. Crops that feed the world 7: Rice. **Food Security**, v. 4, p. 7–24, 2012. DOI 10.1007/s12571-012-0168-1

SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of RNA-seq data. **BMC Bioinformatics**, 14:91, 2013. doi:10.1186/1471-2105-14-91

STEPONKUS, P.L. Role of the plasma membrane in freezing injury and cold acclimation. **Annual Review of Plant Physiology**, v. 35, p. 543–584, 1984.

STHAPIT, B.R.; WITCOMBE, J.R. Inheritance of tolerance to chilling stress in rice during germination and plumule greening. **Crop Science**, v. 38, p. 660 – 665, 1998.

STRAUSS, G.; HAUSER, H. Stabilization of lipid bilayer vesicles by sucrose during freezing. **Proceedings of the National Academy of Sciences**, v. 83, p. 2422–2426, 1986.

SU, C.F.; WANG, Y.C.; HSIEH, T.H.; LU, C.A.; TSENG, T.H.; YU, S.M. A novel MYBS3-dependent pathway confers cold tolerance in rice. **Plant Physiology**, v. 153, n. 1, p. 145-58, 2010. doi: 10.1104/pp.110.153015.

SUN, Z.; ASMANN, Y.W.; NAIR, A.; ZHANG, Y.; WANG, L.; KALARI, K.R.; BHAGWATE, A.V.; BAKER, T.R.; CARR, J.M.; KOCHER, J.P.; PEREZ, E.A.; THOMPSON, E.A. Impact of Library Preparation on Downstream Analysis and Interpretation of RNA-Seq Data: Comparison between Illumina PolyA and NuGEN Ovation Protocol. **PLoS One**, v. 8, n. 8, e71745, 2013. DOI: 10.1371/journal.pone.0071745

SUNKAR, R.; CHINNUSAMY, V.; ZHU, J.; ZHU, J.K. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. **Trends in Plant Science**, v. 12, p. 301–309, 2007.

SWEENEY, M.; MCCOUCH, S. The Complex History of the Domestication of Rice. **Annals of Botany**, v. 100, n. 5, p. 951-957, 2007.

TAKAHASHI, N. Differentiation of ecotypes in *Oryza sativa* L. Pp. 31 –67 in N. Takahashi and S. Tsunoda, eds. Biology of rice. Japan Science Society, Tokyo, Japan. 1984.

TARIQ, M.A.; KIM, H.J.; JEJELOWO, O.; POURMAND, N. Whole-transcriptome RNAseq analysis from minute amount of total RNA. **Nucleic Acids Research**, v. 39, n. 18, e120, 2011.

TILEBENI, H.G.; YOUSEFPOUR, H.; FARHADI, R.; GOLPAYEGANI, A. Germination Behavior of Rice (*Oryza Sativa* L.) Cultivars Seeds to Difference Temperatures.

Advances in Environmental Biology, v. 6, n. 2, p. 573-577, 2012.

TRAN, L.S.; NAKASHIMA, K.; SAKUMA, Y.; SIMPSON, S.D.; FUJITA, Y.; MARUYAMA, K.; FUJITA, M.; SEKI, M.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. **Plant Cell**, v. 16, p. 2481–2498, 2004.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; PERTEA, G.; KIM, D.; KELLEY, D.R.; PIMENTEL, H.; SALZBERG, S.L.; RINN, J.L.; PACHTER, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 17, n. 3, p. 562-78, 2012. doi: 10.1038/nprot.2012.016.

TRAPNELL, C.; WILLIAMS, B.A.; PERTEA, G.; MORTAZAVI, A.M.; KWAN, G.; VAN BAREN, M.J.; SALZBERG, S.L.; WOLD, B.; PACHTER, L. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. **Nature Biotechnology**, v. 28, n. 5, p. 511–515, 2010. doi:10.1038/nbt.1621

TRAPNELL, C.; PACHTER, L.; SALZBERG, S.L. TopHat: discovering splice junctions with RNA-seq. **Bioinformatics**, v.25, p.1105–1111, 2009.

UDVARDI, M.K.; KAKAR, K.; WANDREY, M.; MONTANARI, O.; MURRAY, J.; ANDRIANKAJA, A.; ZHANG, J.Y.; BENEDITO, V.; HOFER, J.M.I.; CHUENG, F. et al Legume transcription factors: global regulators of plant development and response to the environment. **Plant Physiology**, v. 144, p. 538–549, 2007.

UNCTAD - UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT. Information on rice. 2010. <http://www.unctad.org/infocomm/anglais/rice/market.htm#prod>. Acessado em novembro de 2013.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. 2006. The PLANTS Database (<http://plants.usda.gov>, 17 October 2006). National Plant Data Center, Baton Rouge, LA 70874-4490 USA.

VAN VERK, M.C.; HICKMAN, R.; PIETERSE, C.M.J.; VAN WEES, S.C.M. RNA-Seq: revelation of the messengers. **Trends in Plant Science**, v. 18, n. 4, p. 175-179, 2013.

VANNINI, C.; LOCATELLI, F.; BRACALE, M.; MAGNANI, E.; MARSONI, M.; OSNATO, M.; MATTANA, M.; BALDONI, E.; CORAGGIO, I. Overexpression of the rice Osmyb4 gene increases chilling and freezing tolerance of *Arabidopsis thaliana* plants. **The Plant Journal**, v. 37, n. 1, p. 115-27, 2004.

VAUGHAN, D.A.; MORISHIMA, H.; KADOWAKI, K. Diversity in the *Oryza* genus. **Current Opinion in Plant Molecular Biology**, v. 6, p. 139-146, 2003.

WAAGNER, D.; HOLMSTRUP, M.; BAYLEY, M.; SØRENSEN, J.G. Induced cold-tolerance mechanisms depend on duration of acclimation in the chill-sensitive

Folsomia candida (Collembola). **The Journal of Experimental Biology**, v. 216, p. 1991-2000, 2013. doi:10.1242/jeb.079814

WANG, D.; PAN, Y.; ZHAO, X.; ZHU, L.; FU, B.; LI, B.Z. Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. **BMC Genomics**, 12:149, 2011. doi:10.1186/1471-2164-12-149

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 2009. doi:10.1038/nrg2484

WASSMANN, R.; JAGADISH, S. V. K.; HEUER, S.; ISMAIL, A.; REDONA, E.; SERRAJ, R.; SINGH, G. HOWELL, R. K.; PATHAK, H.; SUMFLETH, K. Climate Change Affecting Rice Production: The Physiological and Agronomic Basis for Possible Adaptation Strategies. In Donald L. Sparks, editor, *Advances in Agronomy*, Vol 101. Burlington: Academic Press, 2009, pp.59-122. ISBN: 978-0-12-374817-1

WATANABE, Y. Genomic constitution of genus *Oryza*. In: Matsuo, T. Et al., (Ed.) *Science of the Rice Plant: Genetics*. Tokyo: Food and Agricultural Police Researcher Center, v. 3, p. 29-39, 1997.

WESOLOWSKI, S.; BIRTWISTLE, M.R.; REMPALA, G.A. A Comparison of Methods for RNA-Seq Differential Expression Analysis and a New Empirical Bayes Approach. **Biosensors**, v. 3, n. 3, p. 238-258, 2013. doi:10.3390/bios3030238

XING, J.; ZHANG, Y.; HAN, K.; SALEM, A.H.; SEN, S.K.; HUFF, C.D.; ZHOU, Q.; KIRKNESS, E.F.; LEVY, S.; BATZER, M.A.; JORDE, L.B. Mobile elements create structural variation: analysis of a complete human genome. **Genome Research**, v. 19, p. 1516-1526, 2009.

XU, H.; GAO, Y.; WANG, J. Transcriptomic analysis of rice (*Oryza sativa*) developing embryos using the RNA-Seq technique. **PLoS ONE**, v. 7, n. 2, Article ID e30646, 2012.

YAMANAKA, S.; NAKAMURA, I.; WATANABE, K.N.; SATO, Y.I. Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. **Theoretical and Applied Genetics**, v. 108, p. 1200–1204, 2004.

