

**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE BIOCÊNCIAS  
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**

**JORGE LUÍS BANDEIRA DA SILVA FILHO**

**GENÔMICA ESTRUTURAL E TRANSCRIPTÔMICA DE GENES  
CORREGULADORES EM *Jatropha curcas* L. SOB SALINIDADE**

**RECIFE  
2019**

**JORGE LUÍS BANDEIRA DA SILVA FILHO**

**GENÔMICA ESTRUTURAL E TRANSCRIPTÔMICA DE GENES  
CORREGULADORES EM *Jatropha curcas* L. SOB SALINIDADE**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas, Área de Concentração Biotecnologia, da Universidade Federal de Pernambuco, como requisito para obtenção do título de Mestre em Ciências Biológicas.

**Área de concentração:** Biotecnologia

**Orientador:** Prof°. Dr. Ederson Akio Kido

**Coorientador:** Dr. Manassés Daniel da Silva

**RECIFE  
2019**

Catálogo na fonte:  
Bibliotecário Bruno Márcio Gouveia - CRB-4/1788

Silva Filho, Jorge Luís Bandeira  
Genômica estrutural e transcriptômica de genes correguladores em *Jatropha curcas*  
L. sob salinidade / Jorge Luís Bandeira Silva Filho. – 2019.

103 f. : il.

Orientadora: Prof. Dr. Ederson Akio Kido.

Coorientador: Dr. Manassés Daniel da Silva.

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Ciências Biológicas, Recife, 2019. Inclui referências.

1. Euphorbiaceae. 2. Pinhão-Manso. 3. Genética vegetal. I. Kido, Ederson Akio. (Orientador). II. Silva, Manassés Daniel da (Coorientador). III. Título.

616.9041

CDD (22.ed.)

UFPE/CB – 2018 – 332

**JORGE LUÍS BANDEIRA DA SILVA FILHO**

**GENÔMICA ESTRUTURAL E TRANSCRIPTÔMICA DE GENES  
CORREGULADORES EM *Jatropha curcas* L. SOB SALINIDADE**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas, Área de Concentração Biotecnologia, da Universidade Federal de Pernambuco, como requisito para obtenção do título de Mestre em Ciências Biológicas.

Aprovada em: 31/07/2019.

**BANCA EXAMINADORA**

---

Prof. Dr. Éderson Akio Kido / Universidade Federal de Pernambuco

---

Prof. Dr. Valdir de Queiroz Balbino / Universidade Federal de Pernambuco

---

Prof<sup>a</sup> Dr<sup>a</sup>. Flavia Figueira Aburjaile / Universidade Federal de Pernambuco

## **AGRADECIMENTOS**

Agradeço primeiramente ao Fundo de Amparo a Pesquisa de Pernambuco (FACEPE) por prover a oportunidade para desenvolvimento deste trabalho.

Agradeço também ao meu orientador Ederson Akio Kido e ao meu Co orientador Manassés Daniel pela ajuda ao longo do desenvolvimento deste projeto.

Agradeço aos colegas de laboratório por toda ajuda e incentivo durante todo o período.

Agradeço aos meus colegas mais próximos Lucas, Marina, Ana, Felipe e Catarina pela ajuda, incentivo e amizade.

Agradeço a minha melhor amiga Adrissa por toda ajuda, conselhos e companhia ao longo destes anos.

Agradeço especialmente a toda minha família pela paciência, conselhos e apoio.

Não menos importante, agradeço a Universidade Federal de Pernambuco pela oportunidade de estudo e infraestrutura para realização deste projeto.

## RESUMO

O pinhão-manso (*Jatropha curcas* L.) é uma planta da família Euphorbiaceae, cujas sementes apresentam teor de óleo entre 33 – 38%, com alta concentração de ácidos graxos insaturados, que são atrativos para a indústria do biodiesel, além disto, a cultura apresenta tolerância à seca porém, é relativamente sensível a salinização dos solos. Redes complexas de regulação gênica controlam a expressão gênica global e as modificações pós-traducionais, que influem na composição metabólica, sendo um fator determinante nas respostas aos estímulos impostos aos vegetais. Uma classe de proteínas que atua indiretamente na transcrição de genes-alvo são os *correguladores* (CoREGs), controlando a expressão gênica como co-ativadores ou co-repressores, ativando ou reprimindo genes. Estudos sobre estes grupos de genes são escassos, principalmente considerando a abrangência das famílias CoREGs na família botânica Euphorbiaceae. O presente estudo realizou uma identificação global de potenciais CoREGs em bibliotecas RNA-Seq de raízes de dois acessos *J. curcas* (tolerante - Jc183 e sensível - Jc171), após exposição (3h) ao sal NaCl (150 mM), respectiva classificação em famílias CoREGs, e associação com respostas de genes diferencialmente expressos [DE; p-value  $\leq$  0,0001, FDR  $\leq$  0,005, Log<sub>2</sub> FC  $\geq$  1 (UR – induzido) ou  $\leq$  -1 (DR - reprimido)]. Transcritos CoREGs DE da resposta do acesso Jc171 (sensível ao sal) foram 239, sendo 25 considerados UR e 214 DR, com destaque para representantes das famílias AUX/IAA, Jumonji e PHD. As famílias AUX/IAA e Jumonji, as quais responderam por mais de 50% dos CoREGs DE observados, são basicamente constituídas de proteínas envolvidas no silenciamento de genes e na modelação da cromatina, o que provavelmente afetou a expressão de diversos genes envolvidos com a resposta ao estresse salino. Genes diferencialmente expressos destas duas famílias são candidatos para estudos futuros visando contribuir na elucidação do processo de expressão em resposta ao estresse salino, alguns deles como transgenes em estudos de transgenia. O acesso Jc183 (tolerante ao sal) não evidenciou CoREGs DE, exceto um. A análise de enriquecimento de FTs, tendo os CoREGs UR como alvos, previu três FTs relacionados com CoREGs AUX/IAA, os quais também interagiriam cada um deles com ao menos outros 45 transcritos UR da resposta de Jc171. Esses FTs capazes de modular genes considerados UR seriam bons candidatos também para eventos de transgenia. Adicionalmente, oito alvos CoREGs (famílias Jumonji, SNF2, mBF1, GNAT, SET e mTERF) foram selecionados

e validados por RT-qPCR, tendo sido confirmados cinco resultados da expressão *in silico*. Espera-se que os resultados possam contribuir para um melhor entendimento da importância das proteínas CoREGs nas respostas aos estresses. Esta compreensão é importante para os programas de melhoramento genético, no propósito de desenvolvimento de acessos mais tolerantes uma vez que, famílias representativas de CoREGs neste trabalho se mostraram importantes na resposta ao estímulo salino. Os resultados podem sugerir bons candidatos à ensaios de transgenia para produção de cultivares com melhores valores econômicos.

Palavras-chave: Estresse abiótico. Data-mining. RNA-Seq. Euphorbiaceae. RT-qPCR.

## ABSTRACT

*Jatropha curcas* L. is a plant of the Euphorbiaceae family, whose seeds have an oil content of 33 - 38%, with a high concentration of unsaturated fatty acids, which are attractive to the biodiesel industry. The crop has tolerance to drought but is relatively sensitive to soil salinization. Complex networks of gene regulation control global gene expression and post-translational modifications, which influence metabolic composition, and are a determining factor in responses to stimuli imposed on plants. One class of proteins that indirectly act on target gene transcription are co-regulators (CoREGs), controlling gene expression as co-activators or co-repressors, activating or repressing genes. Studies on these gene groups are scarce, especially considering the scope of CoREGs families in the botanical family Euphorbiaceae. The present study performed a global identification of potential CoREGs in RNA-Seq libraries of roots of two *J. curcas* accessions (tolerant - Jc183 and sensitive - Jc171), after exposure (3h) to NaCl salt (150 mM), respective classification in families. CoREGs, and association with differentially expressed gene responses [DE;  $p$ -value  $\leq 0.0001$ , FDR  $\leq 0.005$ , Log2 HR  $\geq 1$  (RH - induced) or  $\leq -1$  (DR - repressed)]. CoREGs DE transcripts of the Jc171 (salt-sensitive) access response were 239, of which 25 were considered UR and 214 DR, highlighting representatives of the Aux / IAA, Jumonji, and PHD families. The Aux / IAA and Jumonji families, which accounted for more than 50% of the observed CoREGs, are proteins involved in gene silencing and chromatin modeling, which probably affected the expression of several genes involved with the response to saline stress. Differentially expressed genes from these two families are candidates for future studies to contribute to elucidate the expression process in response to saline stress, some of them as transgenes in transgenic studies. Jc183 (salt tolerant) access did not show CoREGs DE except one. FT enrichment analysis, targeting UR CoREGs, predicted three CoREGs Aux / IAA-related TFs, which would also interact with at least 45 other UR transcripts from the Jc171 response. These TFs capable of modulating UR genes would also be good candidates for transgenic events. In addition, eight CoREGs targets (Jumonji, SNF2, mBF1, GNAT, SET, and mTERF families) were selected and validated by RT-qPCR and five in silico expression results were confirmed. It is hoped that the results may contribute to a better understanding of the importance of CoREGs proteins in stress responses. This understanding is important for breeding programs for developing more tolerant accessions, since

representative families of CoREGs in this work proved to be important in the response to saline stimulation. The results may suggest good candidates for transgenic assays for pod cultivation with better economical values.

Key-words: Abiotic-stress. Data-mining. RNA-Seq. Euphorbiaceae. RT-qPCR.

## LISTA DE FIGURAS

<b>Figura 1</b> – Distribuição da espécie <i>J. curcas</i> globalmente .....	18
<b>Figura 2</b> – Aspectos morfológicos de <i>J. curcas</i> .....	20
<b>Figura 3</b> – Aspectos morfológicos de <i>M. esculenta</i> .....	23
<b>Figura 4</b> – Aspectos morfológicos de <i>R. communis</i> .....	24
<b>Figura 5</b> – Total de artigos para famílias de genes CoREG usando-se palavras-chave (nome da família por extenso + <i>plant</i> ).....	41
<b>Figura 6</b> – Total de artigos para famílias de genes CoREG usando-se palavras-chave (nome da família por extenso + <i>plant</i> + <i>stress</i> ) .....	41
<b>Figura 7</b> – UniGenes CoREGs de <i>J. curcas</i> identificados a partir de similaridades em análise BLASTx.....	49
<b>Figura 8</b> – Heatmaps baseados na clusterização hierárquica para representantes da família CoREG AUX/IAA. ....	52
<b>Figura 9</b> – Heatmaps baseados na clusterização hierárquica para representantes das famílias CoREG Jumonji e PHD.....	53
<b>Figura 10</b> – Heatmaps baseados na clusterização hierárquica para representantes das demais famílias. ....	54
<b>Figura 11</b> – Árvore fenética de proteínas CoREGs da família AUX/IAA. ....	57
<b>Figura 12</b> – Árvore fenética de proteínas CoREGs da família Jumonji. ....	58
<b>Figura 13</b> – Diagrama de Venn ilustrando termos GO enriquecidos.....	59
<b>Figura 14</b> – Análise de enriquecimento de FTs.....	61
<b>Figura 15</b> – FTs enriquecidos, da família bZIP. ....	61

## LISTA DE TABELAS

<b>Tabela 1</b> – Informações genéticas gerais sobre a montagem de genoma mais completa atualmente de <i>J. curcas</i> .....	28
<b>Tabela 2</b> – Informações genéticas gerais sobre a montagem de referência do genoma de <i>Manihot esculenta</i> .....	29
<b>Tabela 3</b> – Informações genéticas gerais sobre a montagem do genoma de referência de <i>Ricinus communis</i> . .....	30
<b>Tabela 4</b> – Relação de famílias de proteínas CoREGs identificadas por Burdo <i>et al.</i> (2014) em <i>Zea mays</i> (milho), comparativamente aos observados na base de dados iTAK para <i>Manihot esculenta</i> (mandioca) .....	40
<b>Tabela 5</b> – Distribuição ordenada das Famílias CoREGs de <i>Jatropha curcas</i> de acordo com a representatividade em UniGenes .....	50
<b>Tabela 6</b> – Número de CoREGs de <i>Manihot esculenta</i> e de similares de <i>Jatropha curcas</i> estimados em cada um dos 18 cromossomo de <i>M. esculenta</i> .....	55
<b>Tabela 7</b> – Lista de eficiência baseada em CT ( <i>circle of treshold</i> ) de cada par de <i>primers</i> testado.....	62
<b>Tabela 8</b> – Expressão relativa de genes CoREGs para Jc171 tratada com estresse salino (150 Mm 3 h).....	63

## LISTA DE ABREVIATURAS

$\mu\text{M}$	Micromolar
ABA	Ácido abscísico
ARID	Domínio de interação rico em AT
AUX/IAA	Auxina / Ácido Indol 3 Acético
BLAST	Ferramenta de pesquisa de alinhamento básico local
CaCl <sub>2</sub>	Cloreto de cálcio
CO <sub>2</sub>	Gás carbônico
CoREG	Corregulador
DAG	Dias após germinação
DDBJ	Banco de DNA do Japão
DE	Diferencialmente expresso
DGE	Expressão gênica Digital
DR	Reprimido
EMBL	Laboratório de biologia molecular europeu
EST	Etiqueta de sequência expressa
FC	Variação de proporção
FDR	Taxa de descoberta falsa
FPKM	Fragmento por quilobase de cDNA por milhão
FT	Fator de transcrição
GNAT	Gcn5 N acetiltransferase
GO	Ontologia gênica
GRN	Redes regulatórias de genes
GS	Genômica estrutural
HAT	Histona acetiltransferase
HMG	Grupo de Alta mobilidade
iTAK	Identificador e classificador de fatores de transcrição e proteínas kinases de plantas
IWS1	Interact With Spt6
Jmj	Jumonji

## LISTA DE ABREVIATURAS

KEGG	Enciclopédia de genes e genomas de Kyoto
LUG	Leunig
Mb	<i>Mega bytes</i>
MBF1	Fator 1 de ligação à multiproteína
MED6	Mediador 6
MED7	Mediador 7
MgSO <sub>4</sub> ·7H <sub>2</sub> O	Sulfato de magnésio
mM	Milimolar
mRNA	RNA mensageiro
mTERF	Fator de transcrição de terminação mitocondrial
NaCl	Cloreto de sódio
Não DE	Não diferencialmente expresso
NCBI	National Center for Biotechnology Information
NGS	Sequenciamento de próxima geração
NR	Não redundante
PHD	<i>Plant homeodomain</i>
PlantRegMap	Plataforma de análise e dados de regulação de plantas
Pseudo ARR B	<i>Type B phospho accepting response regulator</i>
RB	Retinoblastoma
RIN	RNA Integrated Number
RNA Seq	Sequenciamento de RNA
RT-qPCR	PCR quantitativo em tempo real
SAGE	Análise seriada de expressão gênica
SET	SU (VAR) 3 9
SSR	Repetição de sequência simples
SWI/SNF	<i>SWItch/Sucrose não fermentável</i>
TE	Elementos transponíveis
TR	Regulado transcricional
TRAF	Receptor associado ao fator de necrose
UR	Induzido

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>15</b>
1.1 OBJETIVOS.....	17
1.1.1 <b>Objetivo geral.....</b>	<b>17</b>
1.1.2 <b>Objetivos específicos.....</b>	<b>17</b>
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>18</b>
2.1 PINHÃO-MANSO.....	18
2.2 ESPÉCIES TAXONOMICAMENTE RELACIONADAS .....	21
2.3 ESTRESSE SALINO EM PLANTAS.....	24
2.4 ALINHAMENTO DE SEQUÊNCIAS E BANCOS DE BIOMOLÉCULAS .....	26
2.5 VISÃO GERAL DAS METODOLOGIAS DE TRANSCRIPTÔMICA.....	30
2.6 ESTUDOS EM TRANSCRIPTÔMICA DE PLANTAS .....	35
<b>2.6.1 Estudos de transcriptômica em <i>J. curcas</i> .....</b>	<b>36</b>
2.7 GENES CORREGULADORES DE TRANSCRIÇÃO .....	38
<b>3 METODOLOGIA.....</b>	<b>43</b>
3.1 ORIGEM DOS DADOS.....	43
3.2 IDENTIFICAÇÃO DE UNIGENES COREGS .....	44
3.3 IDENTIFICAÇÕES DE POTENCIAIS COREGS DIFERENCIALMENTE EXPRESSOS .....	44
3.4 MAPEAMENTO DE COREGS EM GENOMA DE <i>M. ESCULENTA</i> .....	45
3.5 ÁRVORE FENÉTICA DE FAMÍLIAS COREGS AUX/IAA E JUMONJI .....	45
3.6 ANÁLISE DE ONTOLOGIA GÊNICA E DE FATORES DE TRANSCRIÇÃO.....	45
3.7 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO .....	46
3.8 VALIDAÇÃO DE COREGS DIFERENCIAL EXPRESSOS VIA RT-QPCR .....	46
<b>4 RESULTADOS .....</b>	<b>48</b>
4.1 IDENTIFICAÇÃO DE UNIGENES COREGS EM <i>J. CURCAS</i> .....	48
4.2 PERFIL DE EXPRESSÃO DE COREGS EM <i>J. CURCAS</i> SOB TRATAMENTO SALINO .....	51
4.3 MAPEAMENTO DE COREGS EM GENOMA DE <i>M. ESCULENTA</i> .....	55
4.4 ÁRVORES FENÉTICAS DAS FAMÍLIAS AUX/IAA E JUMONJI.....	56
4.5 ANÁLISE DE ENRIQUECIMENTO DE TERMOS DE ONTOLOGIA GÊNICA....	58
4.6 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO .....	60
4.7 VALIDAÇÃO DE COREGS DIFERENCIAL EXPRESSOS VIA RT-QPCR .....	62

<b>5 DISCUSSÃO</b> .....	<b>64</b>
5.1 IDENTIFICAÇÃO DE UNIGENES COREGS .....	64
5.2 PERFIL DE EXPRESSÃO DE COREGS EM <i>J. CURCAS</i> SOB TRATAMENTO SALINO .....	64
5.2.1 Família AUX/IAA .....	65
5.2.2 Família Jumonji .....	66
5.2.3 Família PHD.....	67
5.2.4 Família SNF2 .....	68
5.2.5 Família SET .....	69
5.3 MAPEAMENTO DE COREGS EM GENOMA DE <i>M. ESCULENTA</i> .....	69
5.4 ANÁLISES FENÉTICAS DAS FAMÍLIAS AUX/IAA E JUMONJI.....	70
5.5 ANÁLISE DE ENRIQUECIMENTO DE TERMOS DE ONTOLOGIA GÊNICA....	71
5.6 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO .....	71
<b>6 CONSIDERAÇÕES FINAIS</b> .....	<b>73</b>
<b>REFERÊNCIAS</b> .....	<b>74</b>

## 1 INTRODUÇÃO

O pinhão-manso (*Jatropha curcas* L.) é uma planta da família Euphorbiaceae, cujas sementes correspondem a 53 - 79% do peso do fruto apresentando teor de óleo entre 33 – 38%, com alta concentração de ácidos graxos insaturados (~73%), sendo estes atrativos para a indústria do biodiesel. A exploração comercial de *J. curcas* no Brasil é recente tendo plantações com menos de duas décadas de idade, não dispondo ainda de cultivares elites ou sistemas de produção plenamente definidos (CARELS, 2009). Apesar de apresentar tolerância ao estresse hídrico, *J. curcas* é relativamente sensível a salinização dos solos. Em áreas candidatas ao seu cultivo, em regiões semiáridas / áridas, de baixa precipitação pluviométrica ou áreas irrigadas com água de qualidade não satisfatória para cultivo, solos salinos e sódicos podem ocorrer naturalmente afetando o crescimento das plantas e conseqüentemente sua produtividade.

As plantas, sendo organismos sésseis, respondem às condições ambientais visando completar o ciclo de vida, orquestrando a expressão gênica, através de interações de proteínas reguladoras da transcrição com as propriedades estruturais de seus DNAs. Análises genômicas de larga escala revelam complexas redes de regulação gênica (*gene regulatory networks*, GRNs) a controlar a expressão gênica global e as modificações pós-traducionais (alosteria), influenciando a composição metabólica (URANO *et al.*, 2010). Essas redes apresentam diferentes tipos de regulação genéticas e epigenéticas, acarretando mudanças na expressão dos fatores de transcrição (FTs), nas modificações em histonas, nas distribuições dos nucleossomos, nas metilações de DNA e na atuação dos RNAs não codificadores de proteínas (npcRNAs, *non-protein-coding* RNAs), os quais influenciam as rotas de genes responsáveis aos estresses (URANO *et al.*, 2010).

Além dos FTs, que correspondem a maior classe de genes reguladores da expressão, outra classe atua indiretamente na transcrição de genes-alvo. As proteínas desta classe não se ligam aos DNAs, nos elementos regulatórios em *cis* presentes nos promotores, como fazem os FTs. Tais proteínas são denominadas correguladores (CoREGs) ou *transcriptional regulators* (TRs), controlando a expressão gênica como co-ativadores ou correpressores, ativando ou reprimindo genes, por mecanismos de interação proteína-proteína, como as interações com FTs, ou restringindo/ liberando

o acesso ao DNA, atuando como modificadores de histonas (FANG *et al.*, 2014) ou remodeladores de cromatina (HAN *et al.*, 2012).

A transcriptômica representa a manifestação global da informação genética acessada por uma célula, em resposta a um estímulo, proporcionando a geração de perfis de expressão dos mais variados genes. Apesar de alguns trabalhos em transcriptômica enfocarem isoladamente alguns genes CoREGs (YANG *et al.*, 2012; QIN, 2013; WANG *et al.*, 2013), em respostas de plantas a estresses abióticos, ou mesmo estes terem sido associados com respostas de tolerâncias a estresses em plantas transformadas, pouco se sabe sobre as famílias CoREGs na família botânica Euphorbiaceae.

Assim, analisar a expressão gênica de genes CoREGs em plantas de *J. curcas* em condições de estresse salino são importantes para os programas de melhoramento, pois podem contribuir para o desenvolvimento de materiais tolerantes ao estresse em questão. Assim, esta pesquisa visou identificar e classificar em famílias transcritos RNA-Seq de *J. curcas* codificadores de CoREGs, expressos em raízes após exposição de três horas ao NaCl (150 mM), visando um melhor entendimento da resposta ao estresse salino.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

Identificar e classificar genes CoREGs de *J. curcas*, a partir de transcritos gerados de bibliotecas RNA-Seq em resposta ao estresse salino ou de bancos públicos, procurando identificar genes candidatos a marcadores moleculares funcionais com potencial uso no melhoramento genético da cultura

### 1.1.2 Objetivos específicos

a) Identificar nas bibliotecas de RNA-Seq de *J. curcas* (em resposta ao estresse salino de 3 h de exposição a 150 mM de NaCl), transcritos relativos às famílias gênicas de CoREGs;

b) Identificar CoREGs diferencialmente expressos nas bibliotecas RNASeq de *J. curcas* e as famílias CoREGs numericamente mais relevantes a fim de traçar uma relação entre estas famílias e o estímulo salino;

c) Identificar Termos GO enriquecidos para grupos de CoREGs induzidos e reprimidos para compreender as funções mais presentes em cada um destes dois grupos;

d) Identificar relações entre fatores de transcrição e CoREGs diferencialmente expressos induzidos, além de observar as relações destes fatores de transcrição e outros transcritos induzidos nas bibliotecas RNA-Seq;

e) Validar alvos CoREGs de importantes famílias de moduladores de cromatina, confirmando as análises *in silico*.

## 2 REFERENCIAL TEÓRICO

### 2.1 PINHÃO-MANSO

A *Jatropha curcas* L. (Figura 2), também conhecida como pinhão-manso é uma espécie da família Euphorbiaceae pertencente ao gênero *Jatropha*. Ela é conhecida popularmente como pinhão-manso. Sua distribuição geográfica se dá ao longo das regiões tropicais e sub-tropicais sendo uma espécie nativa da América (FAIRLESS, 2007). A espécie foi descrita inicialmente pelo botânico sueco Carl Linnaeus em 1753 (GOSWAMI *et al.*, 2013). As Euphorbias em geral, são conhecidas pela produção de fitotoxinas (substância venenosa produzida pelas plantas), e seiva branca de aspecto leitoso. Existem 170 espécies conhecidas de *Jatropha*, a maioria das espécies é conhecida e amplamente cultivadas em todo o trópico (BRITTAINE *et al.*, 2010).