YE, H.; DU, H.; TANG, N.; LI, X.; XIONG, L. Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. **Plant Molecular Biology**, v. 71, p. 291 – 305, 2009.

YOSHIDA, S. Fundamentals of rice crop science. International Rice Research Institute, Los Banos, p. 1 –63, 1981.

YUN, K.Y.; PARK, M.R.; MOHANTY, B.; HERATH, V.; XU, F.Y.; MAULEON, R.; WIJAYA, E.; BAJIC, V.B.; BRUSKIEWICH, R.; DE LOS REYES, B.G. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of *japonica* rice to chilling stress. **Bmc Plant Biology**, 2010.

doi:10.1186/1471-2229-10-16

YUSUF, D.; BUTLAND, S.L.; SWANSON, M.I.; BOLOTIN, E.; TICOLL, A. et al. The Transcription Factor Encyclopedia. **Genome Biology**, v. 13, n. 3, p. R24, 2012. doi: 10.1186/gb-2012-13-3-r24

ZARET, K.S.; CARROLL, J.S. Pioneer transcription factors: establishing competence for gene expression. **Genes & Development**, v. 25, p. 2227-2241, 2011. doi: 10.1101/gad.176826.111

ZHANG, S.; ZHENG, J.; LIU, B.; PENG, S.; LEUNG, H.; ZHAO, J.; WANG, X.; YANG, T.; HUANG, Z. Identification of QTLs for cold tolerance at seedling stage in rice (*Oryza sativa* L.) using two distinct methods of cold treatment. **Euphytica**, v. 195, p. 95–104, 2014. DOI 10.1007/s10681-013-0977-0

ZHANG, F.; HUANG, L.; WANG, W.; ZHAO, X.; ZHU, L.; FU, B.; LI, Z. Genome-wide gene expression profiling of introgressed *indica* rice alleles associated with seedling cold tolerance improvement in a *japonica* rice background. **BMC Genomics**, 13:461, 2012. doi:10.1186/1471-2164-13-461

ZHOU, Q.Y.; TIAN, A.G.; ZOU, H.F.; XIE, Z.M.; LEI, G.; HUANG, J.; WANG, C.M.; WANG, H.W.; ZHANG, J.S.; CHEN, S.Y. Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. **Plant Biotechnology Journal**, v. 6, p. 486–503, 2008.

APÊNDICES

Apêndice A: Lista dos fatores de transcrição superexpressos nas 12 combinações analisadas.

FATORES DE TRANSCRIÇÃO SUPEREXPRESSOS			
AC x AT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
NAC	LOC_Os03g02800.1	Os03g0119966	no apical meristem protein
WOX	LOC_Os04g56780.1	Os04g0663600	homeobox domain containing protein
TALE	LOC_Os11g06020.1	Os11g0158832	homeobox domain containing protein
TCP	LOC_Os04g44440.1	Os04g0526000	TCP family transcription factor
MYB	LOC_Os09g26170.1	Os09g0431300	MYB family transcription factor
GATA	LOC_Os12g42970.1	Os12g0624900	GATA zinc finger domain containing protein
MYB	LOC_Os01g12860.1	Os01g0229000	MYB family transcription factor
G2-like	LOC_Os01g74020.1	Os01g0971800	MYB family transcription factor
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
C3H	LOC_Os06g46400.1	Os06g0677700	ATCPSF30/CPSF30
Dof	LOC_Os10g35300.1	Os10g0496000	dof zinc finger domain containing protein
C3H	LOC_Os06g49080.1	Os06g0704300	zinc finger C-x8-C-x5-C-x3-H type family protein
G2-like	LOC_Os03g55590.1	Os03g0764600	MYB family transcription factor
MYB_related	LOC_Os08g06110.2	Os08g0157600	MYB family transcription factor
WRKY	LOC_Os04g50920.1	Os04g0597300	WRKY37
FAR1	LOC_Os01g63380.1	Os01g0852650	transposon protein
ARF	LOC_Os08g40900.1	Os08g0520550	auxin response factor
ARF	LOC_Os07g08600.1	Os07g0183700	auxin response factor
HB-other	LOC_Os08g32990.1	Os08g0426000	DNA-directed RNA polymerase subunit
NAC	LOC_Os02g36880.1	Os02g0579000	No apical meristem protein
WRKY	LOC_Os01g43650.1	Os01g0626400	WRKY11
EIL	LOC_Os04g38400.1	Os04g0456900	ethylene-insensitive 3
bHLH	LOC_Os09g33580.1	Os09g0510500	BEE 1
FAR1	LOC_Os12g12380.1	Os12g0225300	transposon protein
Dof	LOC_Os03g16850.1	Os03g0276300	dof zinc finger domain containing protein

AC x BAC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
WRKY	LOC_Os02g43560.1	Os02g0652100	WRKY34
SBP	LOC_Os08g40260.1	Os08g0513700	OsSPL15 - SBP-box gene family member
ERF	LOC_Os03g60120.1	Os03g0815800	AP2 domain containing protein
WRKY	LOC_Os03g55080.1	Os03g0758050	WRKY3
C3H	LOC_Os01g45730.1	Os01g0645000	zinc finger C-x8-C-x5-C-x3-H type family protein
FAR1	LOC_Os07g42400.1	Os07g0615900	transposon protein
bZIP	LOC_Os01g58760.1	Os01g0801901	bZIP transcription factor domain containing protein
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
C2H2	LOC_Os06g07020.1	Os06g0166200	ZOS6-01 - C2H2 zinc finger protein
ZF-HD	LOC_Os11g03420.1	Os11g0128300	ZF-HD protein dimerisation region containing protein
HB-other	LOC_Os02g05450.1	Os02g0147800	Homeobox domain containing protein
MIKC	LOC_Os06g06750.1	Os06g0162800	OsMADS5 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
C2H2	LOC_Os04g08060.1	Os04g0162500	ZOS4-03 - C2H2 zinc finger protein
FAR1	LOC_Os11g12490.1	Os11g0231800	transposon protein
TALE	LOC_Os05g03884.1	Os05g0129700	homeobox protein knotted-1-like 6
CAMTA	LOC_Os04g31900.1	Os04g0388500	calmodulin-binding transcription activator 4

AT x BAT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
B3	LOC_Os03g42370.1	Os03g0621650	B3 DNA binding domain containing protein
bHLH	LOC_Os04g31290.1	Os04g0381700	helix-loop-helix DNA-binding domain containing protein
ERF	LOC_Os06g03670.1	Os06g0127100	dehydration-responsive element-binding protein
bHLH	LOC_Os01g72370.1	Os01g0952800	helix-loop-helix DNA-binding domain containing protein
NAC	LOC_Os04g52810.1	Os04g0619000	no apical meristem protein
ERF	LOC_Os02g45420.1	Os02g0676800	AP2 domain containing protein
WRKY	LOC_Os08g09800.1	Os08g0198000	WRKY105
WRKY	LOC_Os05g40070.1	Os05g0478700	WRKY84
bHLH	LOC_Os04g23550.1	Os04g0301500	basic helix-loop-helix family protein
HD-ZIP	LOC_Os01g57890.1	Os01g0788800	Homeobox domain containing protein
ERF	LOC_Os09g28440.1	Os09g0457900	AP2 domain containing protein
C2H2	LOC_Os12g38960.1	Os12g0578800	ZOS12-07 - C2H2 zinc finger protein
ERF	LOC_Os09g35030.1	Os09g0522200	dehydration-responsive element-binding protein
ERF	LOC_Os09g35010.1	Os09g0522000	dehydration-responsive element-binding protein
ERF	LOC_Os09g35020.1	Os09g0522100	AP2 domain containing protein

BAC x BAT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
NAC	LOC_Os01g15640.1	Os01g0261200	no apical meristem protein
CO-like	LOC_Os09g33550.1	Os09g0509700	CCT/B-box zinc finger protein
C2H2	LOC_Os01g68160.1	Os01g0909500	ZOS1-22 - C2H2 zinc finger protein
bHLH	LOC_Os11g38870.1	Os11g0601700	helix-loop-helix DNA-binding domain containing protein
CO-like	LOC_Os06g19444.1	Os06g0298200	CCT/B-box zinc finger protein
CO-like	LOC_Os06g44450.1	Os06g0654900	CCT/B-box zinc finger protein
B3	LOC_Os08g23570.1	Os08g0325134	B3 DNA binding domain containing protein
NAC	LOC_Os10g42130.1	Os10g0571600	no apical meristem protein
WRKY	LOC_Os09g25060.1	Os09g0417600	WRKY76
MYB	LOC_Os01g19330.1	Os01g0298400	MYB family transcription factor
NAC	LOC_Os01g66120.1	Os01g0884300	No apical meristem protein
bHLH	LOC_Os06g08500.1	Os06g0184000	BHLH transcription factor
B3	LOC_Os10g39190.1	Os10g0537100	B3 DNA binding domain containing protein
NAC	LOC_Os11g03370.1	Os11g0127600	no apical meristem protein
HSF	LOC_Os02g32590.1	Os02g0527300	HSF-type DNA-binding domain containing protein
WRKY	LOC_Os01g40430.1	Os01g0586800	WRKY27
NAC	LOC_Os05g10620.1	Os05g0194500	no apical meristem protein
NAC	LOC_Os04g38720.1	Os04g0460600	no apical meristem protein
MYB	LOC_Os04g42950.1	Os04g0508500	MYB family transcription factor
bZIP	LOC_Os03g59460.1	Os03g0809200	transcription factor
NAC	LOC_Os11g05614.1	Os11g0154500	no apical meristem protein
M-type	LOC_Os01g69850.1	Os01g0922800	OsMADS65 - MADS-box family gene with MIKC* type-box
bZIP	LOC_Os01g46970.1	Os01g0658900	transcription factor
bZIP	LOC_Os11g05640.1	Os11g0154900	bZIP transcription factor domain containing protein
HSF	LOC_Os01g43590.1	Os01g0625300	HSF-type DNA-binding domain containing protein
ERF	LOC_Os05g41760.1	Os05g0497200	AP2 domain containing protein
WRKY	LOC_Os03g21710.1	Os03g0335200	WRKY79
C2H2	LOC_Os12g38940.1	Os12g0578600	ZOS12-06 - C2H2 zinc finger protein
WRKY	LOC_Os05g25770.1	Os05g0322900	WRKY45
HD-ZIP	LOC_Os06g48290.1	Os06g0698200	homeobox associated leucine zipper
WRKY	LOC_Os01g60490.1	Os01g0820400	WRKY22
NAC	LOC_Os03g04070.1	Os03g0133000	no apical meristem protein
HSF	LOC_Os02g13800.1	Os02g0232000	HSF-type DNA-binding domain containing protein

HSF	LOC_Os01g39020.1	Os01g0571300	HSF-type DNA-binding domain containing protein
HD-ZIP	LOC_Os02g43330.1	Os02g0649300	homeobox associated leucine zipper
C2H2	LOC_Os03g60560.1	Os03g0820300	ZOS3-21 - C2H2 zinc finger protein
B3	LOC_Os01g68370.1	Os01g0911700	B3 DNA binding domain containing protein
ERF	LOC_Os09g13940.1	Os09g0309700	AP2 domain containing protein
HSF	LOC_Os06g35960.1	Os06g0553100	HSF-type DNA-binding domain containing protein
NAC	LOC_Os12g05990.1	Os12g0156100	No apical meristem protein
NAC	LOC_Os12g03050.1	Os12g0123800	no apical meristem protein
MYB_related	LOC_Os04g49450.1	Os04g0583900	MYB family transcription factor
ERF	LOC_Os07g47790.1	Os07g0674800	AP2 domain containing protein
NAC	LOC_Os02g38130.1	Os02g0594800	no apical meristem protein
NF-YB	LOC_Os02g49410.1	Os02g0725900	histone-like transcription factor and archaeal histone
ERF	LOC_Os09g28440.1	Os09g0457900	AP2 domain containing protein
C2H2	LOC_Os03g60570.1	Os03g0820400	ZOS3-22 - C2H2 zinc finger protein
C2H2	LOC_Os11g47630.1	Os11g0702400	ZOS11-10 - C2H2 zinc finger protein
WRKY	LOC_Os01g40260.1	Os01g0584900	WRKY77
G2-like	LOC_Os01g74020.1	Os01g0971800	MYB family transcription factor
HSF	LOC_Os01g53220.1	Os01g0733200	HSF-type DNA-binding domain containing protein

BNC x BAC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
MYB_related	LOC_Os08g06110.2	Os08g0157600	MYB family transcription factor
RAV	LOC_Os01g04750.1	Os01g0140700	B3 DNA binding domain containing protein
bHLH	LOC_Os04g51070.1	Os04g0599300	helix-loop-helix DNA-binding domain containing protein
B3	LOC_Os04g27990.1	Os04g0347400	B3 DNA binding domain containing protein
MIKC	LOC_Os01g52680.1	Os01g0726400	OsMADS32 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os10g26460.1	Os10g0404300	bHLH family protein
FAR1	LOC_Os01g16660.1	Os01g0273200	transposon protein
bZIP	LOC_Os02g10860.1	Os02g0202950	bZIP transcription factor domain containing protein
bHLH	LOC_Os04g31290.1	Os04g0381700	helix-loop-helix DNA-binding domain containing protein
MYB_related	LOC_Os02g46030.1	Os02g0685200	MYB family transcription factor
LBD	LOC_Os03g05510.1	Os03g0149100	DUF260 domain containing protein
LBD	LOC_Os01g39040.1	Os01g0571500	DUF260 domain containing protein
bHLH	LOC_Os07g39940.1	Os07g0588400	helix-loop-helix DNA-binding domain containing protein
bHLH	LOC_Os03g26210.1	Os03g0379300	helix-loop-helix DNA-binding domain containing protein
bZIP	LOC_Os06g50480.1	Os06g0719500	bZIP transcription factor domain containing protein
bHLH	LOC_Os01g72370.1	Os01g0952800	helix-loop-helix DNA-binding domain containing protein
MYB	LOC_Os05g37060.1	Os05g0442400	MYB family transcription factor
MIKC	LOC_Os12g31748.1	Os12g0501700	OsMADS20 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os06g30090.1	Os06g0496400	helix-loop-helix DNA-binding domain containing protein

BNC x BNT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
NAC	LOC_Os09g38010.1	Os09g0552900	no apical meristem protein
HSF	LOC_Os02g13800.1	Os02g0232000	HSF-type DNA-binding domain containing protein
CO-like	LOC_Os06g19444.1	Os06g0298200	CCT/B-box zinc finger protein
HSF	LOC_Os06g35960.1	Os06g0553100	HSF-type DNA-binding domain containing protein
SBP	LOC_Os02g08070.1	Os02g0177300	OsSPL5 - SBP-box gene family member
GRF	LOC_Os02g47280.1	Os02g0701300	growth-regulating factor
NF-YB	LOC_Os02g49410.1	Os02g0725900	histone-like transcription factor and archaeal histone
bHLH	LOC_Os07g36460.1	Os07g0549600	helix-loop-helix DNA-binding domain containing protein
ERF	LOC_Os09g13940.1	Os09g0309700	AP2 domain containing protein
NAC	LOC_Os12g03050.1	Os12g0123800	no apical meristem protein

NAC	LOC_Os04g38720.1	Os04g0460600	no apical meristem protein
NAC	LOC_Os10g42130.1	Os10g0571600	no apical meristem protein
NAC	LOC_Os02g38130.1	Os02g0594800	no apical meristem protein
WRKY	LOC_Os01g40260.1	Os01g0584900	WRKY77
WRKY	LOC_Os01g40430.1	Os01g0586800	WRKY27
C3H	LOC_Os02g10080.1	Os02g0194200	zinc finger C-x8-C-x5-C-x3-H type family protein
C2H2	LOC_Os12g38940.1	Os12g0578600	ZOS12-06 - C2H2 zinc finger protein
B3	LOC_Os01g68370.1	Os01g0911700	B3 DNA binding domain containing protein
NAC	LOC_Os05g34310.1	Os05g0415400	no apical meristem protein
NF-YC	LOC_Os01g39850.1	Os01g0580400	histone-like transcription factor and archaeal histone
MYB_related	LOC_Os02g46030.1	Os02g0685200	MYB family transcription factor
Dof	LOC_Os01g15900.1	Os01g0264000	expressed protein
C2H2	LOC_Os08g20580.1	Os08g0300366	ZOS8-04 - C2H2 zinc finger protein
MYB_related	LOC_Os04g49450.1	Os04g0583900	MYB family transcription factor