**Figura 1.** Distribuição da espécie *J. curcas* globalmente evidenciada pelas áreas escurecidas

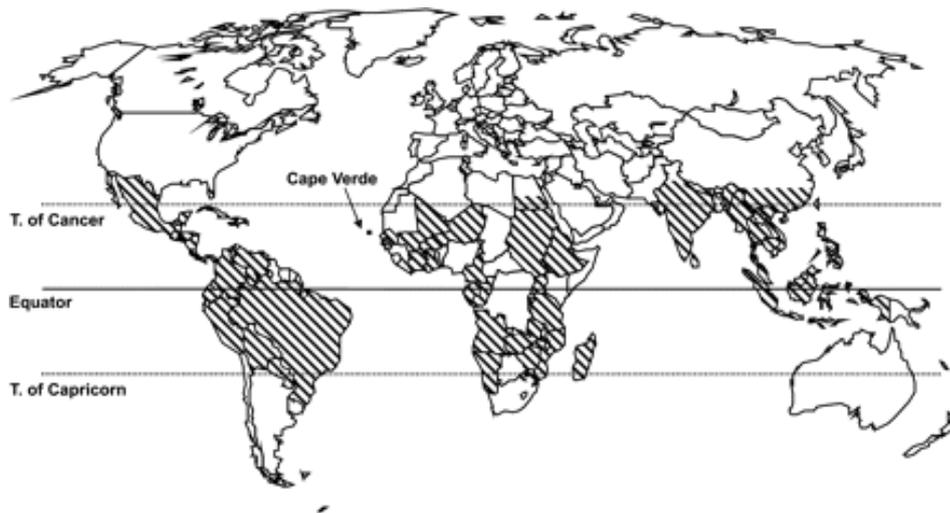


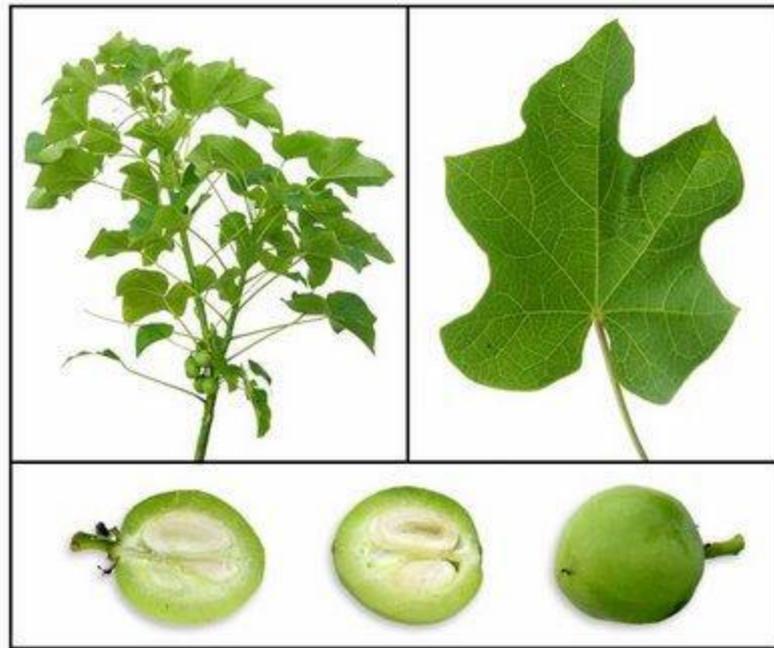
Imagem disponível em King, A.J. *et al* (2009)

Sua morfologia consiste numa espécie arbustiva e de crescimento moderado que leva cerca de quatro anos para atingir sua idade produtiva. Esta ainda pode se estender por cerca de 40 anos (DE ARRUDA *et al.* 2004). Pode chegar a uma altura de quatro metros em média, sendo uma espécie oleaginosa viável em relação a produção e obtenção de biodiesel e com certa tolerância ao estresse hídrico. Segundo Nunes (2007), a maior cultura produtora de óleo vegetal no Brasil é a soja com potencial de produção de óleo vegetal de 412 Kg / ha. Desta forma, o pinhão-manso

apresenta-se como potencial competidor ou substituto na produção de óleo vegetal, uma vez que é capaz de produzir em condições ideais, no mínimo, duas toneladas de óleo por hectare (DE ARRUDA *et al.* 2004; HORSCHUTZ *et al.* 2012). Em seu óleo vegetal é possível encontrar principalmente os ácidos graxos oleico (34,3 - 45,8%; 18: 1), ácido linoleico (29,0-44,2%; 18: 2), ácido palmítico (14,1-15,3%; 16,0) e ácido esteárico (3,7-9,8%; 18: 0) (GUBITZ, 1999).

Seu tronco possui um diâmetro médio de 20 cm e suas raízes são em grande parte de pequeno porte, pivotantes central, com quatro raízes laterais e muitas raízes secundárias. Seu caule apresenta-se de textura lisa e com um miolo mole. Sua rusticidade a faz amplamente distribuída pelo Brasil, sendo encontrada desde a região nordeste até áreas da região sudeste (CORTESÃO, 1956). Os tecidos vasculares das hastes e ramos contêm látex. (BRITTAINE *et al.*, 2010). A espécie é monoica, possuindo flores masculinas e femininas na mesma planta. As flores masculinas possuem coloração amarelada e estão representadas em maior número, localizando-se nas extremidades das ramificações. Em contrapartidas as flores femininas encontram-se logo abaixo das masculinas. Apresenta polinização cruzada, primordialmente entomófila, entre diferentes flores da mesma planta ou de plantas diferentes, sendo parcialmente autocompatível (Figura 2) (VIRGENS *et al.*, 2017; BRITTAINE *et al.*, 2010).

**Figura 2.** Aspectos morfológicos de *J. curcas*; Direta: Folha adulta; Esquerda: Ramos; Abaixo: Sementes



Fonte: <http://www.lideragronomia.com.br/2012/11/pinhao-manso-pinhao-bravo.html>

É uma espécie perene, com baixo efeito alelopático, podendo assim ser utilizada em sistemas agroflorestais, como alternativa para recuperação de áreas degradadas (GOUR, 2006). Apenas as sementes de variedades não tóxicas (caracterizada como isenta ou baixo conteúdo de éster de forbol) provenientes do México, podem ser consumidas por humanos, após processamento de torra (*roasting*). Embora as sementes sejam consideradas a parte mais tóxica da planta, todas as partes da planta de *Jatropha* contêm toxinas como ésteres de forbol, curcinas e inibidores da tripsina. (BRITTAINE *et al.*, 2010).

Além de potencial para biocombustíveis, a espécie também apresenta utilidade no controle da erosão do solo, fabricação de sabão e purgativo para gado bovino, demonstrando assim ser importante para economias locais e expandidas (ARRUDA *et al.*, 2004). Possui folhas com moderado grau de toxicidade para o consumo, porém, as suas sementes depois de tratadas podem ser utilizadas como alimento para animais (ARRUDA *et al.*, 2004). Além disso, suas flores atraem abelhas tendo assim algum potencial melífero. (OPENSHAW, 2000).

Com o advento das tecnologias limpas e uma política de emissão de carbono aumentando sua influência globalmente a cada ano, o interesse em combustíveis renováveis de baixo teor de poluição tem se intensificado. Assim a *J. curcas* mostra-

se uma boa alternativa para desenvolvimento de biodiesel, atendendo a esta crescente demanda. Por conta da facilidade de seu cultivo já são noticiadas pesquisas no intuito do melhoramento desta espécie. Em contraste às principais culturas oleaginosas utilizadas atualmente, não há variedades agronomicamente melhoradas disponíveis para *J. curcas*. E, apesar de *J. curcas* exibir relevante grau de toxicidade relacionada a farinha de sementes, sabe-se que existem variedades comestíveis no México e que atualmente não estão sendo exploradas. Estas variedades comestíveis apresentam considerável potencial para melhoramento que vem sendo discutido (KING *et al.*, 2009). Além da variedade mexicana também existem mais duas que são cultivadas no mundo, a variedade Nicaragua e a Cabo Verde (HENNING, 2004; BRITTAINE *et al.*, 2010).

Segundo Pramanik (2003), as vantagens na substituição parcial de diesel por óleo extraído das sementes de *J. curcas*. É visto que a viscosidade do óleo da espécie é largamente diminuída numa administração 30~40:70~60 (óleo de *Jatropha* / Diesel) mostrando eficiência equiparável a combustível diesel sem mistura. Este é um parâmetro importante a ser avaliado pois, segundo o autor, influencia diretamente no desempenho de um motor. Pramanik demonstra que ainda 50% de uma solução combustível pode ser composta de óleo de *J. curcas* sem perda de eficiência de um motor a diesel. Também há programas de melhoramento desenvolvidos para as culturas de pinhão-manso no Brasil por parte da Embrapa Agroenergia. Estes programas buscam estratégias para a inserção do pinhão-manso na cadeia produtiva do biodiesel (Embrapa Agroenergia, 2009) para a caracterização de coleções de trabalho e desenvolvimento de genótipos melhorados para condições de cada região do Brasil.

## 2.2 ESPÉCIES TAXONOMICAMENTE RELACIONADAS

A família Euphorbiaceae possuem cerca de 6000 espécies (JUDD *et al.*, 2009) dentre as quais, algumas são bastante populares e com certa relevância na cultura mundial, sendo elas a mandioca (*Manihot esculenta*) e a mamona (*Ricinus communis*). *M. esculenta* tem proximidade filogenética com as espécies do gênero *Jatropha*, onde *J. curcas* se encontra. Tal relação pode ser evidenciada por diversos estudos como em Asif *et al.* (2010) que por meio de análises filogenéticas de genes codificadores de proteínas de cloroplasto, desenvolvendo análises de máxima

parcimônia, máxima verossimilhança e evolução mínima, identificaram a proximidade entre os gêneros *Jatropha* e *Manihot*. Segundo Wen *et al.* (2010), a transferibilidade de marcadores de repetição simples (SSR) entre mandioca e pinhão-manso chega a 55% para marcadores ESTs-SSR. Whankaew *et al.* (2011) mostra a transferibilidade de 144 marcadores SSR entre *J. curcas*, *M. esculenta* e *Hevea brasiliensis* (seringueira) onde, pelo menos, 34 marcadores poderiam amplificar o DNA de todas as espécies analisadas.

A mandioca pertence ao *Manihot*, que abriga outras 97 espécies (ROGERS, 1973). É uma cultura de valor econômico elevando sendo um produto da dieta básica para mais de 500 milhões de pessoas que habitam entre os trópicos (BEST & HENRY 1992). O gênero se distribui em grande parte pelos neotrópicos com uma origem evolutiva ainda não bem explicada, tendo-se especulações de que tenha surgido no meio Oeste da América do Sul (OLSEN *et al.*, 1999). Possuem flores unissexuadas, monoclamídeas, sendo o único perianto existente denominado de cálice. Superior a inflorescência encontra-se as flores masculinas enquanto as flores femininas encontram-se distribuídas na base. Seu sistema radicular é tuberoso e é responsável pelo armazenamento das reservas de amido. Além disto, pode possuir também um sistema radicular fibroso. Seu caule é de modelo subarborescente e pode ser indiviso no ciclo vegetativo e ramificado no ciclo reprodutivo. Torna-se lenhoso no estágio adulto da planta adquirindo característica quebradiça. A espécie apresenta folhas simples, inseridas no caule, em disposição alterno-espiralada, lobada e longamente peciolada. O comprimento deste pecíolo pode variar segundo o tipo de cultivar e o estágio de desenvolvimento da espécie. A cor do mesmo pode variar então entre verde, vermelha, rosa e amarelo (DE MATTOS, 2006) (Figura 3).

**Figura 3.** Aspectos morfológicos de *M. esculenta*. Esq. morfologia completa do arbusto. Dir. Aspectos das raízes tuberosas.



Imagem da esquerda disponível em [www.tropilab.com](http://www.tropilab.com); Imagem da direita disponível em [upload.wikimedia.org](http://upload.wikimedia.org)

Diferente de mandioca, a mamona está classificada em um ramo taxonomicamente mais distante de *J. curcas*. Apesar disto, a espécie, sendo da mesma família, apresenta similaridade genética que pode ser vista em estudos como em Sharma & Chauhan (2011) que demonstram cerca de 70% de transferibilidade de marcadores SSR entre *R. communis* e *J. curcas*. Sharma & Chauhan (2012) aplicaram sequências em *J. curcas* a partir de *primers* desenhados para 32 genes de *R. communis* responsáveis pela biossíntese de ácidos graxos. O teste resultou em uma transferabilidade de 65% dos 48 *primers* produzidos.

A espécie *Ricinus communis* L. é conhecida comumente como mamona ou carrapateira. É uma espécie também pertencente à família Euphorbiaceae, assim como a mandioca e o pinhão-mansão sendo a única espécie do gênero *Ricinus*. Sua morfologia é descrita como uma planta arbustiva perene e com ampla distribuição geográfica através do globo. Aclimatou-se relativamente fácil às condições do Norte e Nordeste brasileiro (FIGUEIREDO, 2004). Também há uma grande variabilidade morfológica que é encontrada nos indivíduos desta espécie. Suas lâminas foliares podem apresentar colorações esverdeadas e/ou avermelhadas e variar quanto ao tamanho.

A textura de seus frutos apresenta-se comumente lisa ou com espinhos podendo ter sementes com formatos, volumes, pesos e cores variadas (DE AZEVEDO *et al.* 1997) (Figura 4). As sementes da mamona possuem um endosperma rico em óleo e proteínas. Este óleo é constituído de ácido ricinoleico (ácido D-12-hidroxi-octadec-cis-9-enóico), um ácido graxo hidroxilado (MOREIRA *et al.*, 1996). Este

óleo é uma exclusividade da espécie por ser o único óleo natural que também é hidroxilado sendo de uso frequente na indústria de tintas, vernizes plásticos e cosméticos (VAN DE LOO *et al.*, 1995; SUAREZ *et al.*, 2007).

**Figura 4.** Aspectos morfológicos de *R. communis*. A esq. Desenho esquemático de caule, folhas, frutos e sementes. A dir. Foto de espécime adulta evidenciando folhas e frutos.



Imagem da esquerda, domínio público, fonte: Wikipédia. Foto a direita por Ricardo André Frantz, 31/12/2005.

### 2.3 ESTRESSE SALINO EM PLANTAS

Devido ao aumento da população global, pressões crescentes têm sido responsáveis pela expansão de áreas agricultáveis seguido de áreas de solo degradados, neste último, solos afetados pela salinização (RIBEIRO *et al.*, 2003). A baixa taxa de precipitação pluviométrica associada à baixa atividade bioclimática aliado também a drenagens deficientes e utilização de água de má qualidade, conduzem à formação de solos com alta concentração de sais (CAUÊ DE HOLANDA *et al.*, 2007).

No Brasil, estes aspectos de solo ocorrem no Rio Grande do Sul, na região do Pantanal Mato-grossense e na região semiárida do Nordeste (RIBEIRO *et al.*, 2003). Especialmente, no nordeste brasileiro há grandes áreas com solos salinizados naturalmente e devido as terras mais intensamente cultivadas (SILVA *et al.*, 2011). A região de insuficiência hídrica do Nordeste brasileiro cobre uma superfície de 150 milhões de hectares. Nessa região a maioria dos perímetros de irrigação apresenta

solos com alto teor de salinidade que provocam desde a diminuição nos rendimentos das culturas até o abandono das áreas exploradas (SOUZA *et al.*, 2010).

Aproximadamente 20 % da área total dos perímetros irrigados implantados no Estado de Pernambuco encontram-se com problemas de salinidade e/ou sodicidade (BARROS *et al.*, 2004; MAGALHÃES, 1995). Os solos salinos e sódicos acabam afetando também as espécies de vegetais que estão inseridas nestes solos.

Existem plantas adaptadas aos ambientes salinos. Estas são conhecidas como halófitas (Ex: espécies de manguezais), podendo sobreviver em ambientes onde a concentração de NaCl é superior a 200 mM. Estes grupos de vegetais representam aproximadamente 1,0% de toda flora. As plantas halófitas utilizam algumas estratégias para lidar com os danos causados pela ação do excesso de sal acumulando ou excluindo de forma seletiva íons a nível celular pelos vacúolos, modificando estruturas de membranas, induzindo hormônios e enzimas antioxidantes e alterando as vias fotossintéticas (MUNNS, 2002; MUDGAL *et al.*, 2010).

A maioria das plantas não constam desta tolerância à salinidade e são chamadas de glicófitas, como o pinhão-manso, sendo diretamente afetadas pelo estresse osmótico (FLOWERS & COLMER, 2008; WILLADINO & CAMARA, 2010). O aumento do potencial osmótico do solo aumenta a quantidade de energia necessária utilizada pelos vegetais para a extração de água do solo, além disto, a toxicidade de elementos presentes nestes solos como sódio e os cloretos, em altas concentrações, podem causar distúrbios fisiológicos (BATISTA *et al.*, 2002). Estes sais podem acabar afetando a atividade enzimática em vias metabólicas importantes. Além disto, podem causar a diminuição na eficiência do uso do carbono bem como a decomposição de proteínas e estruturas da membrana (HASEGAWA *et al.*, 2000; FLAGELLA *et al.*, 2004).

Em Gao *et al.* (2008), é visto que em ensaio de aumento da salinidade em solos (tratamentos com 50, 100, 150 e 200 mM NaCl) para germinação de sementes de *J. curcas*, houve uma redução de até 60% da biomassa dos cotilédones, hipocótilos e radículas. O estresse salino também afeta o pinhão-manso de forma significativa na fase de germinação. Quando ocorre alto teor de NaCl a germinação é inibida devido a diminuição do potencial osmótico, dificultando a absorção de água pelas raízes (RIBEIRO *et al.*, 2003; AMORIM *et al.*, 2002; LOPES & MACEDO, 2008).

Em Silva *et al.* (2010), um ensaio com PEG 6000 [11.96% (m/v)] e NaCl 50 mM durante oito dias evidenciou que plantas de *J. curcas* estressadas pelo PEG sofreram

maiores restrições no crescimento das folhas em comparação com as estressadas por sal. Além disso, somente o tratamento com PEG causou um efeito significativo na integridade da membrana foliar. Ambos os tratamentos causaram algum comprometimento das taxas de assimilação de CO<sub>2</sub>. Embora ambos os tratamentos terem causado reduções significativas no conteúdo de clorofila foliar, a atividade fotoquímica não foi afetada. Já em Niu *et al.* (2012), em ensaio para quantificar as respostas do crescimento de *J. curcas* ao estresse salino [NaCl, MgSO<sub>4</sub>·7H<sub>2</sub>O e CaCl<sub>2</sub> em proporção de 87:8:5 (w/w)] em plantas de 3 meses, a análise do peso seco total das plantas avaliadas (após 54 dias de tratamento) diminuiu de 30 a 50%, tendo o estresse salino reduzido significativamente o crescimento e desenvolvimento foliar.

#### 2.4 ALINHAMENTO DE SEQUÊNCIAS E BANCOS DE BIOMOLÉCULAS

A bioinformática é um campo interdisciplinar que envolve as aplicações de ferramentas da informática para a análise de informações biológicas. Ela se caracteriza por ser em área de integração entre a biologia, a ciência da computação, a matemática e a engenharia, em função da análise e interpretação de dados biológicos. Atualmente existe um número crescente de ferramentas e bancos de dados desenvolvidas nesta área.

Dentre a gama de ferramentas bioinformáticas que podem ser utilizadas para melhor entender os dados biológicos obtidos em grande escala encontra-se aquelas derivadas do BLAST (*Basic Local Alignment Search Tool*; Altschul *et al.*, 1990), e que realiza uma análise da similaridade entre sequências ao alinhar localmente a(s) sequência(s) nucleotídica(s) ou proteica(s) de interesse com sequências de uma base de dados. A ferramenta e suas derivações implementadas na plataforma NCBI pode ser vista em <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Dentre os diferentes algoritmos desenvolvidos para a ferramenta BLAST, os mais comumente utilizados são: BLASTn, utilizado para alinhamentos entre sequências nucleotídicas; BLASTp, utilizado para alinhamentos entre sequências proteicas; BLASTx, utilizado para alinhamentos de potenciais produtos de tradução de uma sequência nucleotídica de interesse em relação às sequências proteicas de uma base de dados de proteínas.

Os resultados obtidos através da realização da análise BLAST são alinhamentos locais de maior pontuação (*score*), após considerar diferentes

parâmetros (identidade entre bases ou aminoácidos pareados, abertura e extensão de gaps, etc) entre uma sequência de consulta e outras de uma base de dados de interesse. O BLAST cria uma lista de segmentos pareados de alta pontuação entre a sequência que foi pesquisada e todas as demais sequências nos bancos de dados selecionados. Um segmento configura-se como *substring* de uma sequência, de forma que, possuindo os mesmos tamanhos (segmentos pareados) não são incluídos *gaps*, principal razão da eficiência da ferramenta BLAST. Os alinhamentos então são pontuados utilizando matrizes de substituição. As pontuações, por sua vez, são calculadas em suas medidas de similaridade, por algoritmos dinâmicos, que refletem nos parâmetros: *score*, *e-value*, *identity*, *length*, *gap* e *missmatch*.

O *score* é a pontuação final baseada na soma dos valores obtidos pelo tamanho do alinhamento (*length*; valor de incremento positivo na soma), identidade (*ident*, baseada na igualdade entre cada componente das sequências alinhadas; valor de incremento positivo na soma), componentes de uma sequência alinhados de forma incorreta (*missmatches*; valor de incremento negativo na soma) e necessidade de abertura e extensão para alinhar componentes originariamente não alinhados de uma sequência (*gaps*; valor de incremento negativo na soma).

Um dos parâmetros mais importante nas análises BLAST é o *e-value* que configura-se, simplificada, como um parâmetro que descreve o número de ocorrências que se pode esperar ver por acaso um alinhamento similar e de determinado *score*, ao se pesquisar um banco de dados de um tamanho específico ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=FAQ#expect](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect)). Assim, quanto mais próximo de zero, mais significativo é o alinhamento observado, sendo conveniente o seu uso para se estabelecer limiar de significância (*threshold*) para se considerar os resultados.

Outro recurso importante da bioinformática são os bancos de dados *on-line*, principalmente os depositórios de sequências de biomoléculas que permitem o armazenamento e o compartilhamento dos dados gerados para todo o mundo.

Como exemplos temos o banco americano de dados de sequências de DNA e proteínas (*Genbank*), o banco europeu EMBL (*European Molecular Biology Laboratory*) e o banco japonês DDBJ (*DNA Bank of Japan*). Todos estes bancos são administrados e integram a matriz do *National Center for Biotechnology Information* (NCBI; <http://www.ncbi.nlm.nih.gov/>).

Do mesmo modo, bancos para espécies específicas assim como para tipos específicos de sequências também estão disponíveis na internet. O iTAK (*Plant Transcription factor & Protein Kinase Identifier and Classifier*; [http://itak.feilab.net/cgi-bin/itak/db\\_home.cgi](http://itak.feilab.net/cgi-bin/itak/db_home.cgi)), conta com depositórios de sequências para FTs (TF, *transcript factor*) e genes CoREGs (TR, *transcription regulator*), possuindo sequências nucleotídicas e peptídicas referentes a cada gene, bem como respectivas anotações, de acordo com os melhores alinhamentos advindos contra sequências depositadas no NCBI.

Sobre os dados genéticos armazenados para o pinhão-mansão, no banco de dados online público NCBI existem três projetos genômicos disponíveis (<https://www.ncbi.nlm.nih.gov/assembly/?term=jatropha%20curcas>), o mais recente (JatCur\_1.0 de 02/06/2014 submetido pela *Chinese Academy of Sciences*) indica um genoma de tamanho médio total de 318.36 Mb, com um conteúdo GC% médio de 33,3%. Carvalho *et al.* (2008) demonstra que o genoma de *J. curcas* é relativamente pequeno (C = 416 Mb), semelhante ao de *Oryza sativa*, constituído de 22 cromossomos metacêntricos e submetacêntricos com tamanhos variáveis de 1,71 à 1,24  $\mu\text{m}$

A Tabela 1 reporta informações genômicas gerais da montagem de referência da base de dados NCBI para *J. curcas*.

**Tabela 1.** Informações genéticas gerais sobre a montagem de genoma mais completa atualmente de *J. curcas* (JatCur\_1.0; acesso GCA\_000696525.1 do GenBank), com cobertura de 189x, e sequenciamentos IlluminaGAII e IlluminaHS.

Comprimento total do genoma montado (pb)	318.527.106
Número de <i>Scaffolds</i>	6.024
<i>Scaffold</i> N50	746.835
<i>Scaffold</i> L50	96
Número de <i>contigs</i>	56.154
<i>Contig</i> N50	27.229
<i>Contig</i> L50	2.293
Número total de cromossomos e plasmídeos	1
Número de sequências componentes (WGS ou clone)	56.154

Apesar de haver manuscrito mais recente (HA *et al.*, 2019), este encontra-se mais fragmentado que o genoma de referência do NCBI apresentado acima.

Em relação a sequências de biomoléculas, conforme pesquisa recente (de 05/07/2019), constavam no NCBI 24.713 sequências de genes (*status* refseq) e 77.776 sequências de proteínas depositadas. A espécie ainda não possui seu genoma montado a nível de cromossomos putativos, se resumindo a sequências *scaffolds* de tamanhos variados.

A presença de genomas sendo montados por diferentes e importantes grupos de pesquisas mostra o evidente interesse desses grupos em descobrir genes importantes para o melhoramento da cultura, e com potencial para um melhor aproveitamento econômico de *J. curcas*. Dentre esses genes, destaca-se, ainda, que a espécie também não possui um banco de genes categorizados para CoREGs, de modo diferente de outras duas espécies taxonomicamente relacionadas e abordadas neste trabalho (*M. esculenta* e *R. communis*).

Os dados atuais na plataforma NCBI referentes a *M. esculenta* representam cinco projetos genômicos em andamento (<https://www.ncbi.nlm.nih.gov/genome/genomes/441>), sendo a montagem mais completa ao nível de prováveis cromossomos, a do acesso PRJNA394209 (BioProjeto: PRJNA234389, *Manihot esculenta* v6, submetido por DOE-Joint Genome Institute), a qual indica um genoma de tamanho médio total de 582,27 Mb, com um conteúdo GC% médio de 34,5%. A Tabela 2 descreve outras informações sobre a montagem citada acima.

**Tabela 2.** Informações genéticas gerais sobre a montagem de referência do genoma de *Manihot esculenta* (acesso da montagem do GenBank: GCA\_001659605.1), com cobertura de 121x, e sequenciamento Illumina + CHICAGO

Comprimento total do genoma montado (pb)	582.278.977
Número de <i>Scaffolds</i>	2.020
<i>Scaffold</i> N50	28.119.335
<i>Scaffold</i> L50	10
Número de <i>contigs</i>	39.916
<i>Contig</i> N50	26.702
<i>Contig</i> L50	5.391
Número total de cromossomos e plasmídeos	19
Número de sequências componentes (WGS ou clone)	39.916

Ainda, em pesquisa recente (05/07/2019), constavam nas bases de dados do NCBI um total de 31.987 sequências gênicas e 95.077 sequências proteicas de *M. esculenta*. A espécie também possui dados referentes a 453 sequências CoREGs na base de dados iTAK (v 6.1), sendo 555 contando com isoformas, as quais estão classificadas em 24 famílias (23/01/2019).

Em relação *R. communis*, a espécie conta com duas montagens de genomas em andamento, sendo a mais recente de 2011/07/07 ([https://www.ncbi.nlm.nih.gov/genome/?term=txid3988\[orgn\]](https://www.ncbi.nlm.nih.gov/genome/?term=txid3988[orgn])), submetida pelo *J. Craig Venter Institute*. A montagem de referência (JCVI\_RCG\_1.1) apresenta um tamanho estimado de 350,621 Mb, conteúdo GC% em 34,4%, mas ainda em *scaffolds*, não totalmente definidos cromossomicamente. A Tabela 3 reporta outras informações gerais sobre a montagem.

**Tabela 3.** Informações genéticas gerais sobre a montagem do genoma de referência de *Ricinus communis* (JCVI\_RCG\_1.1; acesso GCA\_000151685.2 do GenBank), com cobertura de 4.5x, e e sequenciamento Illumina.

Comprimento total do genoma montado (pb)	350.621.860
Número de <i>Scaffolds</i>	25.763
<i>Scaffold</i> N50	496.528
<i>Scaffold</i> L50	167
Número de <i>contigs</i>	49.068
<i>Contig</i> N50	21.409
<i>Contig</i> L50	4.197
Número total de cromossomos e plasmídeos	1
Número de sequências componentes (WGS ou clone)	49.406

Nos bancos de dados NCBI, *R. communis* possui 26.128 sequências gênicas e 66.935 sequências proteicas depositadas (pesquisa realizada em 05/07/2019). Em relação aos dados de CoREGs, a espécie possui no banco ITAK 415 sequências, dentre nucleotídicas e peptídicas, originalmente classificadas em 24 famílias de genes TR (23/01/2019).

## 2.5 VISÃO GERAL DAS METODOLOGIAS DE TRANSCRIPTÔMICA

O transcriptoma é o conjunto completo de transcritos (mRNA) presentes em uma célula ou tecido sob determinada condição biológica (NAGALAKSHMI *et al.*,

2010). Os RNAs são produzidos a partir do genoma, nas células de cada indivíduo, e suas proporções variam conforme o desenvolvimento e a condição fisiológica celular e do espécime, podendo variar ainda de acordo com diferentes perfis de expressão de cada gene (WANG *et al.*, 2009).

O estudo sobre expressão gênica pode ser desenvolvido a partir de desenhos experimentais onde um sistema biológico pode ser testado, como a aplicação de determinado estresse. Esses estudos experimentais podem ser complementares aos observacionais, em que são analisadas as diferenças entre fenótipos de um espécime (RITCHIE *et al.*, 2015). Os perfis de expressão podem ser utilizados em estudos comparativos, no qual se observa o contraste entre a condição em estudo em relação a um controle adequado, com o objetivo de caracterizar as expressões gênicas associadas a uma determinada condição.

Estudos em transcriptômica são de grande importância uma vez que podem fornecer informações sobre patogênese, processos metabólicos, tolerância e suscetibilidade, etc. Ainda, propicia a quantificação dos diferentes níveis de expressão de cada transcrito que ocorre durante o desenvolvimento do organismo podendo ser sob diferentes condições (WANG *et al.*, 2009). Logo, o entendimento dos mecanismos desvendados em um estudo de transcriptômica pode auxiliar na interpretação dos elementos funcionais do genoma, fornecendo uma compreensão específica sobre o desenvolvimento do organismo.

As técnicas utilizadas para caracterização dos transcriptomas vem sendo desenvolvidas há algumas décadas. Entre estas merecem destaques: *Expressed Sequence Tag* (EST; ADAMS *et al.*, 1991), *Serial Analysis of Gene Expression* (SAGE; VELCULESCU *et al.*, 1995), *Microarrays* (DUGGAN *et al.*, 1999), e RNA-Seq (MORIN *et al.*, 2008).

As ESTs (ADAMS *et al.*, 1991) visam o estudo da expressão gênica por meio de sequenciamento parcial de clones complementares aos RNAs, os chamados cDNAs. Logo, uma pretensa abundância é detectada a partir das sequências de RNAs correspondentes (HRDLICKOVA *et al.*, 2017). Sendo um método comum para averiguar o padrão de expressão gênica, o objetivo primário da análise de bibliotecas ESTs configura-se como a geração de sequências gênicas (MEYERS *et al.*, 2004). Apesar da ampla aplicabilidade dessa técnica em estudos diversos, a metodologia, seguindo seus protocolos iniciais, apresentava limitações que impediam o uso em

determinadas aplicações, algumas decorrentes da qualidade do sequenciamento, o que refletia na busca por similaridades.

Por sua vez, na época em que foi apresentado, o método SAGE (VELCULESCU *et al.*, 1995) trouxe uma nova abordagem para o sequenciamento SANGER (SANGER, 1997), possibilitando a redução dos custos graças ao sequenciamento de um pequeno fragmento de cDNA (15 pb), amostrado em teoria em mesmo local, para RNAs com PoliA<sup>+</sup> de transcritos de mesma origem. Sua metodologia consiste em fragmentos (*tags*) de 15 pares de bases de extensão que são isolados de cada cDNA que constitui a biblioteca sob análise, sendo que diversas dessas *tags* seriam clonadas e concatenadas em um mesmo vetor plasmidial. Após sequenciamento, as *tags* seriam anotadas por análises de bioinformática. A análise permite, portanto, resultados qualitativos (transcritos expressos) e quantitativos (contagem das *unitags*, ou *tags* únicas, de diferentes composição). Posteriormente, surgiram otimizações da técnica, gerando *tags* maiores, de 21 pb (LONG SAGE; SAHA *et al.*, 2002) e de 26 pb (SuperSAGE; MATSURA *et al.*, 2005), estas últimas minimizando os problemas de anotação em decorrência do pequeno tamanho das *tags*.

A técnica de microarranjo (*microarray*, DUGGAN *et al.*, 1999), por outro lado, tem como base a hibridização e o sequenciamento SANGER. O método possibilitou a caracterização simultânea de níveis de expressão em milhares de transcritos a partir da hibridização com sequências previamente conhecidos ou de candidatos plotados no *chip* de DNA, sendo, portanto, de tecnologia fechada, ao contrário das demais apresentadas anteriormente. Logo, apesar deste método permitir a identificação de transcritos, ele exige informações prévias da sequência para auxílio na análise dos dados, sendo esta uma das limitações (HRDLICKOVA *et al.*, 2017; RITCHIE *et al.*, 2015; KUDAPA *et al.*, 2014).

Já em 2004, Shimkets *et al.* descreveram 27 métodos para análise da expressão gênica em larga escala. Segundo Terauchi *et al.* (2008), métodos aplicados em transcriptômica podem ser divididos em duas categorias principais: a) aquelas baseadas em hibridização de fitas de nucleotídeos complementares a sequências alvos immobilizadas (por exemplo, microarranjos ou chips de DNA); b) aquelas baseadas no sequenciamento e contagem de transcritos (ex: SAGE e seus derivados).

Em meados de 2005, métodos mais eficazes de sequenciamento começaram a surgir no mercado, tendo um impacto significativo no desenvolvimento de estudos genômicos, os chamados sequenciamentos de nova geração (*Next generation sequencing* – NGS) (MOROZOVA & MARRA, 2008). A característica principal surgida destas novas técnicas foi à capacidade de produzir um enorme volume de dados a baixo custo. Em 2012, enquanto um sequenciamento pelo método Sanger 3730xl poderia custar 2.400 dólares/milhão de bases, métodos NGS custaria entre 0.07 e 10 dólares/milhão de bases (LIU *et al.* 2012).

A efetividade da técnica possibilitou, por exemplo, estudos baseados em sequenciar transcriptomas de espécies sem genoma de referência, quantificar transcritos desconhecidos sem conhecimento prévio de genes, além de possibilitar desvendar informações relevantes sobre *splicing* alternativo e variações de seqüências gênicas (WOLD & MYERS, 2008; WANG, 2009). A capacidade de sequenciar o genoma de diversos organismos de forma prática e relativamente rápida (de uma hora à 14 dias, dependendo da plataforma de sequenciamento) permitiu que mais estudos em transcriptômica fossem produzidos (METZKER, 2009; MARDIS, 2008; LIU *et al.*, 2012).

Segundo Mardis (2008), no final da década passada, haviam três plataformas de sequenciamento em destaque: *Roche/454 FLX*, *Illumina/Solexa Genome Analyzer*, e *Applied Biosystems SOLiD System*. Cada uma destas plataformas incorpora uma interação complexa de enzimologia, química, ótica de alta resolução, *hardware* e engenharia de software, para desenvolver suas análises de forma eficiente. Estas características permitem etapas simplificadas de preparação da amostra antes do sequenciamento do DNA, que fornecem uma economia de tempo e um requisito mínimo para equipamentos associados, em comparação aos de múltiplas etapas necessários para o sequenciamento baseado em clones (MARDIS, 2008).

As tecnologias NGS permitiram assim o desenvolvimentos do sequenciamento completo dos transcriptomas (RNA-Seq, WANG, 2009). Comparado com o tradicional sequenciamento EST de baixo rendimento pela tecnologia Sanger, que só detecta os transcritos mais abundantes, o sequenciamento em larga cobertura (100 - 1.000 leituras por par de bases de uma transcrição) de um experimento típico de RNA-Seq oferece uma visão mais fiel de um transcriptoma, incluindo transcritos raros.

Em contraste com tecnologias alternativas de alto rendimento em transcriptômica, como microarranjos, a técnica de RNA-Seq, a partir de

sequenciamento NGS, atinge a resolução a nível de par de bases, e uma faixa dinâmica muito mais alta de níveis de expressão, além de produzir anotações *de novo* (OZSOLAK & MILOS, 2011; WANG, 2009; MARTIN & WANG, 2011). Adicionalmente, os resultados de RNA-Seq mostram altos níveis de reprodutibilidade (WANG *et al.*, 2009; OZSOLAK & MILOS, 2011) e os estudos de seus transcritos podem fornecer informações de grande utilidade para o estudo de transcriptomas complexos, como a conexão entre exons, (WANG *et al.*, 2009).

De outro modo, ao contrário das abordagens que visam hibridizações, como o microarranjo, a técnica de RNA-Seq não está limitada aos transcritos previamente conhecidos, o que a torna atraente para sequenciamento de transcriptomas de organismos não modelos e sem sequências genômicas conhecidas (MARTIN & WANG, 2011). Desta forma, a RNA-Seq se popularizou com os estudos de expressão gênica, propiciando a identificação de transcritos inéditos, revelando a identidade das sequências, e sendo importante, ao final, para análise de genes desconhecidos e isoformas (HRDLICKOVA *et al.*, 2017).

Logo, a utilização das tecnologias NGS para RNA-Seq propicia níveis elevados de expressão em curto espaço de tempo que, conjuntamente com técnicas de bioinformática possibilitam a montagem e anotação de transcriptomas sem genomas de referência (estratégia *de novo*), o que é fundamental para espécies sem genoma sequenciado ou devidamente montado (OZSOLAK & MILOS, 2011; WANG, 2009)

Embora as análises de alto rendimento como RNA-Seq permitam uma análise global e em ampla escala dos padrões da expressão gênica, os padrões de expressão gênica necessitam ser validados por uma segunda técnica. Neste caso, a técnica de RT-qPCR (HEID *et al.*, 1996), é considerada, por sua eficiência, o padrão ouro nesta validação dos perfis de genes diferencialmente expressos e detectados *in silico* (BUSTIN e NOLAN, 2004). A técnica mede a acumulação de produtos de PCR através de uma sonda fluorogénica de dupla marcação. Este método fornece quantificações precisas de cópias de genes. Ao contrário de outros métodos de PCR quantitativos, a PCR em tempo real não requer o manuseio da amostra pós-PCR, impedindo a potencial contaminação de transporte do produto de PCR e resultando em ensaios de rendimento muito mais rápidos e maiores (HEID *et al.*, 1996).

## 2.6 ESTUDOS EM TRANSCRIPTÔMICA DE PLANTAS

Com o avanço das técnicas de transcriptômica, programas de melhoramento puderam desenvolver cultivares de forma mais fácil e eficiente. Estes tipos de estudos tornaram-se muito importantes para o melhor entendimento do comportamento da maquinaria genética dos indivíduos, e principalmente o entendimento destes maquinários em resposta ao estresse. A relevância das técnicas de transcriptômica no auxílio aos programas de melhoramento de plantas pode ser evidenciada a partir dos exemplos abaixo:

Utsumi *et al* (2012) mostraram em *M. esculenta*, por meio de *microarray* (chip *Agilent* de oligonucleotídeos com 20.000 genes), a identificação de 1.300 transcritos UR (*Induzido*, super-expressos, limiares de *p-value* < 0,01 e *Fold Change* > 2) sob estresse hídrico (16 horas sob 40-80  $\mu\text{mol}$  de fótons  $\text{m}^{-2} \text{s}^{-1}$  a 30 ° C em 50% de umidade relativa);

Ding *et al* (2016) realizaram estudo em transcriptômica, a partir de folhas de mandioca sob condições de plena luz e sombra natural, revelando perfis de expressão gênica similares para genes relacionados com: parede celular, metabolismo celular básico e lipídico, fotossíntese e metabolismo de carboidratos; acessos submetidos a sombra induziram significativamente a expressão de genes envolvidos com reação de luz da fotossíntese, sinalização luminosa e síntese de DNA/estrutura da cromatina; resultados de redução na expressão foram relacionados à biossíntese de antocianinas, choque térmico, ciclo de calvin, glicólise, transporte de elétrons mitocondriais, metabolismo de amido e de sacarose;

Chandrasekaran *et al*, (2010) em estudo transcriptômico com *R. communis*, a partir de tecidos cultivados com ABA (Ácido abscísico, 10  $\mu\text{M}$  em meio MS), e sem ABA (controle)) identificou 2.568 genes diferencialmente expressos (DE, *p-value* 0,05; FDR 0,001; *Fold Change*  $\geq$  1), sendo 1.507 UR e 1.061 DR (*Reprimido*). Genes DE foram envolvidos com metabolismo de açúcares, lipídios, síntese de proteínas e a biossíntese de ABA (importante fito-hormônio na resposta a estresses e no desenvolvimento de plantas);

You *et al*. (2019), em estudo de estresse hídrico com gergelim (*Sesamum indicum*) e estresse imposto em 25 dias após semeadura em solos com diferentes teores de umidade (16%, 13%, 10% e 8%), demonstraram a expressão diferencial de

684 transcritos UR e 1346 DR (FDR <0,01;  $\log_2FC >1$ ), envolvidos principalmente no processamento de proteínas no retículo endoplasmático, na fotossíntese, na transdução de sinais de hormônios vegetais, no metabolismo lipídico, e no metabolismo de aminoácidos; metabólitos acumulados sob estresse hídrico incluíram ácido abscísico (ABA), aminoácidos e ácidos orgânicos; o genótipo suscetível à seca demonstrou ser mais perturbado pelo estresse hídrico nos níveis transcricional e metabólico.

Qiu *et al.* (2017) investigando as respostas fisiológicas e transcriptômicas de *Brassica rapa* ao estresse salino (200 mM de NaCl, 12 horas de exposição) detectaram 1.235 genes DE [limiares combinando probabilidade de divergência (função que estabelece a diferença de uma distribuição de probabilidade para outra em uma variedade estatística)  $\geq 0,8$  e  $\log_2FC \geq 1$ ], sendo 740 UR e 495 DR, estando relacionados à transdução de sinais, síntese de osmólitos, fatores de transcrição e proteínas antioxidantes.

Keshishian *et al.* (2018) examinaram níveis de citocinina em conjunto com alterações do transcriptoma RNA-Seq para *Solanum lycopersicum* em resposta ao estresse salino (150 mM NaCl, 24 horas de exposição), identificando 6.643 transcritos DE (FDR  $\leq 0,05$ ; p-value  $\leq 0,05$ ) e vários genes de citocinina.

Estes exemplos demonstram a importância dos estudos em transcriptômica para interpretação de dados biológicos e entendimento das respostas aos estímulos aplicados, a fim de promover o melhoramento genético das espécies estudadas.

### **2.6.1 Estudos de transcriptômica em *J. curcas***

Em pesquisa recente (03/2019) na base de dados do PubMed, foram identificados 4.007 artigos que utilizavam em seu título ou resumo o termo “*transcriptomic*” e “Plant” em seu texto. Ao adicionar o termo “salt” ao texto foram identificados 157 artigos, e ao adicionar o termo “*Jatropha*” ao texto, somente um artigo, referente a Zhang *et al.* (2014) que será apresentado adiante.

Conforme mencionado, *J. curcas* apresenta certa tolerância ao estresse hídrico, boa adaptação em solos pobres, e não competição com culturas produtoras de alimentos. Assim, a transcriptoma desta espécie pode ser uma ferramenta valiosa na caracterização dos mecanismos de tolerância à estresses e aos climas áridos e semi-áridos (SAPETA *et al.*, 2015; DIVAKARA *et al.*, 2010).

Como exemplo de estudo transcriptômico em *Jatropha curcas*, Costa *et al.* (2010) geraram bibliotecas de cDNA, a partir de *pools* de desenvolvimento (JD: 19, 26, 33 e 40 dias após a polinização), e de germinação do endosperma (JG: 24, 36, 48 e 72 horas após a embebição) de sementes, sendo sequenciados 7.320 ESTs do *pool* JD e 5.929 do *pool* JG, totalizando 13.249 ESTs, os quais geraram 1.606 *contigs* e 5.677 *singletons*. Dessas sequências, 6.694 (91,9%) foram codificadoras de proteínas, com destaque para proteínas quinases de serina/treonina (119 sequências) e fatores de transcrição (89). Dos 20 principais transcritos mais expressos em JD, três codificariam proteínas de armazenamento (globulinas 11S), além de um (para aspartil protease), que processaria precursores de proteínas de armazenamento. Diferenças também foram observadas em relação a elementos transponíveis (TE) nas bibliotecas JD e JG, além de permitir a identificação de expressão diferencial e a descoberta de vários genes, inclusive relacionados ao metabolismo de lipídeos.

Em outro estudo, Natarajan & Parani (2011) aplicaram a técnica de alto rendimento, denominada de pirosequenciamento (NYREN, 1993), em cDNAs não clonados, de raízes, folhas maduras, flores, sementes em desenvolvimento, e embrião maduro de *J. curcas*, sendo geradas 381.957 *reads* de boa qualidade, as quais geraram 17.457 *contigs* e 54.002 *singletons*. Desses, 28 transcritos foram validados por análises de RT-qPCR. Análises na base de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*; KANEHISA & GOTO, 2000) mostraram que 2.320 transcritos estavam relacionados às principais vias bioquímicas, incluindo vias de biossíntese de lipídios.

Por sua vez, Wang *et al.* (2013) performaram uma abordagem RNA-Seq de sequenciamento de alto rendimento (Illumina HiSeq™ 2000), seguido por análise de expressão gênica digital, resultando em uma análise global de transcriptoma e perfil de expressão gênica para plântulas de *J. curcas*, em resposta à exposição ao frio (12 °C por 12h, 24h e 48h). Foram identificados 45.251 unigenes, dos quais 33.848 foram anotados a partir de banco de dados de proteínas. Análises DGE (*Digital gene expression*) individuais, 4.185 genes foram expressos diferencialmente (regulados positivamente ou regulados negativamente no tratamento a frio em relação ao controle;  $p\text{-value} \leq 0.01$ ;  $FDR \leq 0.001$ ;  $\log_2\text{Ratio} \geq 1$ ). O número de genes UR e DR mudou gradualmente após 12 e 24 horas de estresse por frio, enquanto aumentou rapidamente entre 24 e 48 horas. Os resultados informam dados relevantes ao que diz respeito aos mecanismos moleculares da resistência ao frio de *J. curcas*.

No único estudo encontrado na base de dados PubMed relacionando estresse salino e transcriptômica em *J. curcas*, Zhang *et al*, 2014, é vista uma análise global de perfis de expressão gênica em mudas de pinhão-mansão expostas ao estresse. Foi feito sequenciamento para analisar perfis globais de expressão gênica de plantas de pinhão-mansão (raízes e folhas; 2 h, dois dias e sete dias após o início do estresse salino). Para o estresse salino, as mudas foram irrigadas diariamente por sete dias no estágio de seis folhas à solução nutritiva de Hoagland com NaCl 100 mM. Duas réplicas biológicas, um total de 24 amostras foram sequenciadas. A partir do sequenciamento foi visto que 1.504 e 1.115 genes foram significativamente expressos em raízes e folhas respectivamente ( $p < 0,01$ ; *fold change*  $\geq 1,8$ ). Análise de ontologia genética (GO) demonstrou que, muitos processos biológicos foram afetados pelo estresse salino, em particular aquelas pertencentes aos processos metabólicos. Os perfis de expressão gênica indicaram que os genes associados foram responsáveis pela via de sinal do ABA, do etileno e regulação osmótica. Zhang explica que os principais genes regulados detectados celular em raízes foram relacionados à síntese de trealose e modificação da estrutura da parede, enquanto que os relacionados à síntese de rafinose e ao removedor de oxigênio reativo foram mais presentes em folhas.

## 2.7 GENES CORREGULADORES DE TRANSCRIÇÃO

O controle da transcrição é um processo de fundamental importância biológica para todos os organismos, sendo os CoREGs entidades que desempenham papéis de co-ativação ou co-repressão no genoma destes indivíduos (O'MALLEY, 2007). Co-ativadores existem e funcionam em grandes complexos multiproteicos, os quais são recrutados para o gene alvo por receptores nucleares (NR), sendo portadores de capacidades enzimáticas necessárias para a expressão dos genes. As reações relacionadas com a transcrição e mediadas por complexos co-ativadores incluem a modificação e a remodelação da cromatina, iniciação da transcrição, alongamento de cadeias de RNA, *splicing* de RNA e, término da resposta transcricional (O'MALLEY, 2007). Segundo Britten e Davidson (1969), no genoma existem um subgrupo de genes que atuam como “genes mestres” coordenando a expressão de genes subservientes denominados “genes produtores”. Esta hipótese foi verificada por O'Malley (2007) em

um manuscrito sobre genes CoREGs, sugerindo que genes codificadores de proteínas CoREGs fizessem parte do grupo de genes mestre.

Burdo *et al.* (2009) denominaram a coleção de CoREGs como sendo o CoREGoma. Em um segundo trabalho, Burdo *et al.* (2014) evidenciaram que as bases para identificação e caracterização de genes CoREGs em grupos estruturados (de famílias) ainda não estavam bem estabelecidas. Naquele trabalho além de fatores de transcrição em milho, os autores também identificaram genes CoREGs e os classificaram em 24 famílias, a partir de domínios proteicos requeridos e proibidos de cada família gênica conforme descritas por Perez-Rodriguez *et al.* (2009).

Dados como os de Burdo *et al.* (2014) podem ser encontrados na base de dados *online* GRASSIUS (<http://grassius.org/grasscoregdb.html>), a qual armazena sequências CoREGs de diferentes gramíneas.

Em decorrência de grandes lacunas nas pesquisas, muitas espécies carecem de informações completas ou relevantes sobre um possível CoREGoma, como *J. curcas*, ao contrário de *Zea mays* (milho), uma espécie que já apresenta número significativo de estudos e de CoREGs (Tabela 4). Também *M. esculenta*, espécie taxonomicamente próxima à *J. curcas*, já possui algumas informações de CoREGs compiladas (Tabela 4).

Os CoREGs presentes no banco iTAK tem suas identidades e classificação baseadas nas regras consenso (de domínios proteicos requeridos e proibidos de cada família gênica), resumidos das bases de dados PInTFDB (PEREZ-RODRIGUEZ *et al.*, 2009) e PlantTFDB (JIN *et al.*, 2013), além das famílias descritas em PlantTFcat (DAI, X. *et al.*, 2013) e AtTFDB (YILMAZ *et al.*, 2010), como evidências de apoio. Mais de vinte e uma famílias de CoREGs foram caracterizadas em plantas de acordo com os bancos de dados supracitados.

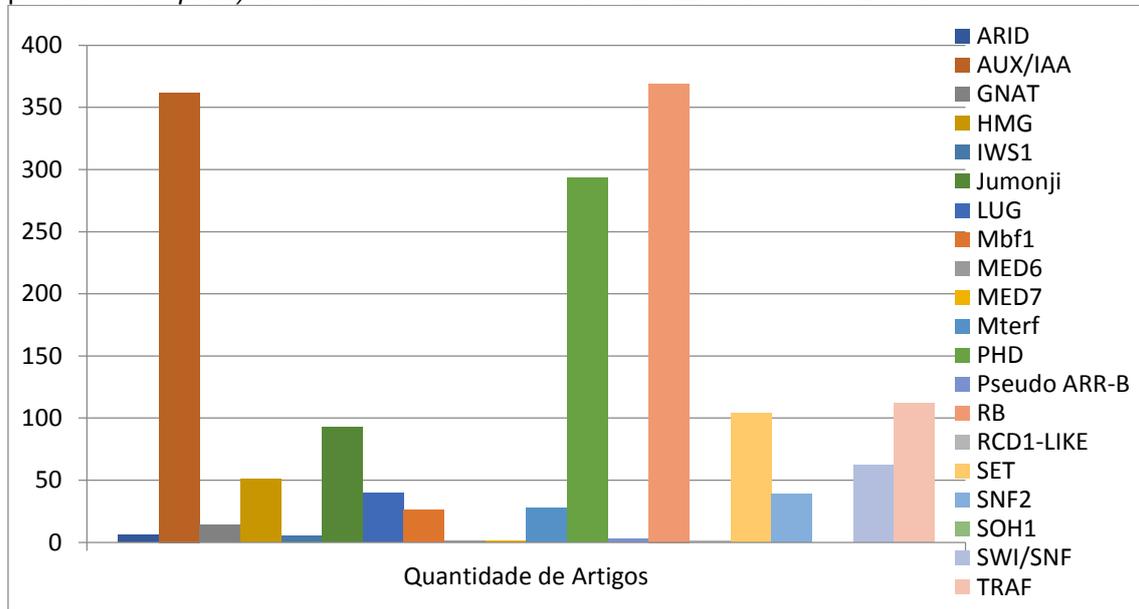
**Tabela 4.** Relação de famílias de proteínas CoREGs identificadas por Burdo *et al.* (2014) em *Zea mays* (milho), comparativamente aos observados na base de dados iTAK para *Manihot esculenta* (mandioca). "X" marca presença da família na espécie em julho de 2018.