BNT x BAT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
bZIP	LOC_Os06g50480.1	Os06g0719500	bZIP transcription factor domain containing protein
bZIP	LOC_Os03g59460.1	Os03g0809200	transcription factor
NAC	LOC_Os07g12340.1	Os07g0225300	NAC domain-containing protein 67
WRKY	LOC_Os08g09800.1	Os08g0198000	WRKY105
NAC	LOC_Os12g03050.1	Os12g0123800	no apical meristem protein
C2H2	LOC_Os03g60560.1	Os03g0820300	ZOS3-21 - C2H2 zinc finger protein
MYB	LOC_Os02g41510.1	Os02g0624300	MYB family transcription factor
HSF	LOC_Os08g43334.1	Os08g0546800	HSF-type DNA-binding domain containing protein
ERF	LOC_Os09g35030.1	Os09g0522200	dehydration-responsive element-binding protein
NAC	LOC_Os11g03370.1	Os11g0127600	no apical meristem protein
GeBP	LOC_Os09g01470.1	Os09g0102300	expressed protein
WRKY	LOC_Os06g44010.1	Os06g0649000	WRKY28
LBD	LOC_Os03g05510.1	Os03g0149100	DUF260 domain containing protein
ERF	LOC_Os09g35010.1	Os09g0522000	dehydration-responsive element-binding protein
MIKC	LOC_Os10g39130.1	Os10g0536100	OsMADS56 - MADS-box family gene with MIKCC type-box
bZIP	LOC_Os11g05640.1	Os11g0154900	bZIP transcription factor domain containing protein
C2H2	LOC_Os03g60570.1	Os03g0820400	ZOS3-22 - C2H2 zinc finger protein
ERF	LOC_Os09g35020.1	Os09g0522100	AP2 domain containing protein
bHLH	LOC_Os04g23550.1	Os04g0301500	basic helix-loop-helix family protein
GRAS	LOC_Os04g35250.1	Os04g0432100	MONOCULM 1
HD-ZIP	LOC_Os02g43330.1	Os02g0649300	homeobox associated leucine zipper
bHLH	LOC_Os06g30090.1	Os06g0496400	helix-loop-helix DNA-binding domain containing protein
LBD	LOC_Os01g39040.1	Os01g0571500	DUF260 domain containing protein

NC x AC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
GATA	LOC_Os12g07120.1	Os12g0168800	GATA zinc finger domain containing protein
C2H2	LOC_Os11g47630.1	Os11g0702400	ZOS11-10 - C2H2 zinc finger protein
MYB	LOC_Os03g29614.1	Os03g0410000	myb-like DNA-binding domain containing protein
MYB_related	LOC_Os05g03550.1	Os05g0126400	MYB family transcription factor
B3	LOC_Os02g25830.1	Os02g0455900	expressed protein
SRS	LOC_Os09g36160.1	Os09g0531600	LRP1
bHLH	LOC_Os08g38080.1	Os08g0487700	helix-loop-helix DNA-binding domain containing protein
FAR1	LOC_Os11g14020.1	Os11g0244900	transposon protein
GRAS	LOC_Os03g09280.1	Os03g0193000	gibberellin response modulator protein
GRAS	LOC_Os10g22430.1	Os10g0369600	gibberellin response modulator protein
Dof	LOC_Os03g42200.1	Os03g0619151	dof zinc finger domain containing protein

FAR1	LOC_Os06g28150.1	Os06g0475800	transposon protein
bZIP	LOC_Os02g14910.1	Os02g0247100	bZIP transcription factor family protein
AP2	LOC_Os04g55970.1	Os04g0653600	AP2-like ethylene-responsive transcription factor AINTEGUMENTA
C2H2	LOC_Os03g10140.1	Os03g0197800	ZOS3-04 - C2H2 zinc finger protein
M-type	LOC_Os01g69850.1	Os01g0922800	OsMADS65 - MADS-box family gene with MIKC* type-box
ERF	LOC_Os06g40150.1	Os06g0604000	AP2 domain containing protein
G2-like	LOC_Os02g22020.1	Os02g0325600	MYB family transcription factor
C2H2	LOC_Os11g47620.1	Os11g0702300	ZOS11-09 - C2H2 zinc finger protein
C2H2	LOC_Os01g39110.1	Os01g0572300	ZOS1-10 - C2H2 zinc finger protein
bZIP	LOC_Os06g50480.1	Os06g0719500	bZIP transcription factor domain containing protein
Nin-like	LOC_Os09g27190.1	Os09g0444100	RWP-RK domain-containing protein
C3H	LOC_Os03g49170.1	Os03g0698800	zinc finger family protein
LSD	LOC_Os12g41700.1	Os12g0611000	LSD1 zinc finger domain containing protein
HD-ZIP	LOC_Os06g48290.1	Os06g0698200	homeobox associated leucine zipper
bHLH	LOC_Os03g51580.1	Os03g0725800	helix-loop-helix DNA-binding domain containing protein
FAR1	LOC_Os11g45530.1	Os11g0681300	transposon protein
M-type	LOC_Os11g43740.1	Os11g0658700	OsMADS68 - MADS-box family gene with MIKC* type-box
WRKY	LOC_Os02g43560.1	Os02g0652100	WRKY34

		NC x NT	
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
bHLH	LOC_Os05g51820.1	Os05g0597000	helix-loop-helix DNA-binding protein
HSF	LOC_Os01g53220.1	Os01g0733200	HSF-type DNA-binding domain containing protein
MYB	LOC_Os07g37210.1	Os07g0558100	MYB family transcription factor
bZIP	LOC_Os01g46970.1	Os01g0658900	transcription factor
C2H2	LOC_Os12g38940.1	Os12g0578600	ZOS12-06 - C2H2 zinc finger protein
ERF	LOC_Os09g13940.1	Os09g0309700	AP2 domain containing protein
HSF	LOC_Os06g35960.1	Os06g0553100	HSF-type DNA-binding domain containing protein
LBD	LOC_Os03g41330.1	Os03g0609500	DUF260 domain containing protein
NAC	LOC_Os04g38720.1	Os04g0460600	no apical meristem protein
bHLH	LOC_Os06g08500.1	Os06g0184000	BHLH transcription factor
GRF	LOC_Os02g47280.1	Os02g0701300	growth-regulating factor
ERF	LOC_Os05g49700.1	Os05g0572000	AP2 domain containing protein
NAC	LOC_Os02g38130.1	Os02g0594800	no apical meristem protein
SBP	LOC_Os06g44860.1	Os06g0659100	OsSPL10 - SBP-box gene family member
C3H	LOC_Os02g10080.1	Os02g0194200	zinc finger C-x8-C-x5-C-x3-H type family protein
HSF	LOC_Os02g13800.1	Os02g0232000	HSF-type DNA-binding domain containing protein
B3	LOC_Os01g68370.1	Os01g0911700	B3 DNA binding domain containing protein
NF-YC	LOC_Os01g39850.1	Os01g0580400	histone-like transcription factor and archaeal histone
Dof	LOC_Os01g15900.1	Os01g0264000	expressed protein
MYB_related	LOC_Os02g46030.1	Os02g0685200	MYB family transcription factor
C2H2	LOC_Os08g20580.1	Os08g0300366	ZOS8-04 - C2H2 zinc finger protein
bZIP	LOC_Os03g59460.1	Os03g0809200	transcription factor
MYB_related	LOC_Os04g49450.1	Os04g0583900	MYB family transcription factor