<b>Família</b>	<b>Milho</b>	<b>Mandioca</b>
ARID		X
AUX/IAA	X	X
BSD	X	
Co-activator p15	X	X
DDT	X	
FHA	X	
GNAT	X	X
HMG	X	X
IWS	X	X
Jumonji		X
Lim	X	
LUG	X	X
MBF1	X	X
MED6	X	X
MED7	X	X
MED26	X	
MED31	X	
PHD		X
Pseudo ARR-B		X
RB	X	X
RCD1-LIKE	X	X
SET		X
SNF2	X	X
SOH1		X
SWI/SNF-SWI3	X	X
SWI/SNF-BAF60	X	X
TAZ	X	X
TRAF	X	X
Ultrapetala	X	
WD40	X	

Como forma de evidenciar as famílias CoREGs mais estudadas e com potencial sobre respostas a estresses bióticos e/ou abióticos foram realizadas buscas (em 06/09/2018) através de palavras-chave na base de dados *online MedLine*. Na Figura 5 é possível verificar a quantidade total de artigos presentes na base de dados, até a

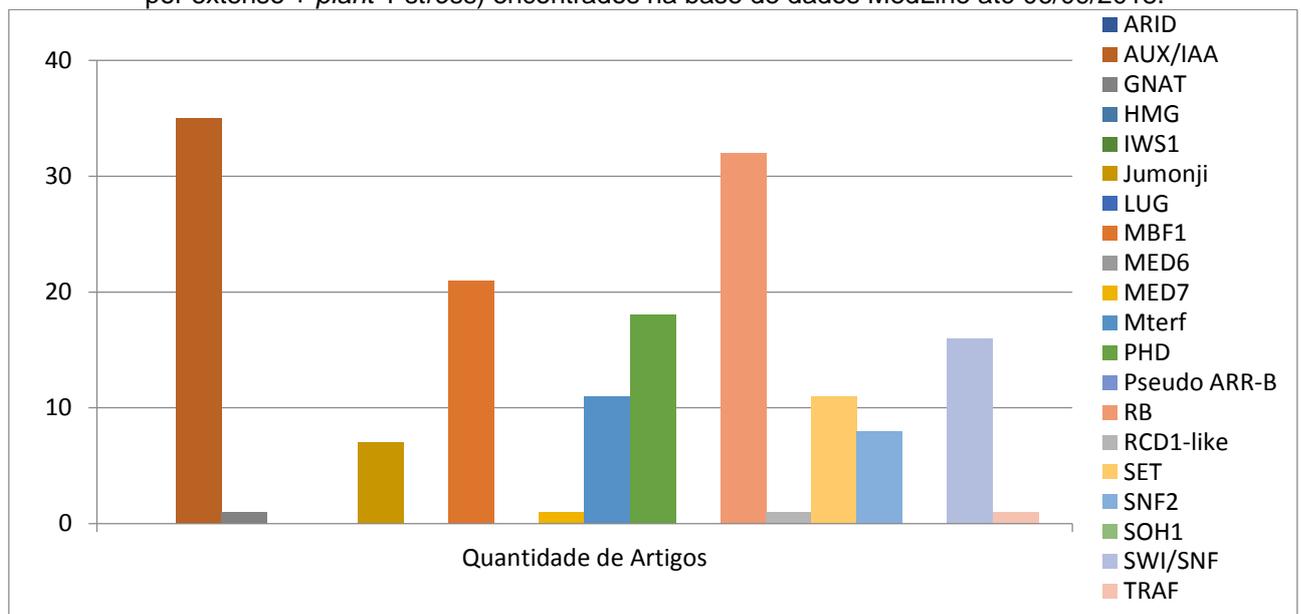
data em questão, para algumas das famílias CoREGs, a partir do uso de palavras-chave [(nome da família) e *plant*], com destaque para RB, AUX/IAA e PHD.

**Figura 5.** Total de artigos para famílias de genes CoREG usando-se palavras-chave (nome da família por extenso + *plant*) encontrados na base de dados MedLine até 06/09/2018.



Na Figura 6 é possível verificar a quantidade de artigos até aquela data, e que relacionavam as palavras-chave, além de estresse [(nome da família), *plant*, *stress*]. Neste caso, as famílias em destaque foram AUX/IAA, RB e mBF1.

**Figura 6.** Total de artigos para famílias de genes CoREG usando-se palavras-chave (nome da família por extenso + *plant* + *stress*) encontrados na base de dados MedLine até 06/09/2018.



No geral, a importância dos CoREGs como co-ativadores e correpressores influenciam todo um conjunto de reações importantes para o desenvolvimento do indivíduo em um dado momento de seu ciclo. Exemplos de funções e outros aspectos relativos a algumas das famílias CoREGs são destacados a seguir.

Família GNAT: como parte da subfamília GNC5, as histonas acetiltransferases (HATs) são fundamentais no desenvolvimento de *A. thaliana* (VLACHONASIOS *et al.*, 2003), além de estarem envolvidas em respostas a estresses, como luz, temperatura, salino e também hídrico (FANG *et al.*, 2014).

Família ARID: constitui característica a presença de domínio (ARID) de ligação ao DNA, sendo conservado em eucariotos superiores, entretanto, nem todas as proteínas contendo ARID se ligam ao DNA de maneira característica da família ARID; suas funções celulares incluem regulação do crescimento, diferenciação e desenvolvimento celular (WILSKER *et al.*, 2005); podem ocorrer em combinação com outros domínios de proteínas, incluindo os domínios Myb, PHD e HMG (RIECHMANN *et al.*, 2000); em *A. thaliana* foram identificadas onze proteínas ARID de cinco subfamílias (WILSKER *et al.*, 2005).

Família HMG: é característico em eucariotos domínio de proteína com aproximadamente 75 resíduos (ANTOSCH *et al.*, 2012); proteínas HMG induzem a dobra do DNA ao se ligar ao DNA linear, atuando como facilitadores arquitetônicos na montagem de complexos nucleoproteicos envolvidos na transcrição, recombinação ou outros (ANTOSCH *et al.*, 2012); genomas de plantas superiores codificariam de 10 a 15 proteínas da família HMG diferentes (STROS *et al.*, 2007).

Percebe-se destes estudos que funções cruciais são desenvolvidas pelos genes e proteínas CoREGs durante o desenvolvimento dos indivíduos. Logo, conhecer estes genes é importante para um melhor entendimento das respostas gênicas dos vegetais, principalmente em situação adversas de estresses ambientais, podendo agregar desempenho e tolerância aos acessos em um programa de melhoramento genético vegetal.

### 3 METODOLOGIA

#### 3.1 ORIGEM DOS DADOS

As plantas utilizadas constaram de dois acessos de *J. curcas* (Jc183 e Jc171; LOZANO-ISLA *et al.*, 2018) que foram cultivadas em estufa (março 2016), no Centro de Ciências Agrárias da Universidade Federal de Alagoas (UFAL, Rio Largo, AL, Brasil), que se encontra em clima classificado (Thorntwaite e Mather, 1955) como úmido, megatérmico, com deficiência moderada de água no verão (dezembro a março) e algum excesso de água no inverno (julho a setembro). O delineamento experimental foi inteiramente casualizado, com dois acessos, dois tratamentos (com e sem sal), e três repetições biológicas para cada condição experimental. Inicialmente, sementes (homogêneas em tamanho e peso) de ambos os acessos foram semeadas em vasos (50 L) contendo 20 kg de areia. Após germinação e os primeiros eófilos (5-10 dias após a germinação - DAG), apenas a planta mais vigorosa foi deixada por vaso. Durante o cultivo, plantas foram irrigadas às 16:00 h a cada três dias com solução nutritiva de Hoagland (20% p/v) (EPSTEIN, 1972). Uma semana antes da aplicação do sal (60 DAG), as plantas receberam solução de Hoagland 100% (força total) todos os dias. No dia anterior à aplicação de sal, plantas foram irrigadas às 16 h. A solução de NaCl (150 mM) foi adicionada à solução de Hoagland, sendo as plantas expostas à solução salina desde às nove a.m. e durante três horas. Plantas irrigadas apenas com a solução de Hoagland constituíram o tratamento controle negativo. Após o tempo de exposição ao NaCl, amostras de raízes foram coletadas, imediatamente congeladas em nitrogênio líquido (N<sub>2</sub>), sendo mantidas a 80°C até a extração do RNA.

O RNA total das amostras de raízes foi isolado usando o Sistema de Isolamento de RNA Total SV (Promega). Estimativas de concentração e da qualidade dos RNAs extraídos foram feitas, respectivamente, por espectrofotometria (Thermo Scientific NanoDrop 2000) e razões de absorvância (DO; 260/280 nm  $\geq$  1,9 e 260/230 nm  $\geq$  1,9), além de visualizações em gel de agarose (1,5% p/v). Amostras de RNAs foram enviadas para o Centro Genômico da Faculdade "Luiz de Queiroz" da Agricultura (ESALQ / USP, Piracicaba, SP, Brasil) e aquelas com RIN (RNA *Integrated Number*)  $\geq$  9,0 (Agilent Bioanalyzer 2100, Santa Clara, CA, EUA) foram utilizadas para confecção das bibliotecas RNA-Seq (12: dois acessos x três repetições biológicas x

dois tratamentos), conforme *TruSeq Stranded mRNA Sample Prep kit* (Illumina Inc, CA, USA, “LS” Protocol). As bibliotecas de RNA-Seq foram sequenciadas (*paired end*, 100 pb, 2x) usando a plataforma Illumina HiSeq2500 (Eurofins MGW, Alemanha).

Pares de *raw* RNA-Seq *Reads* foram avaliadas quanto a qualidade das leituras com o *software* FastQC (v.0.11.5), antes e após filtragem do adaptador e etapas de ajuste (das extremidades pareadas) usando-se parâmetros padrão da ferramenta Trimmomatic (v.0.36; BOLGER *et al.*, 2014). Após excluir adaptadores e bases com baixa qualidade ou indeterminadas, pares de *reads* de alta qualidade (*Phred*  $\geq 30$ , para todas as bases) foram utilizadas na montagem dos transcriptomas usando estratégia *de novo*, a partir do *software* Trinity 2.2.0 (GRABHERR *et al.*, 2011). Os níveis de expressão dos transcritos e de UniGenes foram estimados pelo *software* RSEM (LI & DEWEY, 2011), com auxílio do pacote de alinhamento Bowtie (v4.4.7; LANGMEAD *et al.*, 2009). As matrizes FPKM (fragmento por quilobase de cDNA por milhão de fragmentos mapeados) normalizadas foram geradas a partir das contagens de RSEM, as quais foram usadas para as análises de expressão diferencial realizadas pelo pacote *edgeR* (ROBINSON *et al.*, 2010).

### 3.2 IDENTIFICAÇÃO DE UNIGENES COREGS

Para a identificação de potenciais transcritos codificadores de proteínas CoREGs expressas em raízes de *J. curcas* após exposição ao NaCl foram utilizadas sequências curadas nucleotídicas e proteicas da base de dados *Plant Transcription factor & Protein Kinase Identifier and Classifier*, iTAK (<http://itak.feilab.net/cgi-bin/itak/index.cgi>; ZHENG *et al.*, 2016), oriundas das espécies *Manihot esculenta* (555), *Ricinus communis* (410), e *Arabidopsis thaliana* (329). Transcritos RNASeq de *J. curcas* foram alinhados via BLASTx (*e-value*  $< e^{-10}$ ) contra as sequências proteicas curadas de reguladores transcricionais (CoREGs) da base de dados iTAK.

### 3.3 IDENTIFICAÇÕES DE POTENCIAIS COREGS DIFERENCIALMENTE EXPRESSOS

Potenciais genes CoREGs diferencialmente expressos (DE) foram identificados a partir da aplicação dos limiares *p-value*  $\leq 0,0001$ , FDR (*False Discovery Rate*)  $\leq 0,005$ , e  $\text{Log}_2\text{FC} \geq 1$  (classificado como *induzido*, UR) ou  $\leq -1$  (*reprimido*, DR). A modulação da expressão foi estimada a partir de valores de  $\text{Log}_2\text{FC}$ , sendo FC (*fold*

*change*) > 2, estabelecida como a razão entre as abundâncias (FPKM) do UniGene CoREG na biblioteca estressada em comparação com a correspondente abundância na biblioteca controle negativo. Perfis de expressão de potenciais genes CoREGs em resposta ao estímulo salino foram gerados a partir da clusterização hierárquica feita pelo programa *Cluster 3.0* (v.1.1.4r3; EISEN *et al.*, 1998), e valores de Log<sub>2</sub>FC modulados por ambos os acessos, sendo o *heatmap* visualizado através de imagem no *Java TreeView* (SALDANHA, 2004).

### 3.4 MAPEAMENTO DE COREGS EM GENOMA DE *M. ESCULENTA*

CoREGs totais dos bancos de CoREGs do iTAK de *M. esculenta* foram mapeados no genoma da espécie (obtido do banco de dados *Phytozome*) através de BLASTn (*e-value* < e<sup>-30</sup>). Potenciais CoREGs de *J. curcas* foram analisados quanto suas ocorrências em cromossomos de *M. esculenta* a partir de alinhamento BLASTn (*e-value* < e<sup>-30</sup>). Os dois resultados foram utilizados para fins comparativos.

### 3.5 ÁRVORE FENÉTICA DE FAMÍLIAS COREGS AUX/IAA E JUMONJI

Duas famílias CoREGs bem representadas numericamente (AUX/IAA e Jumonji) foram selecionadas para alinhamento entre CoREGs de *J. curcas* e CoREGs de *M. esculenta* para cada família. Para tanto, transcritos com maiores identidades (≥ 75%) alinhados à CoREGs de *M. esculenta* (BLASTx, *e-value* e<sup>-10</sup>), foram selecionados, totalizando 22 alvos AUX/IAA e 11 alvos Jumonji. As sequências das duas famílias dos CoREGs de *J. curcas* e dos bancos de proteínas das duas famílias CoREGs de *M. esculenta* foram alinhadas separadamente por meio do programa ClustalX 2.0 (LARKIN *et al.*, 2007). Uma árvore fenética foi desenhada utilizando método *Neighbor Joining* (*bootstrap* de 10.000 amostragens) para cada família. As anotações de cada CoREG de *J. curcas* foram confirmadas a partir de BlastP (RefSeq) contra sequências proteicas de *J. curcas* do NCBI.

### 3.6 ANÁLISE DE ONTOLOGIA GÊNICA E DE FATORES DE TRANSCRIÇÃO

O enriquecimento funcional dos termos de ontologia gênica (*Gene Ontology* - GO) identificou aqueles termos significativamente representados (testes exatos de

Fisher,  $p$ -valor  $\leq 0,05$ ) no arquivo de entrada, o qual foi composto, individualmente, por genes modelos de *J. curcas*, correspondentes aos CoREGs DE (UR ou DR) ou CoREGs não DE (ns), tendo por base o melhor alinhamento (*best-hit* em ordem de maior identidade, alinhamento, *score* e *e-value*) após alinhamentos por BLASTn (*e-value*  $< e^{-30}$ ). As análises foram realizadas, em separado, usando a ferramenta *GO Enrichment analysis* ( $p$ -value  $< 0,001$ ) no portal PlantRegMap (Plant Regulator Data and Analysis Platform; <http://plantregmap.cbi.pku.edu.cn>; JIN *et al.*, 2014). Os resultados para cada input file foram visualizados em diagrama de Venn (OLIVEROS, 2007), evidenciando termos GO exclusivos para cada set de genes ou aqueles compartilhados.

### 3.7 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO

De modo similar a metodologia de enriquecimento de termos GO, potenciais FTs relacionando com promotores de genes de *J. curcas* associados aos genes CoREGs DE foram previstos usando-se a ferramenta *TF enrichment analysis* ( $p$ -value  $< 0.001$ ) disponível também no portal PlantRegMap (<http://plantregmap.cbi.pku.edu.cn/>).

### 3.8 VALIDAÇÃO DE COREGS DIFERENCIAL EXPRESSOS VIA RT-QPCR

Pares de primers foram desenhados, tendo por base transcritos RNA-Seq selecionados para ensaios de RT-qPCR, dentre aqueles associados com proteínas CoREGs, a partir da ferramenta online *Primer3* (<http://bioinfo.ut.ee/primer3-0.4.0/>), utilizando-se os seguintes parâmetros: tamanho do amplicon (de 80 a 150 pb); tamanho do primer (de 20 a 24 pb); temperatura de desnaturação do *primer* com mínima 57, máxima de 63 °C e ótima 60 °C; porcentagem de G/C de 45 a 55%, ótimo 50%; temperatura de desnaturação do aplicon (*Amplicon melting*) (65 - 95 °C, ótimo 82 °C); máxima repetição de nucleotídeos: três nucleotídeos; grampo G/C: dois nucleotídeos. *Primers* foram desenhados preferencialmente em junções éxon/éxon ou em éxons diferentes. Para identificação dos éxons dos transcritos utilizou-se a ferramenta online Spleign (<https://www.ncbi.nlm.nih.gov/sutils/spleign>).

Primers propostos (Apêndice A) foram sintetizados pela Bioneer Corporation (Coreia do Sul). A eficiência de cada par de primer nas amplificações RT-qPCR foi estimada a partir de 1  $\mu$ L de cDNA oriundo de diluições seriadas (0,  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ ,

$10^{-4}$ ) obtidas de um bulk composto com parte das amostras de cDNAs, em triplicatas técnicas, sendo que em cada poço de PCR adicionou-se 5  $\mu$ L de Mix (GoTaq® qPCR Master Mix Promega), 3,4  $\mu$ L de água ultra-pura e 0,3  $\mu$ L de cada um dos primers (5  $\mu$ M) do par a ser testado. As reações foram realizadas em real-time PCR Thermocycler LineGene 9600 (Bioer, Hangzhou, China), com as seguintes especificações: desnaturação inicial (95 °C, 2 min., 1 ciclo); PCR Stage (95 °C, 15 s., seguido de 60 °C, 1 min., 40 ciclos); Melting stage (65 - 95 °C, subida de 4 °C.s<sup>-1</sup>).

O *software* LineGene (v.1.1.10) estimou os valores de T<sub>m</sub> e C<sub>q</sub> e as quantificações relativas e absolutas. Para os ensaios de validação foram selecionados apenas os primers com eficiência entre 90 e 110 CT (9). Foram utilizadas placas de 96 poços onde eram testados 3 pares de primers alvos e dois genes de referência [actina (TANG *et al.*, 2016), e  $\beta$ -tubulina (XU *et al.*, 2016) nas configurações de triplicata técnica para cada replica biológica. As reações foram feitas com Kit GoTaq® qPCR Master Mix Promega para RT-qPCR usando, para cada reação, 5  $\mu$ l de mix, 3,4  $\mu$ l de água ultrapura, 0,3  $\mu$ l de cada um primer de seus pares e 1  $\mu$ l de cDNA em concentração 1/10 de cada amostra para seu respectivo poço. As reações foram utilizadas em equipamento real-time PCR *Thermocycler LineGene* 9600 (Bioer, Hangzhou, China). As configurações de corrida para o *software* seguiram da forma: Hold Stage 95 °C – 2" (1 ciclo); PCR Estage 95 °C – 15' seguindo de 60 °C – 1" (45 ciclos); Melting Stage 65 -> 95° C (subida de 4 °C por segundo). O *software* LineGene (v.1.1.10) estimou os valores de T<sub>m</sub> e C<sub>q</sub>. Os dados de expressão relativa foram estimados usando o *software* REST 2009 (v.2.0.13; PFAFFL *et al.*, 2002), aplicando teste de randomização com 2.000 permutações e testando a hipótese de diferenças significativas entre os grupos controle e tratamento.

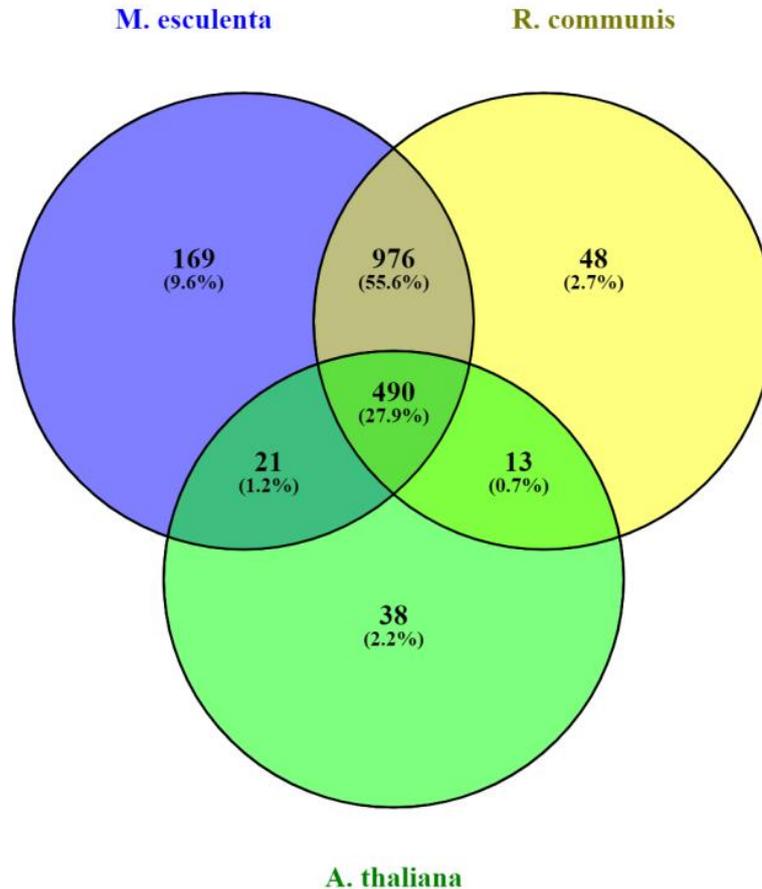
## 4 RESULTADOS

### 4.1 IDENTIFICAÇÃO DE UNIGENES COREGS EM *J. curcas*

A montagem do transcriptoma proporcionou 145,422 transcritos montados (101 Mb) / 126,343 UniGenes (77 Mb). As análises BLASTx ( $e\text{-value} < e^{-10}$ ), feitas de forma independente com sequencias proteicas de coreguladores (CoREGs) das espécies *M. esculenta* e *R. communis* (base de dados iTAK), e *A. thaliana* (base de dados PLNTFDB, Plant Transcription Factor Database; <http://plntfdb.bio.uni-potsdam.de/v3.0/>), relacionaram 2574 transcritos montados (1755 UniGenes) com 1250 proteínas CoREGs.

Individualmente foram 2418 transcritos (1656 UniGenes) associados com 554 proteínas CoREGs de *M. esculenta*; 2261 transcritos (1527 UniGenes) com 359 CoREGs de *R. communis*, e 857 transcritos (562 UniGenes) com 337 de CoREGs de *A. thaliana*. A participação de cada espécie na identificação dos Unigenes codificadores de CoREGs em *J. curcas* é evidenciada no diagrama de Venn da Figura 7, que indica que 490 UniGenes foram identificados em comum a partir das três espécies, 976 em comum entre *M. esculenta* e *R. communis*, 21 entre *M. esculenta* e *A. thaliana*, 13 entre *R. communis* e *A. thaliana*, e, individualmente, 169 a partir de *M. esculenta*, 48 a partir de *R. communis*, e 38 a partir de *A. thaliana*.

**Figura 7.** UniGenes CoREGs de *J. curcas* identificados a partir de similaridades em análise BLASTx ( $e\text{-value} < e^{-10}$ ) com proteínas das espécies *M. esculenta*, *R. communis* e *A. thaliana*. Distribuição de 1755 UniGenes considerando contribuição exclusiva ou compartilhada.



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

A distribuição relativa em famílias CoREGs, considerando cada transcrito CoREG e o respectivo melhor alinhamento de proteína CoREG, em BLASTx, é dada na Tabela 1. Com exceção do grupo Others (outros: agrupamento de potenciais CoREGs que ainda não pertencem a nenhuma das famílias relatadas na literatura mas que futuramente serão categorizados em diferentes famílias, a medida que as caracterizações de domínios conservados e estudos de genômica estrutural e filogenética avançarem em diferentes espécies), as famílias mais representadas em UniGenes foram Jumonji (572), AUX/IAA (349), e PHD (84), após ter sido desconsiderado o grupamento Others.

Entretanto, dentre os 1755 UniGenes CoREGs, aqueles declarados DE após o estímulo salino na resposta de Jc171 corresponderam a 239 UniGenes (24 UR e 215 DR; apêndice B). Estes quando distribuídos em famílias CoREGs evidenciaram as famílias AUX/IAA (72), Jumonji (54), e PHD (29), como as mais representadas em UniGenes DE (Tabela 5). A resposta Jc183 ao sal, por outro lado, somente evidenciou dois UniGenes DE [DN26690\_c0\_g2 (UR) e DN32228\_c0\_g2 (DR)].

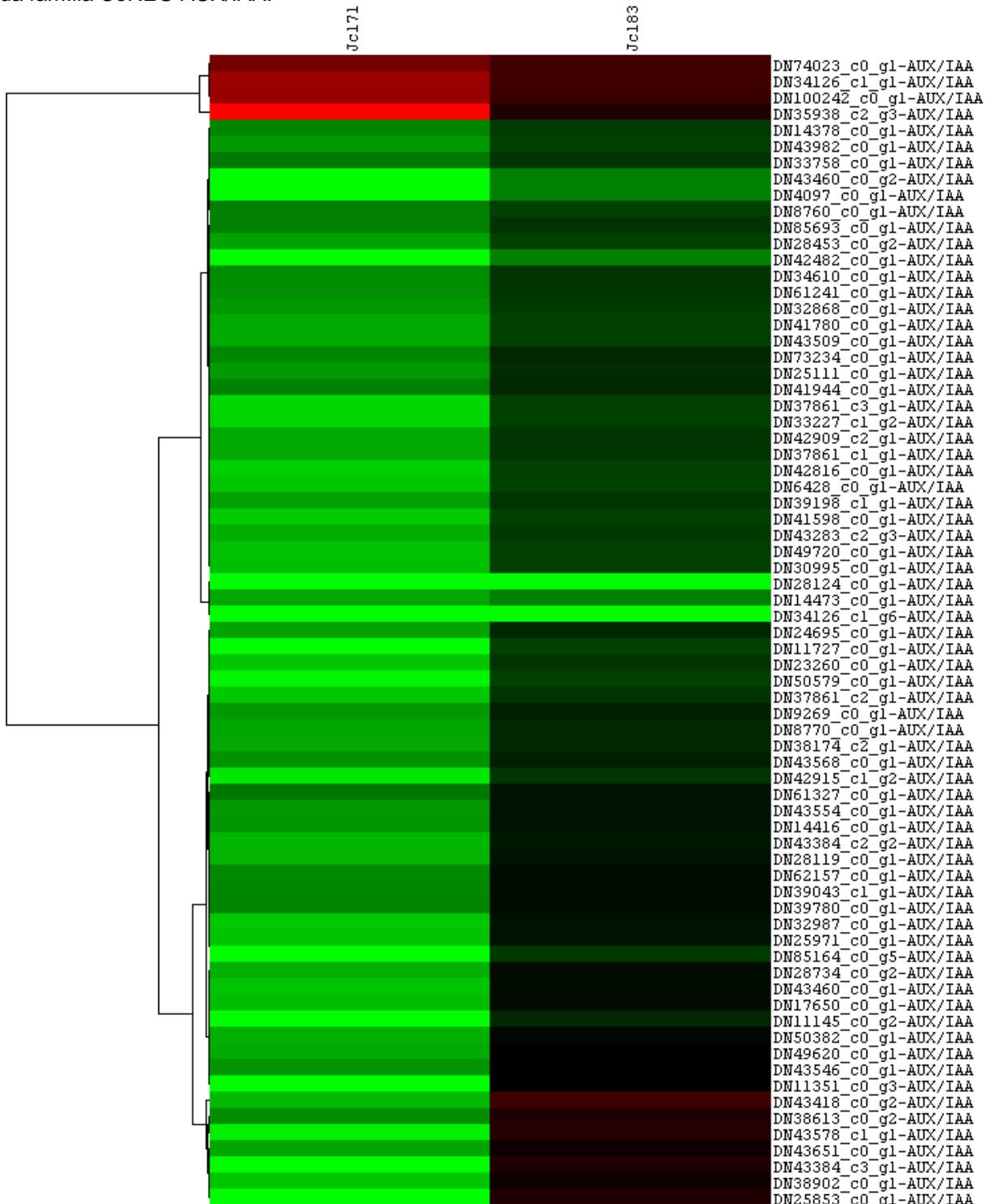
**Tabela 5.** Distribuição ordenada das Famílias CoREGs de *Jatropha curcas* de acordo com a representatividade em UniGenes, identificados em alinhamentos BLASTx (*best-hits*) com proteínas de cada família, e número de UniGenes DE codificando CoREGs em cada família entre parenteses

<b>Família</b>	<b>Unigenes CoREGs (DE)</b>	<b>UR</b>	<b>DR</b>	<b>Proteínas CoREGs alinhadas</b>
Jumonji	572 (55)	4	51	51
AUX/IAA	349 (71)	4	67	93
Others	198 (20)	8	12	263
PHD	84 (29)	0	29	137
MTERF	78 (7)	0	7	84
SET	67 (5)	1	4	129
LUG	61 (8)	1	7	18
GNAT	60 (3)	1	2	103
HMG	54 (0)	0	0	48
TRAF	48 (8)	0	8	48
SNF2	46 (14)	0	14	85
IWS1	31 (3)	0	3	25
ARID	22 (3)	0	3	29
SWI/SNF-BAF60b	21 (4)	2	2	57
MBF1	17 (3)	3	0	8
Pseudo ARR-B	15 (1)	0	1	16
SWI/SNF-SWI3	8 (2)	0	2	9
TAZ	7 (2)	0	2	12
Coactivator P15	6 (0)	0	0	8
RCD1	4 (0)	0	0	10
RB	3 (1)	0	1	6
SOH1	2 (0)	0	0	2
MED6	1 (0)	0	0	5
MED7	1 (0)	0	0	2
<b>Total</b>	<b>1755 (239)</b>	<b>24</b>	<b>215</b>	<b>1248</b>

## 4.2 PERFIL DE EXPRESSÃO DE CoREGs EM *J. curcas* SOB TRATAMENTO SALINO

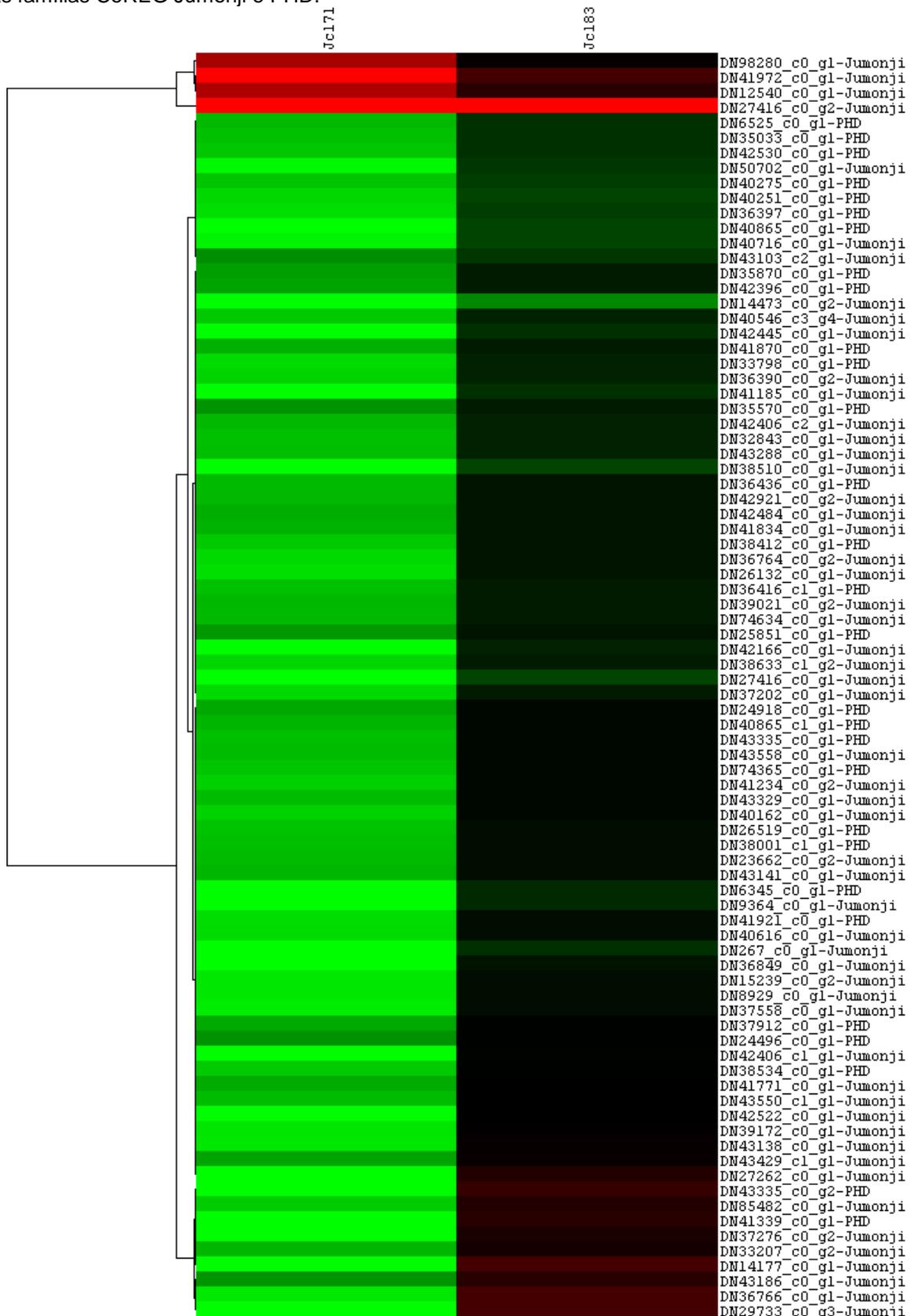
Considerando os valores ( $\text{Log}_2\text{FC}$ ) de modulação da expressão gênica de UniGenes CoREG DE do acesso Jc171, e os valores correspondentes (n.s.) na expressão de Jc183, a co-expressão de Unigenes é ilustrada nas Figuras 10, 11 e 12. De modo geral, a clusterização hierárquica para UniGenes DE representantes da família AUX/IAA (Figura 8), Jumonji e PHD (Figura 9), e demais famílias (Figura 10) apresentaram dois grupos de co-expressão, sendo um deles numericamente menor e composto de Unigenes UR da resposta Jc171, e outro bem maior, com Unigenes DR da resposta Jc171. Uma maior modulação por parte dos Unigenes DE da resposta de Jc171 é evidente, quando comparada com a modulação menos intensa dos mesmos Unigenes, na resposta Jc183. Apesar de não possuir genes DE nas análises de expressão de Jc183, é possível observar tendências a um tipo de expressão UR ou DR em alguns transcritos. Além disto, há a presença de alguns grupos de transcritos contrastantes entre ambos os acessos que podem ser observados nas Figuras 10 e 11. Nestes casos o acesso Jc183 demonstra a tendência a ter os transcritos destes grupos regulados de forma UR, enquanto o contrário é mostrado para Jc171, onde estes grupos de transcritos são DEs DR.

**Figura 8.** Heatmaps baseados na clusterização hierárquica da co-expressão de UniGenes CoREGs DE [p-value: 0.0001; FDR: 0.005;  $\text{Log}_2\text{FC} > 1$  (UR) ou  $< 1$  (DR)] da resposta ao sal do acesso Jc171, em comparação aos correspondentes valores ( $\text{Log}_2\text{FC}$ ) modulados pelo acesso Jc183 para representantes da família CoREG AUX/IAA.



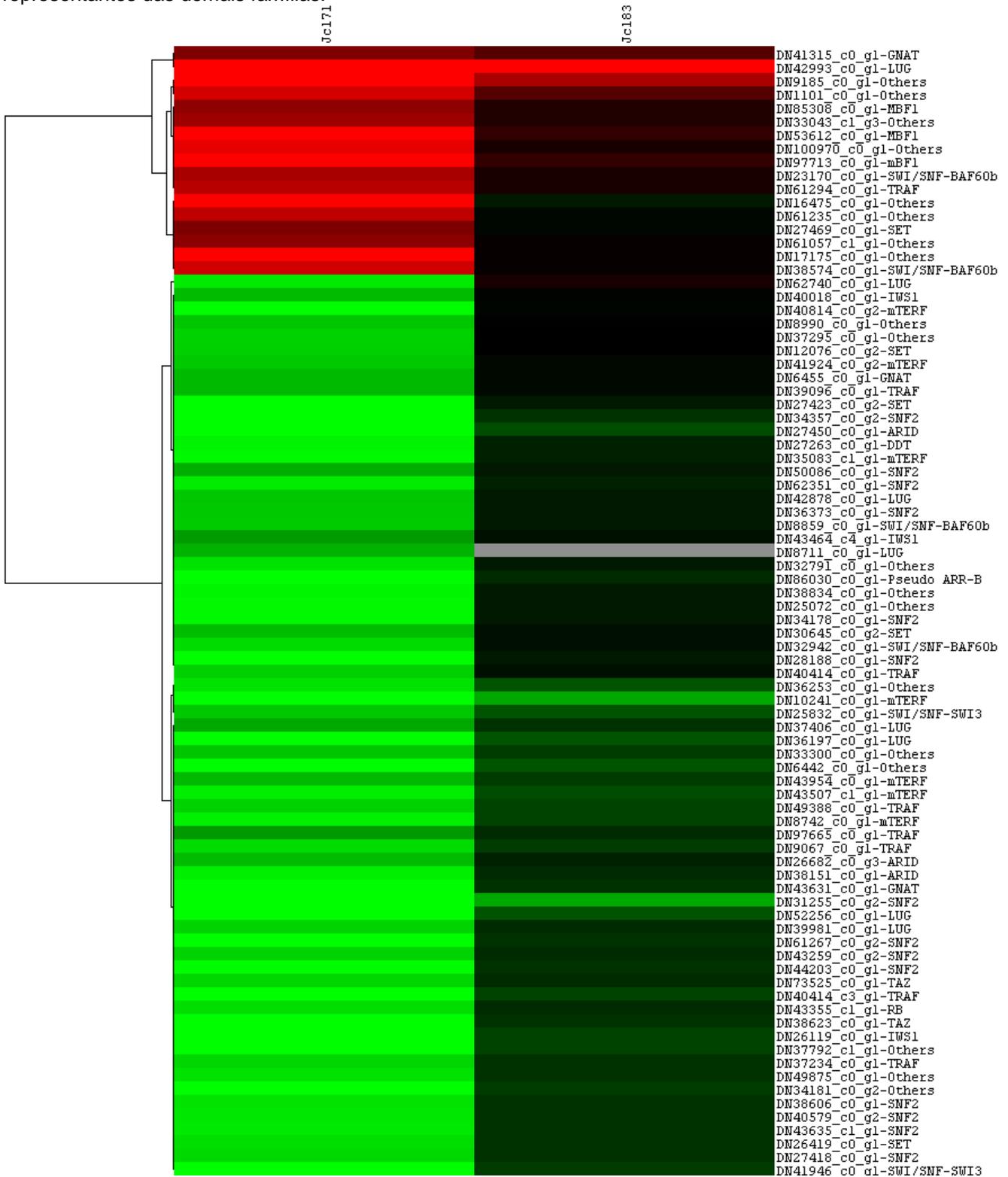
Fonte: Jorge Luís Bandeira da Silva Filho (2019)

**Figura 9.** Heatmaps baseados na clusterização hierárquica da co-expressão de UniGenes CoREGs DE [p-value: 0.0001; FDR: 0.005;  $\text{Log}_2\text{FC} > 1$  (UR) ou  $< 1$  (DR)] da resposta ao sal do acesso Jc171, em comparação aos correspondentes valores ( $\text{Log}_2\text{FC}$ ) modulados pelo acesso Jc183 para representantes das famílias CoREG Jumonji e PHD.



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

**Figura 10.** Heatmaps baseados na clusterização hierárquica da co-expressão de UniGenes CoREGs DE [p-value: 0.0001; FDR: 0.005;  $\text{Log}_2\text{FC} > 1$  (UR) ou  $< 1$  (DR)] da resposta ao sal do acesso Jc171, em comparação aos correspondentes valores ( $\text{Log}_2\text{FC}$ ) modulados pelo acesso Jc183 para representantes das demais famílias.



. Fonte: Jorge Luís Bandeira da Silva Filho (2019)

#### 4.3 MAPEAMENTO DE COREGS EM GENOMA DE *M. ESCULENTA*

O mapeamento de 555 CoREGs de *M. esculenta* nos cromossomos da espécie identificaram 905 locos distribuídos nos 18 cromossomos, ao passo que o mapeamento a partir de 1755 CoREGs de *J. curcas* nos 18 cromossomos de *M. esculenta*, associou, com base nas similaridades a 416 locos (Tabela 6). Desta forma, cerca de 45,97% dos locos previstos para CoREGs de *M. esculenta* no genoma da espécie foram considerados conservados em similaridades em relação aos CoREGs de *J. curcas*. Essa representatividade em cada cromossomo de *M. esculenta* é vista na Tabela 6.

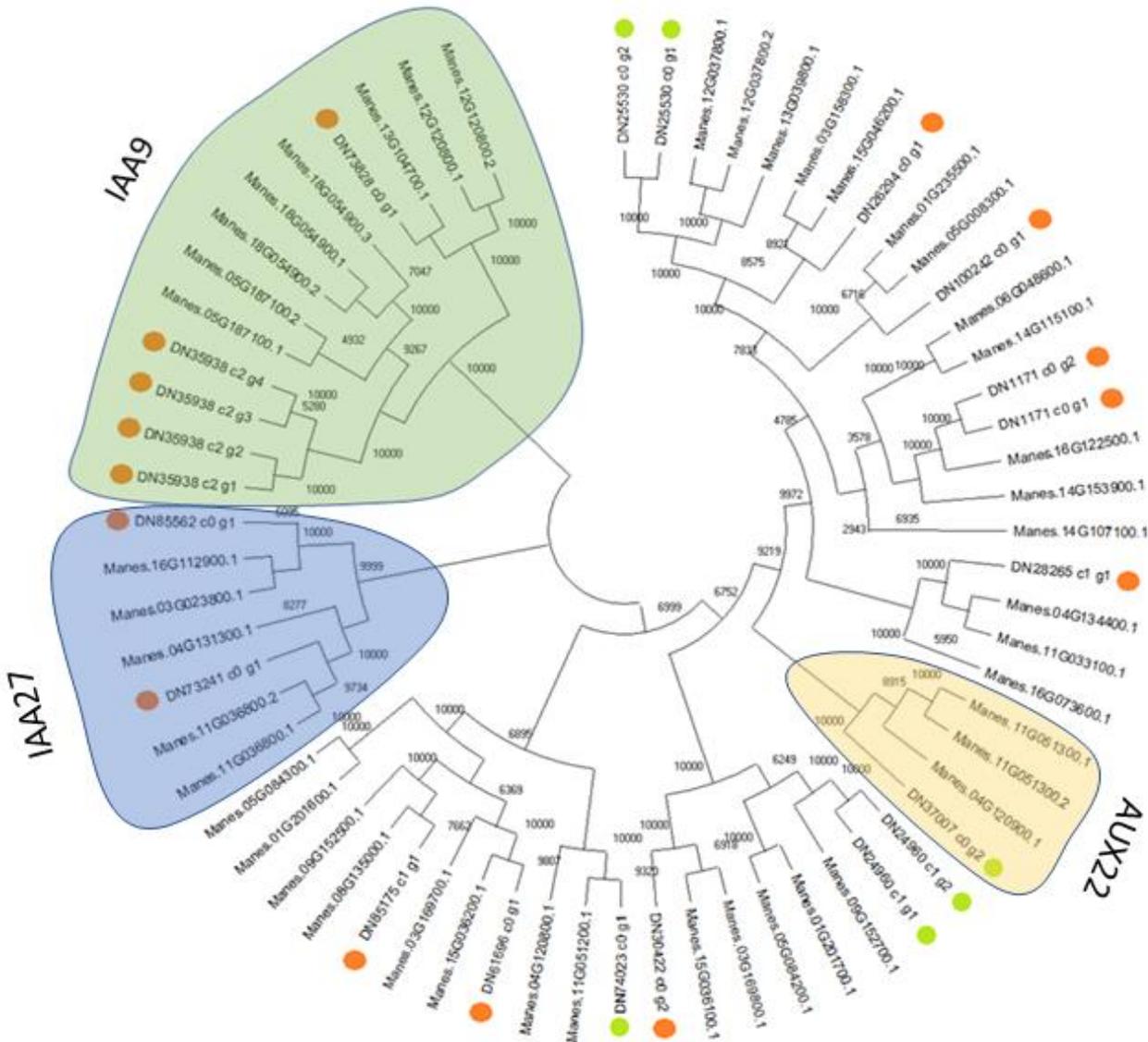
**Tabela 6.** Número de CoREGs de *Manihot esculenta* e de similares de *Jatropha curcas* estimados em cada um dos 18 cromossomo de *M. esculenta*, bem como os respectivos percentuais desta representatividade (razão entre *M. esculenta* / *J. curcas*).

Número do cromossomo	CoREGs detectados a partir de <i>M. esculenta</i>	CoREGs detectados a partir de <i>J. curcas</i>	Mapeamento a partir de <i>J. curcas</i> (%)
Cromossomo 01	73	39	53.4
Cromossomo 02	41	20	48.8
Cromossomo 03	90	44	48.9
Cromossomo 04	45	20	44.4
Cromossomo 05	63	30	47.6
Cromossomo 06	30	12	40.0
Cromossomo 07	34	10	29.4
Cromossomo 08	49	25	51.0
Cromossomo 09	49	23	46.9
Cromossomo 10	38	13	34.2
Cromossomo 11	49	19	38.8
Cromossomo 12	53	22	41.5
Cromossomo 13	46	27	58.7
Cromossomo 14	48	21	43.8
Cromossomo 15	82	43	52.4
Cromossomo 16	48	23	47.9
Cromossomo 17	35	11	31.4
Cromossomo 18	32	14	43.8

#### 4.4 ÁRVORES FENÉTICAS DAS FAMÍLIAS AUX/IAA E JUMONJI

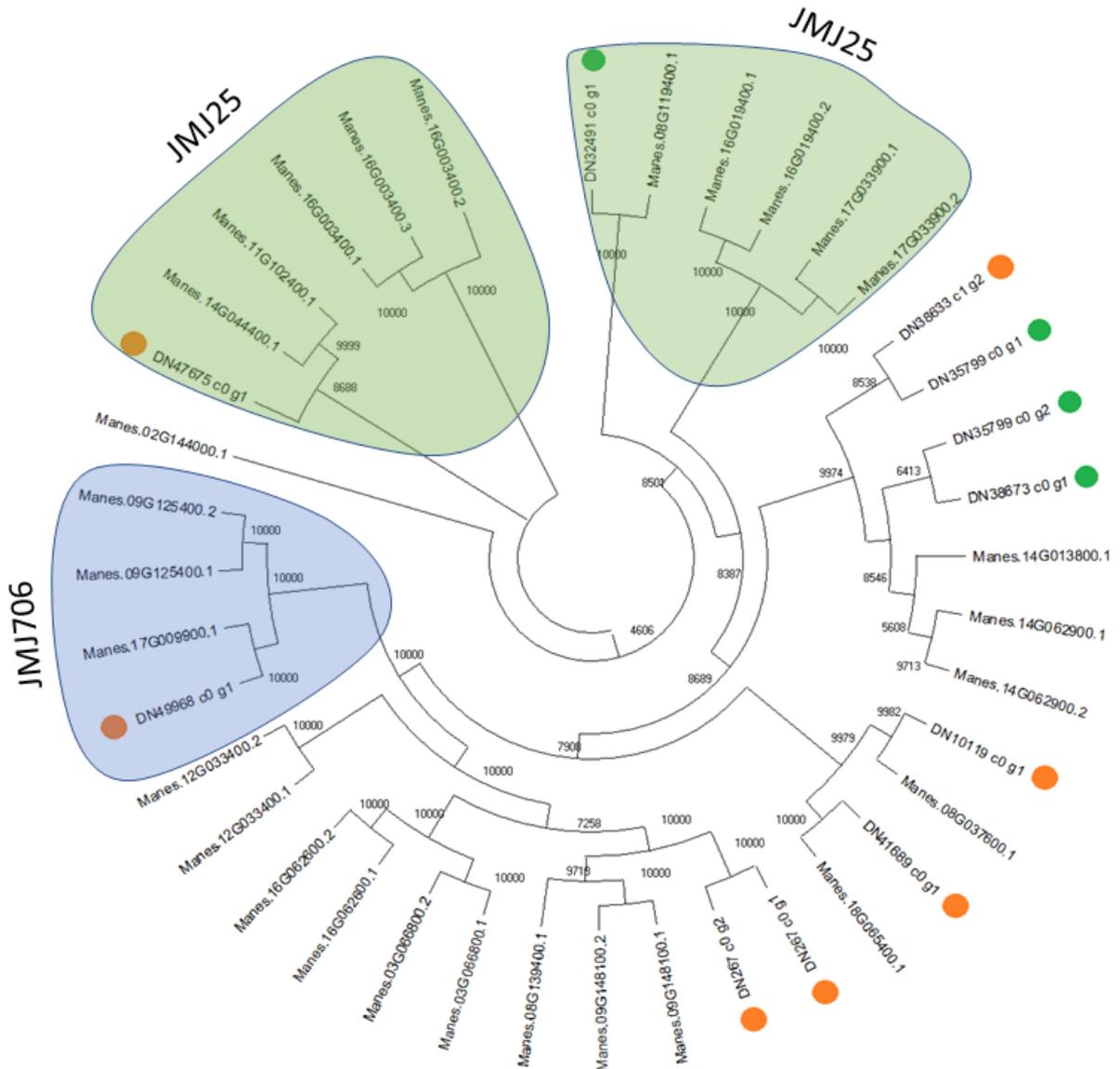
As árvores fenéticas para as famílias AUX/IAA e Jumonji ilustram as similaridades que as proteínas dos potenciais transcritos CoREGs de *J. curcas* demonstraram com as demais proteínas do banco de CoREGs para cada uma das duas famílias analisadas de *M. esculenta*. Além disto, proteínas AUX/IAA identificadas a partir de *R. communis* mostraram-se distribuídas também compartilhando similaridades com proteínas de *M. esculenta*. Na figura 11 (AUX/IAA) há a presença de oito cromossomos de *M. esculenta* envolvidos (3, 4, 5, 8, 11, 12, 15 e 16) dos quais o cromossomo cinco apresenta a maior quantidade de representantes (6). A identificação de cada cromossomo está descrita ao lado esquerdo da letra “G” na identificação de cada ramo terminal. Seis proteínas AUX/IAA foram identificadas a partir de *R. communis*. Três grupos de proteínas AUX/IAA foram identificados bem estabelecidos em grupamentos na árvore (IAA9, IAA27 e AUX22). Na figura 12 (Jumonji), é visto que também há a presença de 5 cromossomos envolvidos (8, 14, 16, 17 e 18) tendo o cromossomo 8 a maior quantidade de representantes (3) e quatro proteínas Jumonji identificadas a partir de *R. communis*. Dois grupos de proteínas Jumonji foram identificados bem estabelecidos na árvore (JMJ25 e JMJ706)

**Figura 11.** Árvore fenética de proteínas CoREGs da família AUX/IAA (sequências  $\geq 75\%$  de identidade) de *J. curcas* e de *M. esculenta*. CoREGs identificados por iniciais: “DN”, oriundos de *J. curcas*; “Manes”, de *M. esculenta*. Grupos de proteínas: IAA9, em palheta verde; IAA27, em palheta azul; AUX22, em palheta laranja. CoREGs de *J. curcas* identificados a partir de: *R. communis*, esferas em verde; *M. esculenta*, esferas em laranja. Método *Neighbor Joining*, com *bootstrap* de 10.000 amostragens (*software* ClustalX 2.0).