NT x AT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
MYB	LOC_Os02g40530.1	Os02g0618400	MYB family transcription factor
WRKY	LOC_Os05g09020.1	Os05g0183100	WRKY67
bZIP	LOC_Os06g50480.1	Os06g0719500	bZIP transcription factor domain containing protein
MIKC	LOC_Os10g39130.1	Os10g0536100	OsMADS56 - MADS-box family gene with MIKCC type-box

NAC	LOC_Os12g03040.1	Os12g0123700	no apical meristem protein
HSF	LOC_Os01g43590.1	Os01g0625300	HSF-type DNA-binding domain containing protein
HSF	LOC_Os05g45410.1	Os05g0530400	HSF-type DNA-binding domain containing protein
NAC	LOC_Os07g12340.1	Os07g0225300	NAC domain-containing protein 67
MYB	LOC_Os04g50680.1	Os04g0593200	MYB family transcription factor
WRKY	LOC_Os01g40430.1	Os01g0586800	WRKY27
C2H2	LOC_Os03g32230.1	Os03g0437200	ZOS3-12 - C2H2 zinc finger protein
HSF	LOC_Os09g35790.1	Os09g0526600	HSF-type DNA-binding domain containing protein
WRKY	LOC_Os01g51690.1	Os01g0714800	WRKY26
GRAS	LOC_Os04g35250.1	Os04g0432100	MONOCULM 1
NAC	LOC_Os04g35660.1	Os04g0437000	no apical meristem protein
M-type	LOC_Os01g69850.1	Os01g0922800	OsMADS65 - MADS-box family gene with MIKC* type-box
MIKC	LOC_Os06g45650.1	Os06g0667200	OsMADS30 - MADS-box family gene with MIKCC type-box
WRKY	LOC_Os06g06360.1	Os06g0158100	WRKY113
MYB	LOC_Os02g41510.1	Os02g0624300	MYB family transcription factor
WRKY	LOC_Os06g44010.1	Os06g0649000	WRKY28
HSF	LOC_Os08g43334.1	Os08g0546800	HSF-type DNA-binding domain containing protein
ERF	LOC_Os03g64260.1	Os03g0860100	AP2 domain containing protein
HD-ZIP	LOC_Os06g48290.1	Os06g0698200	homeobox associated leucine zipper
TALE	LOC_Os03g47016.1	Os03g0673000	homeobox protein knotted-1-like 10
NAC	LOC_Os11g03370.1	Os11g0127600	no apical meristem protein
MYB	LOC_Os05g37060.1	Os05g0442400	MYB family transcription factor
C2H2	LOC_Os03g60560.1	Os03g0820300	ZOS3-21 - C2H2 zinc finger protein
HD-ZIP	LOC_Os02g43330.1	Os02g0649300	homeobox associated leucine zipper
NAC	LOC_Os12g03050.1	Os12g0123800	no apical meristem protein
C2H2	LOC_Os03g60570.1	Os03g0820400	ZOS3-22 - C2H2 zinc finger protein
EIL	LOC_Os02g36510.1	Os02g0574800	ethylene-insensitive 3
bHLH	LOC_Os06g30090.1	Os06g0496400	helix-loop-helix DNA-binding domain containing protein

NT x BNT
Não foram identificados fatores de transcrição nesta combinação

Apêndice B: Lista dos fatores de transcrição subexpressos nas 12 combinações estudadas

FATORES DE TRANSCRIÇÃO SUBEXPRESSOS			
AC x AT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
S1Fa-like	LOC_Os04g33440.2	Os04g0408700	DNA-binding protein S1FA
FAR1	LOC_Os04g30870.1	Os04g0377500	transposon protein
NAC	LOC_Os05g34310.1	Os05g0415400	no apical meristem protein
ERF	LOC_Os05g28350.1	Os05g0351200	AP2 domain containing protein
E2F/DP	LOC_Os06g13670.1	Os06g0245900	E2F family transcription factor protein
C2H2	LOC_Os12g07280.1	Os12g0170400	ZOS12-02 - C2H2 zinc finger protein
ERF	LOC_Os05g36100.1	Os05g0437100	AP2 domain containing protein
bZIP	LOC_Os06g41770.1	Os06g0622700	bZIP transcription factor domain containing protein
bHLH	LOC_Os08g43070.1	Os08g0543700	helix-loop-helix DNA-binding domain containing protein
AP2	LOC_Os04g55970.1	Os04g0653600	AP2-like ethylene-responsive transcription factor AINTEGUMENTA
G2-like	LOC_Os02g14490.1	Os02g0241200	MYB family transcription factor
TALE	LOC_Os02g13310.1	Os02g0226600	homeobox domain containing protein
LBD	LOC_Os01g66590.1	Os01g0889400	DUF260 domain containing protein
ERF	LOC_Os10g41130.1	Os10g0560700	AP2 domain containing protein
Nin-like	LOC_Os06g12360.1	Os06g0228900	pentatricopeptide
Nin-like	LOC_Os06g12360.1	Os06g0228900	pentatricopeptide
WRKY	LOC_Os02g43560.1	Os02g0652100	WRKY34
bZIP	LOC_Os11g05640.1	Os11g0154900	bZIP transcription factor domain containing protein
B3	LOC_Os07g17230.1	Os07g0273700	WRKY123
TCP	LOC_Os02g51280.1	Os02g0747400	TCP-domain protein
M-type	LOC_Os06g23950.1	Os06g0347700	OsMADS59 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os01g18870.1	Os01g0293100	helix-loop-helix DNA-binding domain containing protein
WRKY	LOC_Os05g50610.1	Os05g0583000	WRKY8
bZIP	LOC_Os07g44950.1	Os07g0644100	bZIP transcription factor domain containing protein
ERF	LOC_Os02g13710.1	Os02g0231000	ethylene-responsive transcription factor TINY
WRKY	LOC_Os05g40070.1	Os05g0478700	WRKY84
bZIP	LOC_Os07g48820.1	Os07g0687700	transcription factor
E2F/DP	LOC_Os02g33430.1	Os02g0537500	E2F family transcription factor protein
MYB	LOC_Os01g65370.1	Os01g0874300	MYB family transcription factor
C2H2	LOC_Os01g67970.1	Os01g0907400	ZOS1-20 - C2H2 zinc finger protein
B3	LOC_Os08g23470.1	Os08g0324300	B3 DNA binding domain containing protein
M-type	LOC_Os01g23760.1	Os01g0340100	OsMADS93 - MADS-box family gene with M-beta type-box
bZIP	LOC_Os02g14910.1	Os02g0247100	bZIP transcription factor family protein
M-type	LOC_Os01g74440.1	Os01g0975800	OsMADS79 - MADS-box family gene with M-alpha type-box
B3	LOC_Os02g25820.1	Os02g0455800	B3 DNA binding domain containing protein
MYB	LOC_Os12g38400.2	Os12g0572000	MYB family transcription factor
NAC	LOC_Os08g44820.1	Os08g0562200	no apical meristem protein
WRKY	LOC_Os04g21950.1	Os04g0287400	WRKY51
WRKY	LOC_Os05g46020.1	Os05g0537100	WRKY7
HD-ZIP	LOC_Os02g35770.1	Os02g0565600	homeobox associated leucine zipper
C2H2	LOC_Os03g10140.1	Os03g0197800	ZOS3-04 - C2H2 zinc finger protein
MIKC	LOC_Os04g23910.1	Os04g0304400	OsMADS25 - MADS-box family gene with MIKCC type-box
MYB_related	LOC_Os05g49240.1	Os05g0567600	homeodomain-related
MYB	LOC_Os02g54520.1	Os02g0786400	MYB family transcription factor
GATA	LOC_Os02g43150.1	Os02g0645600	GATA zinc finger domain containing protein
SBP	LOC_Os08g40260.1	Os08g0513700	OsSPL15 - SBP-box gene family member