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

**Figura 12.** Árvore fenética de proteínas CoREGs da família Jumonji (sequências  $\geq 75\%$  de identidade) de *J. curcas* e de *M. esculenta*. CoREGs identificados por iniciais: “DN”, oriundos de *J. curcas*; “Manes”, de *M. esculenta*. Grupos de proteínas: JM25, em palheta verde, e JM706, em palheta azul. CoREGs de *J. curcas* identificados a partir de: *R. communis*, esferas em verde; *M. esculenta*, esferas em laranja. Método *Neighbor Joining*, com *bootstrap* de 10.000 amostragens (*software* ClustalX 2.0).



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

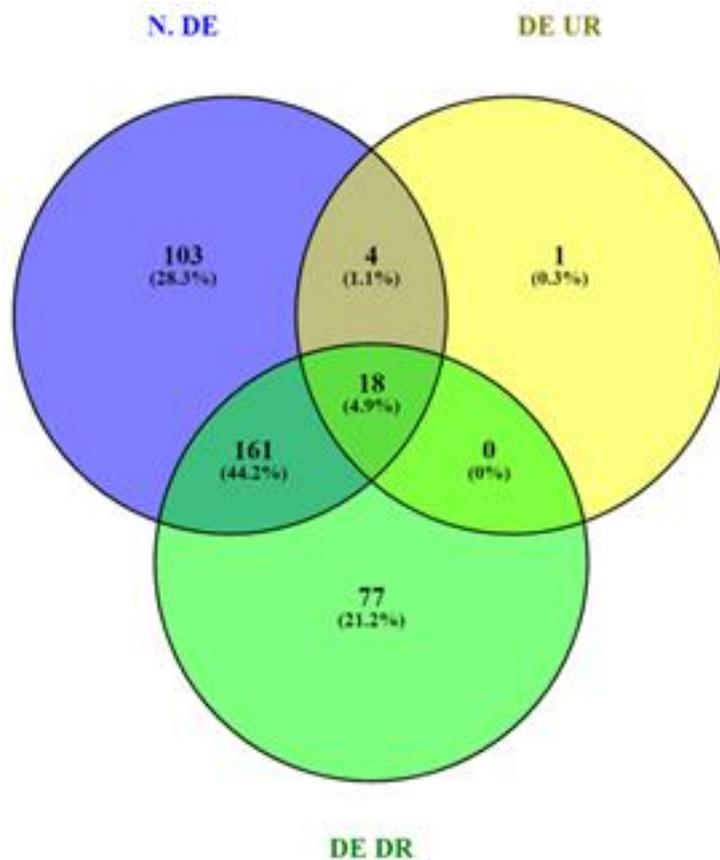
#### 4.5 ANÁLISE DE ENRIQUECIMENTO DE TERMOS DE ONTOLOGIA GÊNICA

As análises de enriquecimento dos termos GO, para cada conjunto, em separado, de UniGenes codificadores de CoREGs (UR, DR ou não DE), revelaram os termos mais representados (enriquecidos) de cada conjunto da resposta Jc171 ao sal. Esses resultados, quando comparados em um diagrama de Venn (Figura 13)

indicaram, em associação aos UniGenes UR, um termo exclusivo (GO: GO:0060090, *binding, bridging*) e quatro compartilhados [com UniGenes CoREG não DE: GO:0003712 (*transcription cofactor activity*); GO:0000989 (*transcription factor activity, transcription factor binding*); GO:0000988 (*transcription factor activity, protein binding*); GO:0003713 (*transcription coactivator activity*)].

Caracterizando os UniGenes CoREG DR, foram associados, de forma exclusiva, 77 termos GO enriquecidos, evidenciando respostas diferentes das encontradas para os Unigenes UR (Apêndice C). Por sua vez, outros 18 termos enriquecidos foram comuns aos três conjuntos de UniGenes codificadores de CoREGs.

**Figura 13.** Diagrama de Venn ilustrando termos GO enriquecidos (*Fisher test, p-value 0.001*) exclusivos ou comuns aos UniGenes codificadores de CoREGs diferencialmente expressos (UR, induzido; DR, reprimido) ou não diferencialmente expressos (N.DE.). DE: p-value: 0.0001; FDR: 0.005; Log<sub>2</sub>FC > 1 (UR) ou < -1 (DR).



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

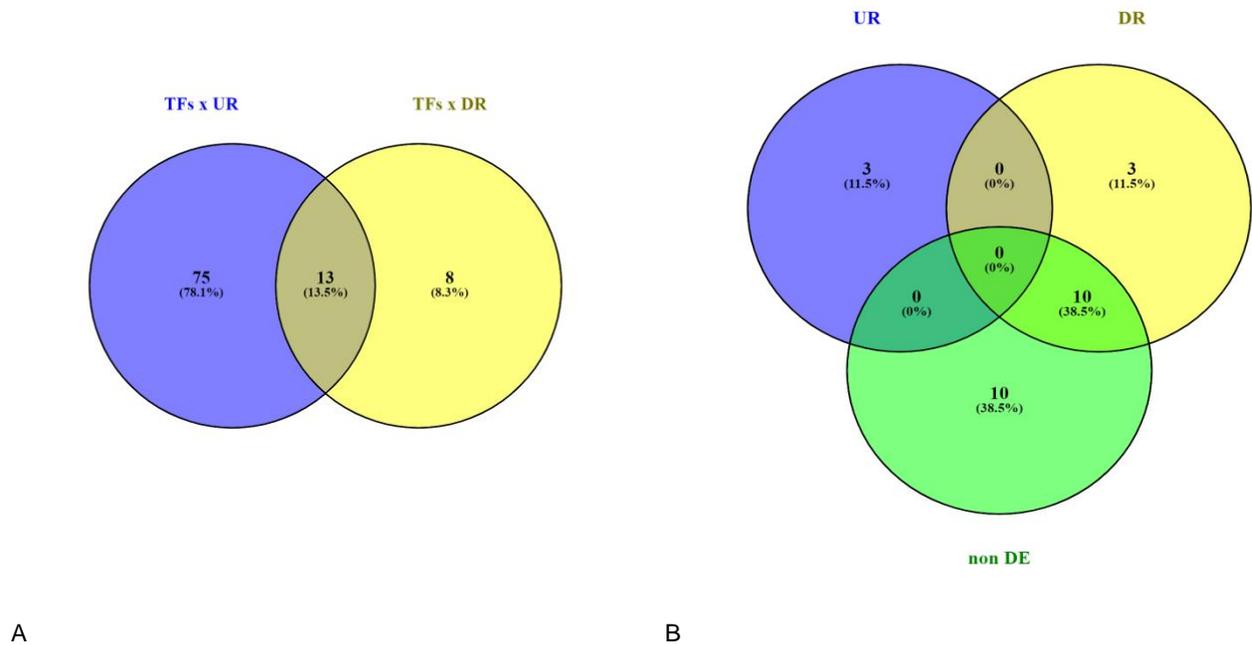
#### 4.6 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO

A predição de fatores de transcrição (FTs) com potencial para interagir com elementos regulatórios em promotores de genes modelo de *J. curcas* envolvendo todos os Unigenes DE da resposta Jc171 foi executada. Os genes modelo de *J. curcas* correspondentes aos Unigenes UR ou DR foram selecionados de alinhamentos BLASTx ( $e\text{-value} < e^{-10}$ ), em que se escolheu o melhor alinhamento de cada Unigene (*best-hit*: ordenando a melhor identidade, alinhamento, *score* e *e-value* respectivamente), do conjunto de proteínas relativas aos genes modelo de *J. curcas*, da base de dados PlantRegMap. Assim, dos 4141 UniGenes declarados DE, 3496 (foram associados a 3024 genes modelo *J. curcas*, sendo 2049 UniGenes UR associados com 1667 genes modelo, e 1447 UniGenes DR com 1355 genes modelo.

A análise com genes modelo associados aos Unigenes UR previu 18899 interações (regulações), a partir de 302 FTs, sendo 88 aqueles considerados enriquecidos ou sobre-representados no conjunto de 1660 alvos pesquisados (Apêndice D). A análise considerando os Unigenes DR previu 13144 interações envolvendo 302 FTs, dos quais 21 foram declarados enriquecidos considerando 1283 input genes (Apêndice E).

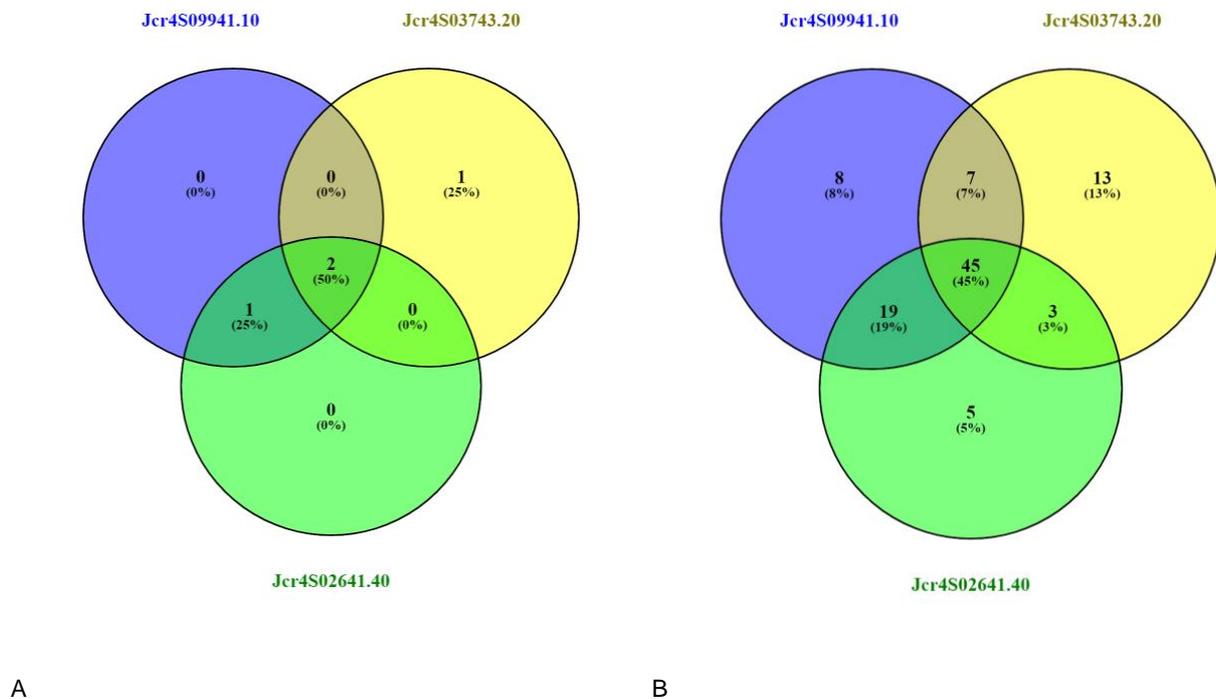
O diagrama de Venn com os FTs enriquecidos identificados em cada análise mostrou que 75 deles regulariam alvos exclusivamente UR, oito deles exclusivamente alvos DR, e 13 deles alvos comuns aos dois conjuntos (Figura 14). Por sua vez, a análise de enriquecimento de FTs associados aos 24 CoREGs UR (21 *input genes*) identificou 129 FTs (254 regulações), dos quais três FTs enriquecidos, todos da família bZIP (Jcr4S02641.40, Jcr4S09941.10, Jcr4S03743.20), que regulariam ao menos três input genes cada, sendo dois deles [Jcr4S00062.70 (AUX/IAA) e Jcr4S00921.20 (MBF1)] alvos em comum aos três FTs (Figura 15A). Esses três FTs enriquecidos foram identificados igualmente enriquecidos ( $p\text{-value} < 0.001$ ) regulando outros UR genes [Jcr4S02641.40 (72 alvos), Jcr4S09941.10 (79 alvos), Jcr4S03743.20 (68 alvos)]. Esses resultados em diagrama de Venn (Figura 15B)

**Figura 14.** Análise de enriquecimento de FTs. A) FTs enriquecidos previstos na interação com os conjuntos de genes alvos UR (*Induzido*) e DR (*Reprimido*); B) FTs enriquecidos previstos na interação com os conjuntos de genes alvos CoREGs com expressão UR, DR ou não DE (non DE).



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

**Figura 15.** FTs enriquecidos, da família bZIP (Jcr4S09941.10, Jcr4S03743.20, Jcr4S02641.40): em (A), influenciando cada um dos três alvos CoREGs (input genes PlantRegMap), sendo dois deles alvos em comum aos três FTs [Jcr4S00062.70 (AUX/IAA) e Jcr4S00921.20 (MBF1)]; em (B), quantidade de transcritos DE UR (*Induzido*) de Jc171 envolvidos com os três TFs em questão.



Fonte: Jorge Luís Bandeira da Silva Filho (2019)

Já a análise dos 214 CoREGs DR (196 *input genes*) previu 260 FTs (1978 regulações), sendo 13 enriquecidos, estes apresentaram-se envolvidos com um número mínimo de 9 e máximo de 80 alvos (Apêndice F). Já a análise com os CoREGs não DE (906 *input genes*) previu 302 FTs (9244 regulações), dos quais 20 foram declarados enriquecidos envolvendo-se com um número mínimo de 10 e máximo de 305 alvos (Apêndice G).

#### 4.7 VALIDAÇÃO DE COREGS DIFERENCIAL EXPRESSOS VIA RT-QPCR

Oito Unigenes CoREG declarados DE na resposta Jc171, de diferentes famílias e regulações, foram selecionados para validação da expressão gênica via RT-qPCR (Tabela 7). Os pares de primers propostos apresentaram valores de eficiência aceitáveis em reações absolutas de RT-qPCR ao amplificar amostras de cDNAs oriundas de diluições seriadas contemplando os tratamentos em análise.

**Tabela 7.** Lista de eficiência baseada em CT (*circle of treshold*) de cada par de *primers* testado. Valores obtidos através de *software LineGene* (v.1.1.10)

Alvo	Eficiência	Intercept	Slope
SET	91,52	32,96	-3,54
Jumonji1	109,53	35,94	-3,11
SNF2a	92,96	33,81	-3,51
MBF1	98,60	30,76	-3,36
GNAT	109,38	32,58	-3,12
SNF2b	109,25	33,45	-3,12
Jumonji2	91,64	33,83	-3,54
mTERF	92,86	35,71	-3,51

De acordo com os resultados, dos oito alvos analisados, cinco confirmaram as expressões DR observadas nas análises *in silico*, enquanto três não confirmaram as expressões (MBF1, GNAT e mTERF; Tabela 8), sendo consideradas de expressão não significativa (n.s.).

**Tabela 8.** Expressão relativa de genes CoREGs para Jc171 tratada com estresse salino (150 Mm 3 h). \* Genes de referência Actina (ACT) e b-Tubulina (bTub). Valores obtidos através da ferramenta REST

Gene	Tipo	Eficiência de reação	Expressão	Erro padrão	95% C.I.	P(H1)	Resultado
bTub	REF	1,000	0,362				
ACT	REF	0,902	2,762				
SET	ALVO	0,915	0,307	0,104 - 1,028	0,026 - 2,434	0,014	DR
Jumonji1	ALVO	1,000	0,194	0,042 - 0,994	0,002 - 5,187	0,018	DR
SNF2a	ALVO	0,927	0,155	0,035 - 0,601	0,008 - 1,027	0	DR
bTub	REF	1,000	0,729				
ACT	REF	0,902	1,371				
SNF2b	ALVO	1,000	0,330	0,101 - 0,947	0,056 - 3,151	0,016	DR
Jumonji2	ALVO	0,916	0,460	0,182 - 1,038	0,111 - 2,711	0,026	DR
bTub	REF	1,000	0,668				
ACT	REF	0,902	1,497				
MBF1	ALVO	0,986	1,290	0,534 - 2,548	0,344 - 15,731	0,600	n.s.
Mterf	ALVO	0,929	0,499	0,159 - 1,882	0,042 - 6,878	0,152	n.s.
GNAT	ALVO	1,000	1,205	0,582 - 2,545	0,244 - 14,781	0,716	n.s.

## 5 DISCUSSÃO

### 5.1 IDENTIFICAÇÃO DE UNIGENES COREGS

Este trabalho apresenta o primeiro conjunto de CoREGS para *J. curcas* e suas respostas transcricionais ao estresse salino. A partir da identificação por similaridade de CoREGS de espécies próximas e *A. thaliana* foi possível observar que, como esperado, houve uma maior quantidade de CoREGs de *J. curcas* identificada a partir das espécies mais próximas (*M. esculenta* e *R. communis*). *A. thaliana* compartilhou uma quantidade relevante unigenes CoREGs identificados com as três espécies.

A maior quantidade de CoREGS identificados foi comum entre *R. communis* e *M. esculenta*, tendo estas duas espécies respondido por 97,8% dos unigenes CoREGs para *J. curcas*. Uma grande parte dos CoREGs identificados foi comum entre as três espécies (27,9%). Apesar da distância filogenética, muitos CoREGS apresentam domínios altamente conservados (LUO *et al.*, 2018; TSUDA *et al.*, 2004; MLYNAROVA *et al.*, 2007), o que pode explicar uma quantidade elevada de sequências similares compartilhadas com *A. thaliana*. Além disto, a partir de *A. thaliana* foi possível identificar 38 (2,1%) Unigenes CoREGs que não eram comum entre as outras duas espécies, adicionando mais informação ao CoREGoma de *J. curcas*. *M. esculenta* foi a espécie mais efetiva na identificação de CoREGs para *J. curcas* com 1656 CoREGs identificados.

Ao analisar as famílias CoREGS presentes nas bibliotecas de *J. curcas*, detectou-se que as famílias predominantes foram Jumonji, AUX/IAA e PHD. Estas famílias serão mais bem discutidas nos próximos tópicos.

### 5.2 PERFIL DE EXPRESSÃO DE CoREGs EM *J. curcas* SOB TRATAMENTO SALINO

Na análise de expressão gênica, apesar do acesso Jc183 não ter demonstrado diferenças significativas em expressão, tanto quanto Jc171 (acesso sensível), que precisou modular mais fortemente seu transcriptoma, inclusive os genes CoREGs, houve uma tendência de semelhança na expressão de ambos os acessos (Figuras 8, 9 e 10). A co-expressão gênica em ambos os acessos (apesar de Jc183 demonstrar tendências de expressão e não transcritos diferencialmente expressos) indica que

grupos de CoREGs regulados de forma semelhante, tanto em Jc171 quanto em Jc183, podem apresentar funções importantes para resposta ao estímulo em salino.

A modulação mais intensa de Jc171 foi reflexo de uma expressão mais reprimida, em relação aos CoREGs, com exceção de alguns genes normalmente envolvidos em respostas a estresses, incluindo aqueles da família AUX/IAA (JUNG *et al.*, 2017; GHANASHYAN & JAIN, 2009; RYU & CHO 2015). Assim, transcritos de Jc171 relacionados com a modulação da cromatina foram reprimidos, em grande parte. Esta resposta pode ser necessária para que a planta direcione o gasto energético do aparato transcricional para outros genes relevantes na resposta ao estresse, ou para que genes importantes permaneçam ativados ou silenciados na resposta ao estímulo salino.

O estresse salino, assim como o estresse osmótico, afeta primariamente o sistema fotossintético e o sistema de regulação basal dos vegetais, diminuindo o aporte de CO<sub>2</sub> e comprometendo o crescimento das plantas (CHAVES, 1991; MUNNS *et al.*, 2006). Segundo Ouyang *et al.* (2007), a repressão de variados genes de *Solanum lycopersicum* sob estresse salino (300 mM NaCl) após 6 h indicou o colapso do vegetal sob estresse, principalmente de genes relacionados ao metabolismo do Nitrogênio. Assim, Jc171 também poderia ter sido afetado de forma demasiada, a ponto de comprometer sua maquinaria genética, sugerindo que parte de seus genes não puderam ser expressos de uma forma positiva.

Algumas particularidades relativas às famílias CoREGs AUX/IAA, Jumonji e PHD, as três que se mais destacaram em quantidade de transcritos DE, além de SNF2 e SET, duas famílias basicamente de remodeladores de cromatina, com exemplos de expressão validadas em ensaios de RT-qPCR, são fornecidas abaixo.

### **5.2.1 Família AUX/IAA**

A maior família CoREG em quantidade de transcritos DE, no presente estudo, foi AUX/IAA, a maior parte deles DR, havendo relação com a concentração de auxina. A auxina desempenha papel crucial nas respostas celulares e no desenvolvimento das plantas. Estas podem detectar e responder rapidamente às mudanças nos níveis de auxina. As respostas envolvem várias classes principais de genes responsivos como a família Auxina/Ácido Indol-3-Acético (AUX/IAA), cujas proteínas normalmente agem para manter os genes silenciados, ao interagir com

proteínas FTs do tipo ARFs (*Auxin Response Factor*), as quais podem se ligar diretamente aos genes (LAVY *et al.*, 2016).

Visto que o estresse salino pode remodelar a arquitetura das raízes, e alterar o acúmulo e a redistribuição de auxinas, esta redistribuição de auxinas em determinados tecidos pode propiciar um menor crescimento vegetal. Em plantas, em zonas teciduais com baixa concentração de auxinas se desencadeia uma resposta AUX/IAA, a qual envolve o silenciamento de diversos genes, através de interação com proteínas FTs ARF. Em tecidos com maior concentração de auxina, se permite a interação de proteínas AUX/IAA com TIR1/ABF, uma proteína que conduz a degradação de AUX/IAA pelos proteossomas (WANG *et al.*, 2015). Desta forma, sem a interação de proteínas AUX/IAA, os FTs ARF podem desenvolver suas atividades de ativação ou repressão (DODD & PÉREZ-ALFOCEA, 2012; RYU E CHO 2015; LAVY *et al.*, 2016).

Logo, sendo os membros da família AUX/IAA CoREGs do tipo silenciadores de genes, e tendo o acesso sensível Jc171 mostrado intensa repressão de genes AUX/IAA (evidenciados na modulação DE do tipo DR), outros genes importantes para a resposta ao estresse evitariam ser silenciados desta maneira.

Por outro lado, a adição de IAA exógeno poderia minimizar alguns dos efeitos adversos do estresse salino (FORNI *et al.*, 2017). Em Jc183, alguns transcritos AUX/IAA, apesar da modulação da expressão não ter sido significativa, mostrou alguma modulação positiva, podendo ser estes motivos para estudos futuros.

### 5.2.2 Família Jumonji

A segunda maior família em CoREGs DE identificados foi Jumonji, sendo em sua maior parte de expressão DR, com alguns poucos UR expressos por Jc171.

Segundo Tsukada *et al.* (2006), as proteínas Jumonji possuem atividade de desmetilação, com um papel na modificação de histonas. Por exemplo, o gene *JMJ705* da proteína C do arroz codifica uma histona-lisina-desmetilase que reverte especificamente marcação repressiva de H3K27me2/3 (LI *et al.* 2013), permitindo a expressão do gene anteriormente silenciado (GRAFI & BARAK, 2015). No perfil de expressão de Jc171 foi observado CoREG (de expressão DR; DN26132\_c0\_g1) do tipo JMJ706. Ambos os CoREGs (JMJ705 e JMJ706) fazem parte do mesmo grupo (KDM4/JHDM3), segundo a análise filogenética desenvolvida por Lu *et al.* (2008).

Adicionalmente, as proteínas Jumonji também desempenhariam papel essencial na transdução de sinal de brassinoesteróides, e no desenvolvimento de órgãos florais e tempo de floração (Yuan et al., 2013). Portanto, essas proteínas além de atuar na configuração das histonas e da cromatina, tem por consequência atuação na regulação e transcrição de genes responsivos a estresses, bem como no desenvolvimento da planta.

Em estudo desenvolvido por Shen *et al.* (2014), a indução do gene *JMJ15* em mutantes de *A. thaliana* com ganho de função reduziu a altura das plantas e propiciou um acúmulo de lignina nas hastes. Por outro lado, mutantes com perda de função de *JMJ15* não apresentaram diferença de fenótipo quando comparados aos acessos selvagens. Além disto, os mutantes com ganho de função mostraram maior tolerância ao sal (plantas de 8 dias, após tratamentos com 100 a 150 mM NaCl), enquanto que mutantes com perda de função foram mais sensíveis ao sal, em comparação com o tipo selvagem.

A análise transcriptômica desenvolvida em mesmo estudo revelou que a indução de *JMJ15* regularia muitos genes marcados por H3K4me3 e H3K4me2, entre os quais, genes reguladores de transcrição envolvidos nas respostas ao estresse, resultando em aumento da tolerância. Desta forma, tendo em vista o grande número de transcritos Jumonji de expressão DR por Jc171 após o tratamento salino, esse acesso não se beneficiaria desta ação das desmetilases. Os ensaios de validação por RT-qPCR confirmaram a expressão DR de Jc171 para dois alvos Jumonji observados *in silico*.

### 5.2.3 Família PHD

Terceira família de CoREGs em quantidade de transcritos DE. As proteínas PHD são histona acetiltransferases (HATs) que modificam covalentemente as caudas N-terminais das histonas. A família se caracteriza por proteínas com domínios conservados, de motivos Cis 4-His-Cis 3, e que coordenam dois íons de zinco em uma configuração de "cross brace", onde cada íon de zinco é coordenado por pares alternados de ligantes Cis / His (MANSFIEL *et al.*, 2011). Estes domínios, comumente encontrados em proteínas associadas à cromatina, estão envolvidos essencialmente com modelação da cromatina e interação com histonas (AASLAND *et al.*, 1995). Como subunidades dos complexos de histona acetiltransferase ou histona desacetilases, as

proteínas PHD são necessárias para a ativação transcricional ou repressão transcricional (WEI *et al.*, 2009). Basicamente, todos os transcritos DE de Jc171 (após exposição ao tratamento salino) foram de expressão DR.

Wei *et al.* (2009) superexpressando o gene *GmPHD2* (de função de supressão de genes) em plantas de *A. thaliana* transgênicas após tratamento salino (150 mM NaCl) observaram maior taxa de germinação e melhor crescimento que as plantas do tipo selvagem. Logo, alguns genes PHD quando induzidos na resposta salina podem ser benéficos. Desta forma, a maior sensibilidade demonstrada por Jc171 pode ser decorrente também do perfil DR para representantes desta família, que basicamente não modularam no perfil Jc183 (alguns com tendência de indução).

#### 5.2.4 Família SNF2

Todos os representantes desta família foram de expressão DR na resposta ao sal de Jc171. As proteínas SNF2 são necessárias para a transcrição de certos genes, com envolvimento na fase de alongamento do transcrito pela RNA polimerase (Pol) II, a partir de interação com histonas (KWON & WAGNER, 2007; SCHWABISH & STRUHL, 2007). Também reprimem a transcrição pela RNA Pol II, e a transcrição de rDNA, além de estarem envolvidas no silenciamento telomérico, e no reparo do DNA pelo DSB (*double-strand break*) (MARTENS & WINSTON, 2002; SHIVASWAMY & IYER, 2008; TROTTER *et al.*, 2008).

Análises funcionais demonstraram que proteínas SNF2 desempenham papéis importantes no desenvolvimento das plantas e em respostas a estresse (HU *et al.*, 2013). Mlynarova *et al.* (2007) demonstra que a indução de gene *SNF2* do tipo CHR12 reduziu o crescimento e desenvolvimento de plantas *A. thaliana* expostas aos estresses de seca, calor e salinidade, enquanto o *knockout* do gene conferiu melhor crescimento de raízes e desenvolvimento das plantas relação às selvagens. No entanto, o desligamento de CHR12 sob estresse severo pode ser prejudicial para o vegetal uma vez que está relacionado à parada do crescimento, e em algumas ocasiões é útil para que a planta direcione seu gasto energético para responder ao estresse imposto. Logo, o fato dos representantes desta família apresentarem expressão DE do tipo DR pode ser interessante em concentrações de NaCl baixas. A validação da expressão DR por RT-qPCR foi confirmada para dois alvos representantes desta família.

### 5.2.5 Família SET

CoREGs desta família são essencialmente histona-lisina metiltransferases (BERR *et al.*, 2011), com relatos importantes associados ao desenvolvimento reprodutivo da planta (controle do tempo de floração, determinação do ciclo celular, morfogênese foliar, organogênese floral, *imprinting* parental e desenvolvimento de sementes; BERR *et al.*, 2011). Na literatura, o único representante SET envolvido na sinalização de estresse, no caso desidratação, esteve envolvido tanto na via dependente de ABA como independente da ABA, sendo este representante o fator TRX de *Arabidopsis* (ATX1), que tri-metila a histona H3 na lisina 4 (H3K4me3) (DING *et al.*, 2011).

Basicamente, todos os representantes desta família foram declarados DR (somente um representante UR), na resposta ao sal de Jc171, em plantas de 60 DAG, portanto, em desenvolvimento vegetativo.

### 5.3 MAPEAMENTO DE CoREGs EM GENOMA DE *M. esculenta*

Os dados obtidos do mapeamento puderam elucidar a distribuição do CoREGoma de *M. esculenta* ao longo do genoma, além de demonstrar também os cromossomos que tiveram as maiores quantidades de CoREGs detectados, sendo eles os cromossomos 03 (90), 15 (82) e 01 (73).

Os dados puderam evidenciar que os genes codificadores de proteínas CoREGs mapeados em genoma de *M. esculenta* a partir das sequências de *J. curcas* representaram uma fração entre 29,4 a 58,7% dos CoREGs totais para os cromossomos de *M. esculenta*. Os cromossomos 12, 15 e 01 foram os mais representativos em termos de quantidade de CoREGs mapeados a partir de CoREGs de *J. curcas*. Desta forma, sendo ambas as espécies taxonomicamente próximas, sugere-se que os CoREGs de *J. curcas* distribuam-se de forma semelhantes ao longo de seus cromossomos. Além disto, a porcentagem representativa de CoREGs de *J. curcas* mapeados em genoma de *M. esculenta* pode indicar quais porções do genoma são mais conservadas que outras entre as duas espécies.

#### 5.4 ANÁLISES FENÉTICAS DAS FAMÍLIAS AUX/IAA E JUMONJI

A análise fenética da família AUX/IAA ressaltou três grupos de CoREGs bem definidos em similaridades com CoREGs curadas de *M. esculenta*, e que incluiu CoREGs de *J. curcas*, alguns dos quais identificados a partir de *R. communis*. Os grupos, que englobaram os conjuntos completos de proteínas curadas, foram de proteínas IAA9, IAA27 e AUX22. A análise corroborou as anotações dos potenciais CoREGs de *J. curcas*.

Dos três grupos de proteínas, em respostas de plantas a estresses, há relato envolvendo proteína IAA9, em estudo de Li *et al.* (2016), demonstrando que após poda de inflorescências de plantas jovens (*Nicotiana tabacum*) ocorre um desbalanceamento de proteínas IAA (em especial, IAA9), modificando a distribuição de auxina ao longo do vegetal. Tal fato ocasionou menor grau de silenciamento gênico, que seria executado pelas proteínas IAA9, contribuindo com um melhor desenvolvimento de raízes laterais. No presente trabalho, um potencial transcrito para IAA9 foi declarado UR na expressão de Jc171 (DN100242\_c0\_g1); este, no entanto, não permitiu amplificação no ensaio de RT-qPCR.

A análise fenética com representantes da família Jumonji posicionou candidatos CoREGs de *J. curcas* em agrupamentos com conjuntos completos de proteínas curadas para JMJ25 e JMJ706, sendo que para JMJ25 houve subgrupos distintos.

A importância desses dois grupos de proteínas Jumonji no desenvolvimento vegetal encontra respaldo na literatura. Adounnet *et al.* (2017) mostraram que mutações em um gene homólogo à JMJ25, o JMJ24, com a perda de função, produziu uma série de defeitos no desenvolvimento de plantas mutantes. Estes defeitos afetaram genes ativos no estágio vegetativo. JMJ24 também foi mostrado interagindo com vias de RNAi envolvidos com o silenciamento gênico.

Por sua vez, Sun e Zhou (2008) observaram que em plantas de *Oryza sativa* (arroz) o gene JMJ706 codificaria uma desmetilase H3K9, a qual foi associada à heterocromatina e à regulação do desenvolvimento das flores. A proteína JMJ706 reverteria especificamente a di e trimetilação da lisina 9 da histona H3 (H3K9). Mutações que causariam a perda de função do gene levariam ao aumento de di- e trimetilações de H3K9 e afetariam o desenvolvimento de espiguetas, morfologia floral e número de órgãos reprodutivos.

## 5.5 ANÁLISE DE ENRIQUECIMENTO DE TERMOS DE ONTOLOGIA GÊNICA

Na análise de termos GO enriquecidos aqueles em comum entre CoREGs UR, DR ou não DE (diferencialmente expressos) ressaltaram principalmente respostas basais, celular e hormonal, conforme esperado. Os termos enriquecidos mais informativos corresponderam aqueles exclusivos decorrentes da categorização GO de proteínas CoREGs cuja abundância foi expressivamente reduzida após o tratamento salino. Muitos destes termos GO ressaltam nitidamente a regulação negativa (de expressão gênica, de transcrição, de biossíntese metabólica etc.), conforme as ações esperadas de CoREGs, por exemplo, das famílias Jumonji e AUX/IAA, comentadas anteriormente.

## 5.6 ANÁLISE DE ENRIQUECIMENTO DE FATORES DE TRANSCRIÇÃO

Dos FTs enriquecidos, tendo como alvos o conjunto de transcritos UniGenes declarados DE da resposta de Jc171 após tratamento salino, aqueles das famílias ERF e DOF foram os mais representados como prováveis FTs reguladores de genes responsivos ao sal. Em geral, os TFs ERF foram mais representativos na regulação de genes de expressão UR, enquanto FTs DOF foram mais representativos regulando genes DE de expressão DR.

FTs ERF apresentam-se como sendo responsivos ao etileno, sendo proteínas da subfamília APETALA2, e exclusivamente presente em plantas. Esses FTs compartilham domínios de aminoácidos conservados que podem se ligar a dois elementos ativos cis similares, a GCC box e ao elemento responsivo a desidratação DRE (SINGH *et al.*, 2002). Eles possuem relevante importância durante o desenvolvimento das plantas e em respostas aos estímulos do ambiente (CHEN J. *et al.*, 2008; CHEN G. *et al.*, 2008).

As proteínas da família DOF (*DNA-binding One Zinc Finger*) têm propriedades de ligação ao DNA semelhantes, em razão do domínio de ligação ao DNA ser altamente conservado. Estes FTs desempenham papéis importantes em muitos processos fundamentais em plantas superiores, como fotossíntese, respostas ao estresse, germinação de sementes, indução de flores e ritmos circadianos mediados pela luz (NOGUERO, 2013; KISU, 1995). Ainda, segundo Ma *et al.* (2015), os genes

DOF respondem a diferentes estresses abióticos (salinidade, seca e temperaturas extremas).

A análise de FTs enriquecidos tendo os CoREGs DE de expressão UR como alvos mostrou que três potenciais FTs enriquecidos (Jcr4S02641.40, Jcr4S09941.10, Jcr4S03743.20), todos integrantes da família bZIP, a qual está envolvida com sinalização de estresses, processos de defesa a patógenos, desenvolvimento de flores, e maturação de sementes (JAKOBY *et al.*, 2002).

Os potenciais TFs enriquecidos também interagiriam com quantidades expressivas de alvos gênicos de expressão UR (no mínimo 45 genes), a grande maioria de genes não codificadores de proteínas CoREGs, mas responsáveis ao tratamento salino aplicado. Estes TFs seriam candidatos interessantes como transgenes em ensaios futuros de transformação, visando obtenção de plantas com maior tolerância ao estresse salino.

## 6 CONSIDERAÇÕES FINAIS

Este trabalho apresentou a visão global do primeiro CoREGoma de *J. curcas*, ou seja, o primeiro painel do conjunto de genes CoREGs expresso em raízes de pinhão-manso após 3h de exposição ao tratamento salino de 150 mM de NaCl

No presente estudo, o perfil de expressão de UniGenes declarados DE de Jc171 em resposta ao sal foi em grande parte devido aos transcritos de regulação DR (cerca de 90%), sendo que dos 239 DE de proteínas CoREGs, 126 (52,7%) foram de representantes das famílias AUX/IAA (72) e Jumonji (54). Desses 126 transcritos, 118 (93,6%) foram de UniGenes DE de expressão DR, sendo 68 da família AUX/IAA e 50 da família Jumonji. As expressões potenciais CoREGs foram confirmadas em ensaio de RT-qPCR

As duas famílias, cujas abundâncias decresceram de modo significativo, são basicamente de proteínas envolvidas no silenciamento de genes e na modelação da cromatina, e que provavelmente afetou a expressão de diversos genes envolvidos com a resposta ao estresse salino, evidenciados na resposta Jc171 ao sal.

Vários desses genes diferencialmente expressos são candidatos para estudos futuros visando contribuir na elucidação do processo de expressão em resposta ao estresse salino, dentre outros, bem como como transgenes em estudos de transgenia. Também candidatos a transgenes são os três FTs previstos interagindo com cerca de 45 alvos moleculares dentre CoREGs e não CoREGs, todos de expressão UR em resposta ao sal.

Trabalhos futuros devem ser conduzidos com tratamentos salinos mais concentrados e em diferentes tempos de exposição, e mais prolongados, a fim de melhor avaliar a resposta Jc171, e a provável resposta tolerante do acesso Jc183. Sugere-se também, esforços em estudos envolvendo silenciamento gênico, tendo como candidatos CoREGs regulados negativamente, especialmente aqueles descritos como membros das famílias AUX/IAA e Jumonji. Os resultados demonstrados aliados as informações da literatura para as famílias CoREGs destacadas criam novas perspectivas para o melhoramento vegetal de *J. curcas*. O desenvolvimento de cultivares melhores adaptados à ambientes áridos contribui para o crescimento econômico da cultura.

## REFERÊNCIAS

- ADAMS, Mark D. *et al.* Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**, v. 252, n. 5013, p. 1651-1656, 1991.
- AUDONNET, Laure; SHEN, Yuan; ZHOU, Dao-Xiu. JMJ24 antagonizes histone H3K9 demethylase IBM1/JMJ25 function and interacts with RNAi pathways for gene silencing. **Gene Expression Patterns**, v. 25, p. 1-7, 2017.
- AGROENERGIA, EMBRAPA. Pinhão manso: matéria-prima potencial para produção de biodiesel. 2009.
- ALTSCHUL, Stephen F. *et al.* Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403-410, 1990.
- ANTOSCH, Martin; MORTENSEN, Simon A.; GRASSER, Klaus D. Plant proteins containing high mobility group box DNA-binding domains modulate different nuclear processes. **Plant physiology**, v. 159, n. 3, p. 875-883, 2012.
- ASIF, Mehar H. *et al.* Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. **Tree genetics & genomes**, v. 6, n. 6, p. 941-952, 2010.
- BARROS, Maria de FC *et al.* Recuperação de solos afetados por sais pela aplicação de gesso de jazida e calcário no Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 2004.
- BATISTA, M.J. *et al.* Drenagem como instrumento de dessalinização e prevenção da salinização de solos. **Brasília: CODEVASF, 216p**, 2002.
- BERR, Alexandre; SHAFIQ, Sarfraz; SHEN, Wen-Hui. Histone modifications in transcriptional activation during plant development. **Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms**, v. 1809, n. 10, p. 567-576, 2011.
- BEST, R.; HENRY, G. Report of the first meeting of the international network for cassava genetic resources. **Cassava Genetic Resources. Cali, CO, USA: Cent Int Agric Trop**, p. 3-11, 1992.
- BIENZ, Mariann. The PHD finger, a nuclear protein-interaction domain. **Trends in biochemical sciences**, v. 31, n. 1, p. 35-40, 2006.
- BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.
- BRITTAINE, Richard *et al.* **Jatropha: a smallholder bioenergy crop: the potential for pro-poor development**. Food and Agriculture Organization of the United Nations (FAO), 2010.
- BRITTEN, Roy J.; DAVIDSON, Eric H. Regulamento de genes para células superiores: uma teoria. **Ciência**, v. 165, n. 3891, p. 349-357, 1969.

BURDO, Brett *et al.* The Maize TFome—development of a transcription factor open reading frame collection for functional genomics. **The Plant Journal**, v. 80, n. 2, p. 356-366, 2014..

BUSTIN, Michael; CATEZ, Frédéric; LIM, Jae-Hwan. The dynamics of histone H1 function in chromatin. **Molecular cell**, v. 17, n. 5, p. 617-620, 2005.

BUSTIN, Stephen A.; NOLAN, Tania. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. **Journal of biomolecular techniques: JBT**, v. 15, n. 3, p. 155, 2004.

CAIRNS, Bradley R. Chromatin remodeling machines: similar motors, ulterior motives. **Trends in biochemical sciences**, v. 23, n. 1, p. 20-25, 1998.

CAMPOS, Mariana Lins de Oliveira *et al.* Photosynthesis and antioxidant activity in *Jatropha curcas* L. under salt stress. **Brazilian Journal of Plant Physiology**, v. 24, n. 1, p. 55-67, 2012.

CARELS, Nicolas. *Jatropha curcas*: a review. **Advances in botanical research**, v. 50, p. 39-86, 2009.

CARVALHO, Carlos Roberto *et al.* Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. **Plant Science**, v. 174, n. 6, p. 613-617, 2008.

CAUÊ DE HOLANDA, Alan *et al.* Desenvolvimento inicial de espécies arbóreas em ambientes degradados por sais. **Revista de biologia e ciências da terra**, v. 7, n. 1, 2007.

CHAI, Bob *et al.* Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. **Genes & development**, v. 19, n. 14, p. 1656-1661, 2005.

CHANDRASEKARAN, Umashankar; XU, Wei; LIU, Aizhong. Transcriptome profiling identifies ABA mediated regulatory changes towards storage filling in developing seeds of castor bean (*Ricinus communis* L.). **Cell & bioscience**, v. 4, n. 1, p. 33, 2014.

CHOWRASIA, Soni *et al.* Identification of jumonjiC domain containing gene family among the *Oryza* species and their expression analysis in FL478, a salt tolerant rice genotype. **Plant Physiology and Biochemistry**, v. 130, p. 43-53, 2018.

CONTRERAS-CORNEJO, Hexon Angel *et al.* *Trichoderma* spp. improve growth of *Arabidopsis* seedlings under salt stress through enhanced root development, osmolite production, and Na<sup>+</sup> elimination through root exudates. **Molecular Plant-Microbe Interactions**, v. 27, n. 6, p. 503-514, 2014.

CORTESÃO, M. *Culturas tropicais: plantas oleaginosas*. Lisboa: Clássica, 1956. 231 p.

COSTA, Gustavo GL *et al.* Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. **BMC genomics**, v. 11, n. 1, p. 462, 2010.

DAI, Xinbin *et al.* PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. **BMC bioinformatics**, v. 14, n. 1, p. 321, 2013.

DE ARRUDA, FRANCINEUMA PONCIANO *et al.* Cultivo de pinhão manso (*Jatropha curca* L.) como alternativa para o semi-árido nordestino. **Revista brasileira de oleaginosas e fibrosas**, v. 8, n. 1, 2004.

DE AZEVEDO, D. M. P. *et al.* Recomendações técnicas para o cultivo da mamoneira (*Ricinus communis* L.) no Nordeste do Brasil. **Embrapa Algodão-Circular Técnica (INFOTECA-E)**, 1997.

DE MATTOS, P. L. P.; FARIAS, Alba Rejane Nunes; FERREIRA FILHO, José Raimundo. **Mandioca: o produtor pergunta, a Embrapa responde**. Brasília, DF: Embrapa Informação Tecnológica; Cruz das Almas: Embrapa Mandioca e Fruticultura Tropical, 2006., 2006.

DING, Yong; AVRAMOVA, Zoya; FROMM, Michael. The Arabidopsis trithorax-like factor ATX1 functions in dehydration stress responses via ABA-dependent and ABA-independent pathways. **The Plant Journal**, v. 66, n. 5, p. 735-744, 2011.

DING, Zehong *et al.* Transcriptome response of cassava leaves under natural shade. **Scientific reports**, v. 6, p. 31673, 2016.

DIVAKARA, B. N. *et al.* Biology and genetic improvement of *Jatropha curcas* L.: a review. **Applied Energy**, v. 87, n. 3, p. 732-742, 2010.

DODD, Ian C.; PÉREZ-ALFOCEA, Francisco. Microbial amelioration of crop salinity stress. **Journal of Experimental Botany**, v. 63, n. 9, p. 3415-3428, 2012.

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **PNAS**, v. 95, 14863-14868, 1998

EPSTEIN, Emanuel *et al.* **Mineral nutrition of plants: principles and perspectives**. 1972.

FAIRLESS, Daemon. Biofuel: the little shrub that could-maybe. **Nature News**, v. 449, n. 7163, p. 652-655, 2007.

FANG, Hui *et al.* Expression analysis of histone acetyltransferases in rice under drought stress. **Biochemical and biophysical research communications**, v. 443, n. 2, p. 400-405, 2014.

FLAGELLA, Zina *et al.* Effect of saline water on oil yield and quality of a high oleic sunflower (*Helianthus annuus* L.) hybrid. **European Journal of Agronomy**, v. 21, n. 2, p. 267-272, 2004.

FLOWERS, Timothy J.; COLMER, Timothy D. Salinity tolerance in halophytes. **New Phytologist**, v. 179, n. 4, p. 945-963, 2008.

FORNI, Cinzia; DUCA, Daiana; GLICK, Bernard R. Mechanisms of plant response to salt and drought stress and their alteration by rhizobacteria. **Plant and Soil**, v. 410, n. 1-2, p. 335-356, 2017.

GAO, S. *et al.* Effects of salt stress on growth, antioxidant enzyme and phenylalanine ammonia-lyase activities in *Jatropha curcas* L. seedlings. **Plant Soil Environ**, v. 54, n. 9, p. 374-381, 2008.

GHANASHYAM, Challa; JAIN, Mukesh. Role of auxin-responsive genes in biotic stress responses. **Plant signaling & behavior**, v. 4, n. 9, p. 846-848, 2009.

GOVIND, Geetha *et al.* Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. **Molecular Genetics and Genomics**, v. 281, n. 6, p. 591-605, 2009.

GOSWAMI, N. K.; SAHARIA, D.; KAR, A. Traditional uses of *Jatropha curcas* Linnaeus [Euphorbiaceae] as medicine by different communities in Northeast India. **Pleione**, v. 7, n. 1, p. 66-72, 2013

GRABHERR, Manfred G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology**, v. 29, n. 7, p. 644, 2011.

GRAFI, Gideon; BARAK, Simon. Stress induces cell dedifferentiation in plants. **Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms**, v. 1849, n. 4, p. 378-384, 2015.

GÜBITZ, Georg M.; MITTELBACH, Martin; TRABI, Manuela. Exploitation of the tropical oil seed plant *Jatropha curcas* L. **Bioresource technology**, v. 67, n. 1, p. 73-82, 1999.

HASEGAWA, Paul M. *et al.* Plant cellular and molecular responses to high salinity. **Annual review of plant biology**, v. 51, n. 1, p. 463-499, 2000.

HA, Jungmin *et al.* Genome sequence of *Jatropha curcas* L., a non-edible biodiesel plant, provides a resource to improve seed-related traits. **Plant biotechnology journal**, v. 17, n. 2, p. 517-530, 2019

HEID, Christian A. *et al.* Real time quantitative PCR. **Genome research**, v. 6, n. 10, p. 986-994, 1996.

HERNANDEZ, J. A. *et al.* Salt-induced oxidative stress in chloroplasts of pea plants. **Plant Science**, v. 105, n. 2, p. 151-167, 1995.

HENNING, Reinhard K. The *Jatropha* system, integrated rural development by utilization of *Jatropha curcas* L. (JCL) as raw material and as renewable energy. **Weissenberg, Germany**, 2004.

HORSCHUTZ, Ana CO *et al.* Crescimento e produtividade do pinhão-mansão em função do espaçamento e irrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi**, v. 16, n. 10, 2012.

- HRDLICKOVA, Radmila; TOLOUE, Masoud; TIAN, Bin. RNA-Seq methods for transcriptome analysis. **Wiley Interdisciplinary Reviews: RNA**, v. 8, n. 1, p. e1364, 2017.
- HU, Yongfeng *et al.* Analysis of rice Snf2 family proteins and their potential roles in epigenetic regulation. **Plant physiology and biochemistry**, v. 70, p. 33-42, 2013.
- JAKOBY, Marc *et al.* bZIP transcription factors in Arabidopsis. **Trends in plant science**, v. 7, n. 3, p. 106-111, 2002.
- JIN, Jinpu *et al.* PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. **Nucleic acids research**, v. 42, n. D1, p. D1182-D1187, 2013.
- JUDD, Walter S. *et al.* **Sistemática Vegetal:- Um Enfoque Filogenético**. Artmed Editora, 2009.
- JUNG, Harin *et al.* OsIAA6, a member of the rice AUX/IAA gene family, is involved in drought tolerance and tiller outgrowth. **Plant Science**, v. 236, p. 304-312, 2015.
- KALKHOVEN, Eric *et al.* The PHD type zinc finger is an integral part of the CBP acetyltransferase domain. **Molecular and cellular biology**, v. 22, n. 7, p. 1961-1970, 2002.
- KANEHISA, Minoru; GOTO, Susumu. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27-30, 2000.
- KHEIRA, Abdrabbo A. Abou; ATTA, Nahed MM. Response of *Jatropha curcas* L. to water deficits: Yield, water use efficiency and oilseed characteristics. **Biomass and bioenergy**, v. 33, n. 10, p. 1343-1350, 2009.
- KESHISHIAN, Erika A. *et al.* Salt and oxidative stresses uniquely regulate tomato cytokinin levels and transcriptomic response. **Plant Direct**, v. 2, n. 7, p. e00071, 2018.
- KING, Andrew J. *et al.* Potential of *Jatropha curcas* as a source of renewable oil and animal feed. **Journal of experimental botany**, v. 60, n. 10, p. 2897-2905, 2009.
- KUDAPA, Himabindu *et al.* Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. **PloS one**, v. 9, n. 1, p. e86039, 2014.
- KWON, Chang Seob; WAGNER, Doris. Unwinding chromatin for development and growth: a few genes at a time. **Trends in Genetics**, v. 23, n. 8, p. 403-412, 2007.
- LANGMEAD, Ben *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome biology**, v. 10, n. 3, p. R25, 2009.
- LARKIN, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. Clustal W and Clustal X version 2.0. **Bioinformatics**, 23:2947-2948, 2007.