bZIP	LOC_Os06g50480.1	Os06g0719500	bZIP transcription factor domain containing protein
MYB	LOC_Os01g62410.1	Os01g0841500	MYB family transcription factor
GRAS	LOC_Os11g31100.1	Os11g0507300	gibberellin response modulator protein
DBB	LOC_Os06g05890.1	Os06g0152200	B-box zinc finger family protein
ARF	LOC_Os06g46410.1	Os06g0677800	auxin response factor
GRF	LOC_Os02g45570.1	Os02g0678800	growth regulating factor protein
M-type	LOC_Os09g02780.1	Os09g0116000	OsMADS77 - MADS-box family gene with M-alpha type-box
LBD	LOC_Os08g44940.1	Os08g0563400	DUF260 domain containing protein
Nin-like	LOC_Os06g12360.1	Os06g0228950	pentatricopeptide
Nin-like	LOC_Os06g12360.1	Os06g0228950	pentatricopeptide
WRKY	LOC_Os03g55164.1	Os03g0758900	WRKY4
bZIP	LOC_Os07g10890.1	Os07g0209800	transcription factor
NAC	LOC_Os12g29330.1	Os12g0477400	no apical meristem protein
GRAS	LOC_Os01g65900.1	Os01g0881500	chitin-inducible gibberellin-responsive protein
TCP	LOC_Os04g11830.1	Os04g0194600	TCP family transcription factor
bHLH	LOC_Os08g38080.1	Os08g0487700	helix-loop-helix DNA-binding domain containing protein
bHLH	LOC_Os04g47080.1	Os04g0557800	anthocyanin regulatory Lc protein
GeBP	LOC_Os09g01140.1	Os09g0100800	expressed protein
MYB	LOC_Os09g36250.1	Os09g0532900	MYB family transcription factor
ERF	LOC_Os10g25170.1	Os10g0390800	AP2 domain containing protein
GRAS	LOC_Os11g47920.1	Os11g0706200	SCARECROW
NF-YA	LOC_Os08g09690.1	Os08g0196700	nuclear transcription factor Y subunit
ERF	LOC_Os03g60120.1	Os03g0815800	AP2 domain containing protein
bZIP	LOC_Os01g36220.1	Os01g0542700	bZIP transcription factor domain containing protein

AC x BAC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
WRKY	LOC_Os03g21710.1	Os03g0335200	WRKY79

AT x BAT		
Não foram identificados fatores de transcrição nesta combinação		

BAC x BAT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
NF-YA	LOC_Os07g06470.1	Os07g0158500	nuclear transcription factor Y subunit
WRKY	LOC_Os07g02060.1	Os07g0111400	WRKY29
bZIP	LOC_Os02g58670.1	Os02g0833600	bZIP transcription factor domain containing protein
MYB	LOC_Os02g17190.1	Os02g0271900	MYB family transcription factor
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
C3H	LOC_Os06g32720.1	Os06g0519400	zinc finger C-x8-C-x5-C-x3-H type family protein
TCP	LOC_Os04g44440.1	Os04g0526000	TCP family transcription factor
MYB	LOC_Os09g23620.1	Os09g0401000	MYB family transcription factor
TALE	LOC_Os06g01934.1	Os06g0108900	homeobox domain containing protein
bHLH	LOC_Os03g56950.1	Os03g0782500	phytochrome-interacting factor 4
LBD	LOC_Os05g03160.1	Os05g0123000	DUF260 domain containing protein
FAR1	LOC_Os08g14880.1	Os08g0246800	transposon protein
bHLH	LOC_Os03g43810.1	Os03g0639300	helix-loop-helix DNA-binding domain containing protein
HSF	LOC_Os09g35790.1	Os09g0526600	HSF-type DNA-binding domain containing protein
bHLH	LOC_Os04g51070.1	Os04g0599300	helix-loop-helix DNA-binding domain containing protein
bHLH	LOC_Os10g23050.1	Os10g0376900	helix-loop-helix DNA-binding protein
MIKC	LOC_Os04g23910.1	Os04g0304400	OsMADS25 - MADS-box family gene with MIKCC type-box
RAV	LOC_Os01g04800.1	Os01g0141000	B3 DNA binding domain containing protein

bHLH	LOC_Os10g39750.1	Os10g0544200	inducer of CBF expression 2
DBB	LOC_Os01g10580.1	Os01g0202500	B-box zinc finger family protein
MYB_related	LOC_Os02g45670.1	Os02g0680700	myb-like DNA-binding domain containing protein
MYB_related	LOC_Os03g14810.1	Os03g0252900	MYB-like transcription factor DIVARICATA
bHLH	LOC_Os04g41229.1	Os04g0489600	helix-loop-helix DNA-binding domain containing protein
B3	LOC_Os04g31730.1	Os04g0386900	expressed protein
LSD	LOC_Os08g03610.1	Os08g0130100	LSD1 zinc finger domain containing protein
bHLH	LOC_Os03g08930.1	Os03g0188400	helix-loop-helix DNA-binding protein
MYB	LOC_Os01g74410.1	Os01g0975300	MYB family transcription factor
WOX	LOC_Os07g48560.1	Os07g0684900	homeobox domain containing protein
C3H	LOC_Os01g45730.1	Os01g0645000	zinc finger C-x8-C-x5-C-x3-H type family protein
bZIP	LOC_Os01g55150.1	Os01g0756200	bZIP transcription factor domain containing protein
MIKC	LOC_Os06g01890.1	Os06g0108500	MADS-box transcription factor
NAC	LOC_Os08g10080.1	Os08g0200600	no apical meristem protein
MYB_related	LOC_Os08g06110.2	Os08g0157600	MYB family transcription factor
bHLH	LOC_Os06g12210.1	Os06g0226500	helix-loop-helix DNA-binding domain containing protein
NF-YA	LOC_Os08g09690.1	Os08g0196700	nuclear transcription factor Y subunit
RAV	LOC_Os01g04750.1	Os01g0140700	B3 DNA binding domain containing protein
NF-YC	LOC_Os03g14669.1	Os03g0251350	core histone H2A/H2B/H3/H4
LBD	LOC_Os01g03890.1	Os01g0129600	DUF260 domain containing protein
LBD	LOC_Os01g32770.1	Os01g0511000	DUF260 domain containing protein
Trihelix	LOC_Os01g21590.1	Os01g0318400	homeodomain

BNC x BAC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
WRKY	LOC_Os08g13840.1	Os08g0235800	WRKY25
GeBP	LOC_Os09g01140.1	Os09g0100800	expressed protein
MIKC	LOC_Os06g45650.1	Os06g0667266	OsmADS30 - MADS-box family gene with MIKCC type-box
B3	LOC_Os03g42280.1	Os03g0620400	B3 DNA binding domain containing protein
HSF	LOC_Os01g53220.1	Os01g0733200	HSF-type DNA-binding domain containing protein
B3	LOC_Os01g13300.1	Os01g0234100	B3 DNA binding domain containing protein
G2-like	LOC_Os01g74020.1	Os01g0971800	MYB family transcription factor
NF-YB	LOC_Os08g07740.1	Os08g0174500	histone-like transcription factor and archaeal histone
B3	LOC_Os07g17230.1	Os07g0273700	WRKY123
WRKY	LOC_Os05g40070.1	Os05g0478700	WRKY84
bZIP	LOC_Os01g64000.1	Os01g0859300	bZIP transcription factor
NF-YB	LOC_Os02g49410.1	Os02g0725900	histone-like transcription factor and archaeal histone
ERF	LOC_Os05g37640.1	Os05g0448700	expressed protein
ERF	LOC_Os05g37640.1	Os05g0448700	expressed protein
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
NF-YC	LOC_Os01g39850.1	Os01g0580400	histone-like transcription factor and archaeal histone
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
FAR1	LOC_Os11g02964.1	Os11g0123066	FAR1 family protein
ERF	LOC_Os07g47790.1	Os07g0674800	AP2 domain containing protein
B3	LOC_Os03g42230.1	Os03g0619600	B3 DNA binding domain containing protein

BNC x BNT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
B3	LOC_Os06g09420.1	Os06g0194400	B3 DNA binding domain containing protein
MYB	LOC_Os02g17190.1	Os02g0271900	MYB family transcription factor
WRKY	LOC_Os06g44010.1	Os06g0649000	WRKY28
MYB	LOC_Os11g47460.1	Os11g0700500	MYB family transcription factor

B3	LOC_Os06g02230.1	Os06g0112300	B3 DNA binding domain containing protein
ERF	LOC_Os04g46250.1	Os04g0547600	AP2 domain containing protein
ERF	LOC_Os02g42585.1	Os02g0638650	AP2 domain containing protein
C3H	LOC_Os05g45020.1	Os05g0525900	zinc finger/CCCH transcription factor
WOX	LOC_Os07g48560.1	Os07g0684900	homeobox domain containing protein
RAV	LOC_Os01g04800.1	Os01g0141000	B3 DNA binding domain containing protein
bHLH	LOC_Os10g39750.1	Os10g0544200	inducer of CBF expression 2
NAC	LOC_Os08g10080.1	Os08g0200600	no apical meristem protein
ERF	LOC_Os09g35010.1	Os09g0522000	dehydration-responsive element-binding protein
C3H	LOC_Os06g32720.1	Os06g0519400	zinc finger C-x8-C-x5-C-x3-H type family protein
MYB	LOC_Os01g74410.1	Os01g0975300	MYB family transcription factor
HSF	LOC_Os08g43334.1	Os08g0546800	HSF-type DNA-binding domain containing protein
LBD	LOC_Os05g03160.1	Os05g0123000	DUF260 domain containing protein
TCP	LOC_Os04g44440.1	Os04g0526000	TCP family transcription factor
RAV	LOC_Os01g04750.1	Os01g0140700	B3 DNA binding domain containing protein
bZIP	LOC_Os01g55150.1	Os01g0756200	bZIP transcription factor domain containing protein
ERF	LOC_Os02g55380.1	Os02g0797100	AP2 domain containing protein
DBB	LOC_Os01g10580.1	Os01g0202500	B-box zinc finger family protein
LSD	LOC_Os08g03610.1	Os08g0130100	LSD1 zinc finger domain containing protein
NF-YA	LOC_Os08g09690.1	Os08g0196700	nuclear transcription factor Y subunit
ERF	LOC_Os02g45420.1	Os02g0676800	AP2 domain containing protein
bHLH	LOC_Os04g23550.1	Os04g0301500	basic helix-loop-helix family protein
MIKC	LOC_Os04g23910.1	Os04g0304400	OsMADS25 - MADS-box family gene with MIKCc type-box
MYB	LOC_Os09g23620.1	Os09g0401000	MYB family transcription factor
MIKC	LOC_Os06g01890.1	Os06g0108500	MADS-box transcription factor
LBD	LOC_Os01g32770.1	Os01g0511000	DUF260 domain containing protein
ERF	LOC_Os07g47790.1	Os07g0674800	AP2 domain containing protein
ERF	LOC_Os09g35030.1	Os09g0522200	dehydration-responsive element-binding protein
HSF	LOC_Os09g35790.1	Os09g0526600	HSF-type DNA-binding domain containing protein
LBD	LOC_Os01g03890.1	Os01g0129600	DUF260 domain containing protein
Trihelix	LOC_Os01g21590.1	Os01g0318400	homeodomain
ERF	LOC_Os02g45450.1	Os02g0677300	dehydration-responsive element-binding protein

BNT x BAT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
NAC	LOC_Os04g52810.1	Os04g0619000	no apical meristem protein
FAR1	LOC_Os05g30720.1	Os05g0370100	transposon protein
ERF	LOC_Os03g08460.1	Os03g0182800	AP2 domain containing protein
WRKY	LOC_Os08g09900.1	Os08g0198900	WRKY118
ERF	LOC_Os05g37640.1	Os05g0448700	expressed protein
ERF	LOC_Os05g37640.1	Os05g0448700	expressed protein
B3	LOC_Os03g42280.1	Os03g0620400	B3 DNA binding domain containing protein
B3	LOC_Os01g13300.1	Os01g0234100	B3 DNA binding domain containing protein
NAC	LOC_Os05g34310.1	Os05g0415400	no apical meristem protein
C2H2	LOC_Os08g20580.1	Os08g0300366	ZOS8-04 - C2H2 zinc finger protein
TALE	LOC_Os03g56110.1	Os03g0771500	homeobox protein knotted-1
B3	LOC_Os07g17230.1	Os07g0273700	WRKY123
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
ERF	LOC_Os05g37640.1	Os05g0448675	expressed protein
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
NF-YC	LOC_Os01g39850.1	Os01g0580400	histone-like transcription factor and archaeal histone
B3	LOC_Os03g42230.1	Os03g0619600	B3 DNA binding domain containing protein

NC x AC			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
B3	LOC_Os02g38470.1	Os02g0598200	B3 DNA binding domain containing protein
WRKY	LOC_Os01g60640.1	Os01g0821600	WRKY21
bZIP	LOC_Os01g11350.1	Os01g0211800	bZIP transcription factor domain containing protein
WRKY	LOC_Os03g55164.1	Os03g0758900	WRKY4
WRKY	LOC_Os02g08440.1	Os02g0181300	WRKY71
FAR1	LOC_Os12g12380.1	Os12g0225300	transposon protein
FAR1	LOC_Os02g18370.1	Os02g0284500	transposon protein
M-type	LOC_Os01g11510.1	Os01g0213300	OsMADS91 - MADS-box family gene with M-beta type-box
ERF	LOC_Os07g47790.1	Os07g0674800	AP2 domain containing protein
bZIP	LOC_Os01g46970.1	Os01g0658900	transcription factor
B3	LOC_Os10g39190.1	Os10g0537100	B3 DNA binding domain containing protein
ARF	LOC_Os04g36054.1	Os04g0442000	auxin response factor 9
M-type	LOC_Os12g21850.1	Os12g0407000	OsMADS73 - MADS-box family gene with M-alpha type-box
B3	LOC_Os06g42630.1	Os06g0632500	B3 DNA binding domain containing protein
GRAS	LOC_Os01g71970.1	Os01g0948200	GRAS family transcription factor containing protein
MYB	LOC_Os01g59660.1	Os01g0812000	MYB family transcription factor
GRF	LOC_Os02g53690.1	Os02g0776900	growth regulating factor protein
WRKY	LOC_Os05g46020.1	Os05g0537100	WRKY7
bHLH	LOC_Os05g04740.1	Os05g0139100	helix-loop-helix DNA-binding domain containing protein
ERF	LOC_Os02g43970.1	Os02g0657000	AP2 domain containing protein
bZIP	LOC_Os02g52780.1	Os02g0766700	bZIP transcription factor
TALE	LOC_Os05g03884.1	Os05g0129700	homeobox protein knotted-1-like 6
NAC	LOC_Os07g27330.1	Os07g0456900	no apical meristem protein
MYB	LOC_Os08g34960.1	Os08g0450900	MYB family transcription factor
FAR1	LOC_Os07g42400.1	Os07g0615900	transposon protein
ARF	LOC_Os06g48950.1	Os06g0702600	auxin response factor 19
Dof	LOC_Os10g35300.1	Os10g0496000	dof zinc finger domain containing protein
WRKY	LOC_Os12g32250.1	Os12g0507300	WRKY96
NAC	LOC_Os12g29330.1	Os12g0477400	no apical meristem protein
MYB_related	LOC_Os06g04960.1	Os06g0141950	expressed protein
ERF	LOC_Os07g42510.1	Os07g0617000	AP2 domain containing protein
NAC	LOC_Os06g01230.1	Os06g0101800	no apical meristem protein
ZF-HD	LOC_Os11g03420.1	Os11g0128300	ZF-HD protein dimerisation region containing protein
NF-X1	LOC_Os01g06550.1	Os01g0158900	NF-X1-type zinc finger protein
SBP	LOC_Os08g40260.1	Os08g0513700	OsSPL15 - SBP-box gene family member
C2H2	LOC_Os09g03500.1	Os09g0122000	ZOS9-01 - C2H2 zinc finger protein
MIKC	LOC_Os06g06750.1	Os06g0162800	OsMADS5 - MADS-box family gene with MIKCC type-box
C2H2	LOC_Os06g07020.1	Os06g0166200	ZOS6-01 - C2H2 zinc finger protein
MYB	LOC_Os07g25370.1	Os07g0434300	myb-like DNA-binding domain containing protein
HB-other	LOC_Os02g05450.1	Os02g0147800	Homeobox domain containing protein
MYB	LOC_Os04g30890.1	Os04g0377932	MYB family transcription factor
FAR1	LOC_Os11g12490.1	Os11g0231800	transposon protein
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
C2H2	LOC_Os04g08060.1	Os04g0162500	ZOS4-03 - C2H2 zinc finger protein
NAC	LOC_Os03g42630.1	Os03g0624600	No apical meristem protein

NC x NT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
MIKC	LOC_Os01g10504.1	Os01g0201700	OsMADS3 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os03g08930.1	Os03g0188400	helix-loop-helix DNA-binding protein
HSF	LOC_Os08g43334.1	Os08g0546800	HSF-type DNA-binding domain containing protein
ERF	LOC_Os03g64260.1	Os03g0860100	AP2 domain containing protein

ERF	LOC_Os06g08340.1	Os06g0181700	AP2 domain containing protein
bHLH	LOC_Os10g39750.1	Os10g0544200	inducer of CBF expression 2
MIKC	LOC_Os04g49150.1	Os04g0580700	OsMADS17 - MADS-box family gene with MIKCC type-box
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
C3H	LOC_Os06g32720.1	Os06g0519400	zinc finger C-x8-C-x5-C-x3-H type family protein
bHLH	LOC_Os04g54900.1	Os04g0641700	bHLH family protein
WOX	LOC_Os07g48560.1	Os07g0684900	homeobox domain containing protein
MYB	LOC_Os05g28320.1	Os05g0350900	myb-like DNA-binding domain containing protein
TALE	LOC_Os03g47016.1	Os03g0673000	homeobox protein knotted-1-like 10
MYB	LOC_Os01g74410.1	Os01g0975300	MYB family transcription factor
LSD	LOC_Os08g03610.1	Os08g0130100	LSD1 zinc finger domain containing protein
RAV	LOC_Os01g04800.1	Os01g0141000	B3 DNA binding domain containing protein
NF-YA	LOC_Os08g09690.1	Os08g0196700	nuclear transcription factor Y subunit
NAC	LOC_Os08g10080.1	Os08g0200600	no apical meristem protein
ERF	LOC_Os02g55380.1	Os02g0797100	AP2 domain containing protein
bZIP	LOC_Os01g55150.1	Os01g0756200	bZIP transcription factor domain containing protein
ERF	LOC_Os09g35030.1	Os09g0522200	dehydration-responsive element-binding protein
DBB	LOC_Os01g10580.1	Os01g0202500	B-box zinc finger family protein
MIKC	LOC_Os06g01890.1	Os06g0108500	MADS-box transcription factor
LBD	LOC_Os05g03160.1	Os05g0123000	DUF260 domain containing protein
LBD	LOC_Os01g32770.1	Os01g0511000	DUF260 domain containing protein
MIKC	LOC_Os04g23910.1	Os04g0304400	OsMADS25 - MADS-box family gene with MIKCC type-box
HSF	LOC_Os09g35790.1	Os09g0526600	HSF-type DNA-binding domain containing protein
MYB_related	LOC_Os10g30690.1	Os10g0443800	MYB family transcription factor
LBD	LOC_Os01g03890.1	Os01g0129600	DUF260 domain containing protein
bHLH	LOC_Os06g12210.1	Os06g0226500	helix-loop-helix DNA-binding domain containing protein
Trihelix	LOC_Os01g21590.1	Os01g0318400	homeodomain

NT x AT			
Fam. do FT	ID MSU	ID RAP-DB	Função gênica
YABBY	LOC_Os10g36420.1	Os10g0508300	YABBY domain containing protein
C2H2	LOC_Os01g57020.1	Os01g0778100	expressed protein
GRAS	LOC_Os05g42130.1	Os05g0500600	MONOCULM 1
bZIP	LOC_Os08g26880.1	Os08g0357300	bZIP transcription factor domain containing protein
bHLH	LOC_Os03g46860.1	Os03g0671800	helix-loop-helix DNA-binding protein
AP2	LOC_Os02g40070.1	Os02g0614300	AP2-like ethylene-responsive transcription factor PLETHORA 2
C2H2	LOC_Os12g07280.1	Os12g0170400	ZOS12-02 - C2H2 zinc finger protein
NAC	LOC_Os09g38000.1	Os09g0552800	no apical meristem protein
C3H	LOC_Os02g10080.1	Os02g0194200	zinc finger C-x8-C-x5-C-x3-H type family protein
NAC	LOC_Os02g15340.1	Os02g0252200	no apical meristem protein
GeBP	LOC_Os01g14720.1	Os01g0249900	transcription regulator
MIKC	LOC_Os03g11614.1	Os03g0215400	OsMADS1 - MADS-box family gene with MIKCC type-box
HD-ZIP	LOC_Os10g01470.1	Os10g0103700	homeobox associated leucine zipper
M-type	LOC_Os01g69850.1	Os01g0922800	OsMADS65 - MADS-box family gene with MIKC* type-box
MYB	LOC_Os07g37210.1	Os07g0558100	MYB family transcription factor
MYB	LOC_Os01g16810.1	Os01g0274800	MYB family transcription factor
C2H2	LOC_Os02g08510.1	Os02g0182300	ZOS2-04 - C2H2 zinc finger protein
WRKY	LOC_Os01g53260.1	Os01g0734000	WRKY23
C2H2	LOC_Os02g31890.1	Os02g0518500	ZOS2-08 - C2H2 zinc finger protein
NAC	LOC_Os04g43560.1	Os04g0515900	no apical meristem protein
bHLH	LOC_Os12g43620.1	Os12g0632600	helix-loop-helix DNA-binding domain containing protein
C2H2	LOC_Os01g70870.1	Os01g0935000	ZOS1-23 - C2H2 zinc finger protein

ERF	LOC_Os03g08460.1	Os03g0182800	AP2 domain containing protein
NAC	LOC_Os08g33670.1	Os08g0433500	no apical meristem protein
C2H2	LOC_Os08g44190.1	Os08g0555700	ZOS8-13 - C2H2 zinc finger protein
WRKY	LOC_Os08g13840.1	Os08g0235800	WRKY25
WRKY	LOC_Os08g09900.1	Os08g0198900	WRKY118
bZIP	LOC_Os01g64000.1	Os01g0859300	bZIP transcription factor
MYB_related	LOC_Os07g25150.1	Os07g0432800	myb-related protein 306
WRKY	LOC_Os05g40070.1	Os05g0478700	WRKY84
TALE	LOC_Os03g56110.1	Os03g0771500	homeobox protein knotted-1
bHLH	LOC_Os02g35660.1	Os02g0564700	PTF1
B3	LOC_Os01g13300.1	Os01g0234100	B3 DNA binding domain containing protein
bHLH	LOC_Os01g01840.1	Os01g0108400	helix-loop-helix DNA-binding domain containing protein
GRAS	LOC_Os11g04570.1	Os11g0141550	GRAS family transcription factor containing protein
NAC	LOC_Os04g52810.1	Os04g0619000	no apical meristem protein
C2H2	LOC_Os12g38960.1	Os12g0578800	ZOS12-07 - C2H2 zinc finger protein
HD-ZIP	LOC_Os01g57890.1	Os01g0788800	Homeobox domain containing protein
B3	LOC_Os03g42280.1	Os03g0620400	B3 DNA binding domain containing protein
GeBP	LOC_Os09g01140.1	Os09g0100800	expressed protein
B3	LOC_Os07g17230.1	Os07g0273700	WRKY123
NF-YC	LOC_Os01g39850.1	Os01g0580400	histone-like transcription factor and archaeal histone
B3	LOC_Os03g42230.1	Os03g0619600	B3 DNA binding domain containing protein

NT x BNT
Não foram identificados fatores de transcrição nesta combinação