LAVIOLA, Bruno Galvêas et al. Genetic improvement of *Jatropha* for biodiesel production. **Ceiba**, v. 51, n. 1, p. 1-10, 2010.

LI, Bo; DEWEY, Colin N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC bioinformatics**, v. 12, n. 1, p. 323, 2011.

LI, Fei et al. Identification of topping responsive proteins in tobacco roots. **Frontiers in plant science**, v. 7, p. 582, 2016.

LIU, Lin et al. Comparison of next-generation sequencing systems. **BioMed Research International**, v. 2012, 2012.

LI, Tiantian *et al.* Jumonji C domain protein JMJ705-mediated removal of histone H3 lysine 27 trimethylation is involved in defense-related gene activation in rice. **The Plant Cell**, v. 25, n. 11, p. 4725-4736, 2013.

LIU, Xuncheng *et al.* PHYTOCHROME INTERACTING FACTOR3 associates with the histone deacetylase HDA15 in repression of chlorophyll biosynthesis and photosynthesis in etiolated *Arabidopsis* seedlings. **The Plant Cell**, v. 25, n. 4, p. 1258-1273, 2013.

LOPES, José Carlos; MACEDO, CMP de. Germinação de sementes de couve chinesa sob influência do teor de água, substrato e estresse salino. **Revista Brasileira de Sementes**, v. 30, n. 3, p. 79-85, 2008.

LOZANO-ISLA, Flavio *et al.* Effects of seed storage time and salt stress on the germination of *Jatropha curcas* L. **Industrial Crops and Products**, v. 118, p. 214-224, 2018.

LU, Falong et al. Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in *Arabidopsis* and rice. **Journal of integrative plant biology**, v. 50, n. 7, p. 886-896, 2008.

LUO, Jie; ZHOU, Jing-Jing; ZHANG, Jin-Zhi. AUX/IAA gene family in plants: molecular structure, regulation, and function. **International journal of molecular sciences**, v. 19, n. 1, p. 259, 2018.

MA, Jing et al. Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. **BMC genomics**, v. 16, n. 1, p. 33, 2015.

MAES, W. H. A. et al. Climatic growing conditions of *Jatropha curcas* L. **Biomass and bioenergy**, v. 33, n. 10, p. 1481-1485, 2009.

MAGALHÃES, A. F. Manejo da fertilidade dos solos irrigados: produtividade, degradação e correção. **Reunião Brasileira de Fertilidade do Solo e Nutrição de Plantas**, v. 21, p. 77-86, 1995.

MANSFIELD, Robyn E. et al. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. **Journal of Biological Chemistry**, v. 286, n. 13, p. 11779-11791, 2011.

MARDIS, Elaine R. Next-generation DNA sequencing methods. **Annu. Rev. Genomics Hum. Genet.**, v. 9, p. 387-402, 2008.

MARGULIES, Marcel et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, n. 7057, p. 376, 2005.

MARTENS, Joseph A.; WINSTON, Fred. Evidence that Swi/Snf directly represses transcription in *S. cerevisiae*. **Genes & development**, v. 16, n. 17, p. 2231-2236, 2002.

MARTIN, Jeffrey A.; WANG, Zhong. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671, 2011.

METZKER, Michael L. Sequencing technologies—the next generation. **Nature reviews genetics**, v. 11, n. 1, p. 31, 2010.

MEYERS BC, *et al.* Methods for transcriptional profiling in plants. Be Fruitful and Replicate. **Plant Physiol** 135: 637–652, 2004

MLYNÁROVÁ, Ludmila; NAP, Jan-Peter; BISSELING, Ton. The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in *Arabidopsis thaliana* upon perceiving environmental stress. **The Plant Journal**, v. 51, n. 5, p. 874-885, 2007.

MOHINUDEEN, C. et al. An Overview of Next-Generation Sequencing (NGS) Technologies to Study the Molecular Diversity of Genome. In: **Microbial Applications Vol. 1**. Springer, Cham, 2017. p. 295-317.

MOREIRA, J. *et al.* Melhoramento da mamoneira (*Ricinus communis* L.). **EMBRAPA-CNPA. Documentos**, 1996.

MORIN R, Bainbridge M, Fejes A, et al. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. **Biotechniques** 45: 81-94.

MOROZOVA, Olena; MARRA, Marco A. Applications of next-generation sequencing technologies in functional genomics. **Genomics**, v. 92, n. 5, p. 255-264, 2008.

MUDGAL, V. *et al.* Biochemical mechanisms of salt tolerance in plants: a review. **International Journal of Botany**, v. 6, n. 2, p. 136-143, 2010.

MUNNS, Rana. Comparative physiology of salt and water stress. **Plant, cell & environment**, v. 25, n. 2, p. 239-250, 2002.

NAGALAKSHMI, Ugrappa; WAERN, Karl; SNYDER, Michael. RNA-Seq: a method for comprehensive transcriptome analysis. **Current protocols in molecular biology**, v. 89, n. 1, p. 4.11. 1-4.11. 13, 2010.

NATARAJAN, Purushothaman; PARANI, Madasamy. De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing. **BMC genomics**, v. 12, n. 1, p. 191, 2011.

NETO FIGUEIREDO, Acácio *et al.* Divergência genética em acessos de mamona (*Ricinus communis* L.) baseada nas características das sementes. **Revista de Biologia e Ciências da Terra**, v. 4, n. 2, 2004.

NIU, Genhua *et al.* Responses of *Jatropha curcas* to salt and drought stresses. **International journal of agronomy**, v. 2012, 2012.

NÓBREGA, Márcia Barreto de Medeiros. **Avaliação de genótipos de mamona (*Ricinus communis* L.) em cruzamentos dialélicos parciais**. 2008. Tese de Doutorado. Universidade de São Paulo.

NUNES, Sidemar Presotto. Produção e consumo de óleos vegetais no Brasil. **Departamento de Estudos Sócio-Econômico Rurais**, n. 159, 2007.

NYRÉN, Pettersson; PETTERSSON, Bertil; UHLÉN, Mathias. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. **Analytical biochemistry**, v. 208, n. 1, p. 171-175, 1993.

O'MALLEY, Bert W. Coregulators: from whence came these "master genes". **Molecular Endocrinology**, v. 21, n. 5, p. 1009-1013, 2007.

OLIVEROS, Juan Carlos. VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>, 2007.

OLSEN, Kenneth M.; SCHAAL, Barbara A. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. **Proceedings of the National Academy of Sciences**, v. 96, n. 10, p. 5586-5591, 1999.

OPENSHAW, Keith. A review of *Jatropha curcas*: an oil plant of unfulfilled promise. **Biomass and bioenergy**, v. 19, n. 1, p. 1-15, 2000.

OUYANG, Bo *et al.* Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. **Journal of experimental botany**, v. 58, n. 3, p. 507-520, 2007.

OZSOLAK, Fatih; MILOS, Patrice M. RNA sequencing: advances, challenges and opportunities. **Nature reviews genetics**, v. 12, n. 2, p. 87, 2011.

PÉREZ-RODRÍGUEZ, Paulino *et al.* PInTFDB: updated content and new features of the plant transcription factor database. **Nucleic acids research**, v. 38, n. suppl\_1, p. D822-D827, 2009.

PFÄFFL, Michael W.; HORGAN, Graham W.; DEMPFFLE, Leo. Relative expression *software* tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic acids research**, v. 30, n. 9, p. e36-e36, 2002.

PRAMANIK, K. Properties and use of *Jatropha curcas* oil and diesel fuel blends in compression ignition engine. **Renewable energy**, v. 28, n. 2, p. 239-248, 2003.

QIU, Nianwei et al. Physiological and transcriptomic responses of Chinese cabbage (*Brassica rapa* L. ssp. *Pekinensis*) to salt stress. **International journal of molecular sciences**, v. 18, n. 9, p. 1953, 2017.

RIBEIRO, M. R.; FREIRE, F. J.; MONTENEGRO, AA de A. Solos halomórficos no Brasil: Ocorrência, gênese, classificação, uso e manejo sustentável. **CURI, N.; MARQUES, JJ; GUILHERME, LRG; LIMA, JM**, p. 165-208, 2003.

RITCHIE, Matthew E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic acids research**, v. 43, n. 7, p. e47-e47, 2015.

ROBINSON, Mark D.; MCCARTHY, Davis J.; SMYTH, Gordon K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.

ROGERS, David J.; APPAN, Subramaniam G. Manihot and Manihotoides (*Euphorbiaceae*): A computer assisted study. **Flora neotropica**, v. 13, 1973.

RYU, Hojin; CHO, Yong-Gu. Plant hormones in salt stress tolerance. **Journal of Plant Biology**, v. 58, n. 3, p. 147-155, 2015.

SAHA, Saurabh et al. Using the transcriptome to annotate the genome. **Nature biotechnology**, v. 20, n. 5, p. 508, 2002.

SALDANHA, Alok J. Java Treeview—extensible visualization of microarray data. **Bioinformatics**, v. 20, n. 17, p. 3246-3248, 2004.

SANGER, Frederick; NICKLEN, Steven; COULSON, Alan R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the national academy of sciences**, v. 74, n. 12, p. 5463-5467, 1977.

SAPETA, Helena et al. Transcriptomics and physiological analyses reveal coordinated alteration of metabolic pathways in *Jatropha curcas* drought tolerance. **Journal of experimental botany**, v. 67, n. 3, p. 845-860, 2015.

SCHMUTZ, Jeremy *et al.* Genome sequence of the palaeopolyploid soybean. **nature**, v. 463, n. 7278, p. 178, 2010.

SCHWABISH, Marc A.; STRUHL, Kevin. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. **Molecular and cellular biology**, v. 27, n. 20, p. 6987-6995, 2007.

SHANI, Eilon *et al.* Plant stress tolerance requires auxin-sensitive AUX/IAA transcriptional repressors. **Current Biology**, v. 27, n. 3, p. 437-444, 2017.

SHARMA, Arti; CHAUHAN, Rajinder Singh. Identification and association analysis of castor bean orthologous candidate gene-based markers for high oil content in *Jatropha curcas*. **Plant molecular biology reporter**, v. 30, n. 4, p. 1025-1031, 2012.

SHARMA, Arti; CHAUHAN, Rajinder Singh. Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha curcas*. **Comparative and functional genomics**, v. 2011, 2011.

SHEN, Yuan *et al.* Over-expression of histone H3K4 demethylase gene JMJ15 enhances salt tolerance in *Arabidopsis*. **Frontiers in plant science**, v. 5, p. 290, 2014.

SHIMKETS RA. **Gene expression profiling**. 1<sup>o</sup> ed. Totowa: Humana Press, Inc, 2004.

SHIVASWAMY, Sushma; IYER, Vishwanath R. Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. **Molecular and cellular biology**, v. 28, n. 7, p. 2221-2234, 2008.

SILVA, EN da *et al.* Comparative effects of salinity and water stress on photosynthesis, water relations and growth of *Jatropha curcas* plants. **Journal of Arid Environments**, v. 74, n. 10, p. 1130-1137, 2010.

SILVA, Evandro Nascimento da *et al.* Salt stress induced damages on the photosynthesis of physic nut young plants. **Scientia Agricola**, v. 68, n. 1, p. 62-68, 2011.

SINGH, Karam B.; FOLEY, Rhonda C.; OÑATE-SÁNCHEZ, Luis. Transcription factors in plant defense and stress responses. **Current opinion in plant biology**, v. 5, n. 5, p. 430-436, 2002.

SOUZA, Y. A. *et al.* Efeito da salinidade na germinação de sementes e no crescimento inicial de mudas de pinhão-manso. **Embrapa Semiárido-Artigo em periódico indexado (ALICE)**, 2010.

ŠTROS, M.; LAUNHOLT, Dorte; GRASSER, Klaus D. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. **Cellular and molecular life sciences**, v. 64, n. 19-20, p. 2590, 2007.

SUAREZ, Paulo AZ *et al.* Transformation of triglycerides into fuels, polymers and chemicals: some applications of catalysis in oleochemistry. **Química Nova**, v. 30, n. 3, p. 667-676, 2007.

SUN, Qianwen; ZHOU, Dao-Xiu. Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development. **Proceedings of the National Academy of Sciences**, v. 105, n. 36, p. 13679-13684, 2008.

TANG, Yuehui *et al.* Genome-wide analysis of the AP2/ERF gene family in physic nut and overexpression of the JcERF011 gene in rice increased its sensitivity to salinity stress. **PLoS One**, v. 11, n. 3, p. e0150879, 2016.

TERAUCHI R, *et al.* SuperSAGE: The most advanced transcriptoma technology for functional genomics. In: Kahl, Günter; Meksem, Khalid (eds.). **The handbook of plant functional genomics**. 1<sup>a</sup> ed. Weinheim: Wiley-VCH Verlag GmbH & Co, cap. 3, p. 37-54, 2008.

THORNTHWAITE, C. W.; MATHER, J. R. The water balance. New Jersey: Drexel Institute of Technology, 1955. 104p. **Publications in Climatology**.

TROTTER, Kevin W. *et al.* The HSA domain of BRG1 mediates critical interactions required for glucocorticoid receptor-dependent transcriptional activation in vivo. **Molecular and cellular biology**, v. 28, n. 4, p. 1413-1426, 2008.

TSUKADA, Yu-ichi *et al.* Histone demethylation by a family of JmjC domain-containing proteins. **Nature**, v. 439, n. 7078, p. 811, 2006.

UTSUMI, Y. *et al.* Transcriptome analysis using a high-density oligomicroarray under drought stress in various genotypes of cassava: an important tropical crop. **DNA research**, v. 19, n. 4, p. 335-345, 2012

VAN DE LOO, Frank J. *et al.* An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. **Proceedings of the National Academy of Sciences**, v. 92, n. 15, p. 6743-6747, 1995.

VELCULESCU, Victor E. *et al.* Serial analysis of gene expression. **Science**, v. 270, n. 5235, p. 484-487, 1995.

VIRGENS, Ivana Oliveira *et al.* *Jatropha curcas* L.: morphophysiological and chemical aspects. **Brazilian Journal of Food Technology**, v. 20, 2017.

VLACHONASIOS, Konstantinos E.; THOMASHOW, Michael F.; TRIEZENBERG, Steven J. Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect *Arabidopsis* growth, development, and gene expression. **The Plant Cell**, v. 15, n. 3, p. 626-638, 2003.

WANG, Chunyang *et al.* Insights into the origin and evolution of the plant hormone signaling machinery. **Plant physiology**, v. 167, n. 3, p. 872-886, 2015.

WANG, Haibo *et al.* Global analysis of transcriptome responses and gene expression profiles to cold stress of *Jatropha curcas* L. **PloS one**, v. 8, n. 12, p. e82817, 2013.

WANG, Youning; LI, Kexue; LI, Xia. Auxin redistribution modulates plastic development of root system architecture under salt stress in *Arabidopsis thaliana*. **Journal of plant physiology**, v. 166, n. 15, p. 1637-1645, 2009.

WANG, Zhong; GERSTEIN, Mark; SNYDER, Michael. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews genetics**, v. 10, n. 1, p. 57, 2009.

WEI, Wei *et al.* A histone code reader and a transcriptional activator interact to regulate genes for salt tolerance. **Plant physiology**, v. 175, n. 3, p. 1304-1320, 2017.

WEN, Mingfu *et al.* Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L. **BMC research notes**, v. 3, n. 1, p. 42, 2010.

WHANKAEW, Sukhuman *et al.* Cross-genera transferability of (simple sequence repeat) SSR markers among cassava (*Manihot esculenta* Crantz), rubber tree

(*Hevea brasiliensis* Muell. Arg.) and physic nut (*Jatropha curcas* L.). **African Journal of Biotechnology**, v. 10, n. 10, p. 1768-1776, 2011.

WILLADINO, Lilia; CAMARA, Terezinha Rangel. Tolerância das plantas à salinidade: aspectos fisiológicos e bioquímicos. **Enciclopédia biosfera**, v. 6, n. 11, p. 1-23, 2010.

WILSKER, Deborah *et al.* Nomenclature of the ARID family of DNA-binding proteins. **Genomics**, v. 86, n. 2, p. 242-251, 2005.

WOLD, Barbara; MYERS, Richard M. Sequence census methods for functional genomics. **Nature methods**, v. 5, n. 1, p. 19, 2007.

XU, Gang *et al.* Transcriptome analysis of flower sex differentiation in *Jatropha curcas* L. using RNA sequencing. **PLoS one**, v. 11, n. 2, p. e0145613, 2016.

YANG, Chang-Qing *et al.* Transcriptional Regulation of Plant Secondary Metabolism F. **Journal of integrative plant biology**, v. 54, n. 10, p. 703-712, 2012.

YILMAZ, Alper *et al.* AGRIS: the Arabidopsis gene regulatory information server, an update. **Nucleic acids research**, v. 39, n. suppl\_1, p. D1118-D1122, 2010.

YOU, Jun *et al.* Transcriptomic and metabolomic profiling of drought-tolerant and susceptible sesame genotypes in response to drought stress. **BMC Plant Biology**, v. 19, n. 1, p. 267, 2019.

YU, Jun *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). **science**, v. 296, n. 5565, p. 79-92, 2002.

YUAN, Lianyu *et al.* Involvement of histone modifications in plant abiotic stress responses. **Journal of integrative plant biology**, v. 55, n. 10, p. 892-901, 2013.

ZHANG, Lin *et al.* Global analysis of gene expression profiles in physic nut (*Jatropha curcas* L.) seedlings exposed to salt stress. **PLoS One**, v. 9, n. 5, p. e97878, 2014.

ZHENG, Binglian *et al.* An ARID domain-containing protein within nuclear bodies is required for sperm cell formation in *Arabidopsis thaliana*. **PLoS genetics**, v. 10, n. 7, p. e1004421, 2014.

ZHENG, Yi *et al.* iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. **Molecular Plant**, v. 9, n. 12, p. 1667-1670, 2016.

## APÊNDICE A

**Apêndice A.** Pares de *primers* propostos usados em ensaios de validação em RT-qPCR. A anotação representa a família para qual o par de *primer* foi desenhado. T<sub>m</sub> indica o time melting de cada *primer* e do amplicon.

ID do transcrito	Anotação	Foward primer	T <sub>m</sub> (°C)	Reverse primer	T <sub>m</sub> (°C)	Amplicon (bp)	T <sub>m</sub> (°C)
TRINITY_DN27423_c0_g2	<i>SET</i>	5'GCAGAACTTGCTACCATAGG 3'	58.93	5'ACAAGAGAAGTGAATGCCTTGG 3'	60.66	111	73.70
TRINITY_DN38633_c1_g2	<i>Jumonji1</i>	5'TTGTTTCAGGACAGGAAGAAGC 3'	60.77	5'TGTAGGGACATCGAATGTAAGG 3'	58.97	105	69.41
TRINITY_DN32491_c0_g1	<i>Jumonji2</i>	5'GATCACAGAGAAGCTTTGATGC 3'	59.08	5'ATCATTCCTTCGACACTGATGG 3'	61.27	110	72.71
TRINITY_DN85308_c0_g1	<i>MBF1</i>	5'CAACTGAGCAAGTTGAGACTGG 3'	60.09	5'GCACTACCTTAAACACGAGGAAGC 3'	59.38	136	73.57
TRINITY_DN75398_c0_g1	<i>GNAT</i>	5'CCCGCTGAACTCTATAAACACG 3'	61.01	5'GTGGAAGTGGGCTTTATGAGC 3'	61.00	132	74.92
TRINITY_DN34178_c0_g1	<i>SNF2A</i>	5'AGGAAGAAGATGGTTTGTCTGG 3'	59.62	5'GTTGGTAATCCCTCATCTTTCC 3'	58.83	87	73.27
TRINITY_DN62351_c0_g1	<i>SNF2B</i>	5'CCTATGCTGGAAGTCTCTGACC 3'	60.27	5'ACCACCGAGTTCTATTGTTTGC 3'	60.41	149	71.56
TRINITY_DN41924_c0_g2	<i>Mterf</i>	5'GATCTTGTTGTCTGAGCAATCG 3'	59.88	5'TAGTCAACCCTCACCAACTTCC 3'	60.39	88	69.39
Actin*	<i>ACT</i>	5'TAATGGTCCCTCTGGATGTG 3'		5'AGAAAAGAAAAGAAAAAAGCAGC 3'		147	60.00
β-tubulin*	<i>β-TUB</i>	5'CTCTGCAACTATGAGTGGTGTAAACG 3'		5'CACGAGAAGTAAGTGGGGCAA 3'		143	60.00

\*Gene de referência: Actina (Tang et al., 2016), e β-tubulin (Xu et al., 2016)

## APÊNDICE B

**Apêndice B.** Unigenes RNA-Seq potenciais CoREGs de *J. curcas* que foram diferencialmente [*p-value* 0.0001; FDR 0.005; Log<sub>2</sub>FC > 1 (UR); Log<sub>2</sub>FC < -1 (DR)] expressos nas bibliotecas do acesso sensível Jc171 em resposta ao tratamento salino (3h, 150 mM NaCl) e seus genes correspondentes nas bibliotecas do acesso tolerante Jc183.

Transcrito	Família CoREG	Expressão	Log <sub>2</sub> FC Jc171	Log <sub>2</sub> FC Jc183
DN26682_c0_g3	ARID	DR	-2.22	-0.40
DN38151_c0_g1	ARID	DR	-2.82	-0.50
DN27450_c0_g1	ARID	DR	-6.43	-0.90
DN34126_c1_g1	AUX/IAA	UR	2.39	1.00
DN74023_c0_g1	AUX/IAA	UR	1.76	1.00
DN35938_c2_g3	AUX/IAA	UR	9.83	0.50
DN100242_c0_g1	AUX/IAA	UR	2.34	0.90
DN85693_c0_g1	AUX/IAA	DR	-1.96	-0.80
DN41944_c0_g1	AUX/IAA	DR	-1.99	-0.60
DN14378_c0_g1	AUX/IAA	DR	-2.04	-0.90
DN73234_c0_g1	AUX/IAA	DR	-2.07	-0.60
DN62157_c0_g1	AUX/IAA	DR	-2.09	-0.20
DN34610_c0_g1	AUX/IAA	DR	-2.18	-0.80
DN61241_c0_g1	AUX/IAA	DR	-2.19	-0.80
DN43982_c0_g1	AUX/IAA	DR	-2.34	-1.00
DN43554_c0_g1	AUX/IAA	DR	-2.36	-0.30
DN9269_c0_g1	AUX/IAA	DR	-2.36	-0.50
DN39198_c1_g1	AUX/IAA	DR	-2.45	-0.80
DN24695_c0_g1	AUX/IAA	DR	-2.49	-0.60
DN28453_c0_g2	AUX/IAA	DR	-2.49	-1.00
DN42909_c2_g1	AUX/IAA	DR	-2.57	-0.80
DN49620_c0_g1	AUX/IAA	DR	-2.6	0.05
DN28734_c0_g2	AUX/IAA	DR	-2.68	-0.20
DN28119_c0_g1	AUX/IAA	DR	-2.76	-0.30
DN17650_c0_g1	AUX/IAA	DR	-2.89	-0.20
DN43460_c0_g1	AUX/IAA	DR	-3.00	-0.20
DN6428_c0_g1	AUX/IAA	DR	-3.01	-1.00
DN23260_c0_g1	AUX/IAA	DR	-3.03	-0.80
DN33227_c1_g2	AUX/IAA	DR	-3.26	-1.00
DN37861_c3_g1	AUX/IAA	DR	-3.29	-1.00
DN50579_c0_g1	AUX/IAA	DR	-3.77	-1.00
DN4097_c0_g1	AUX/IAA	DR	-3.87	-2.00
DN11727_c0_g1	AUX/IAA	DR	-4.03	-1.00
DN43460_c0_g2	AUX/IAA	DR	-4.33	-2.00
DN28124_c0_g1	AUX/IAA	DR	-7.79	-6.00
DN61327_c0_g1	AUX/IAA	DR	-1.84	-0.30
DN33758_c0_g1	AUX/IAA	DR	-1.85	-0.80
DN8760_c0_g1	AUX/IAA	DR	-1.94	-1.00
DN39780_c0_g1	AUX/IAA	DR	-2.04	-0.20

Continuação

Transcrito	Família CoREG	Expressão	Log <sub>2</sub> FC Jc171	Log <sub>2</sub> FC Jc183
DN39043_c1_g1	AUX/IAA	DR	-2.07	-0.20
DN38613_c0_g2	AUX/IAA	DR	-2.18	0.50
DN43568_c0_g1	AUX/IAA	DR	-2.24	-0.50
DN43546_c0_g1	AUX/IAA	DR	-2.27	-0.01
DN14416_c0_g1	AUX/IAA	DR	-2.28	-0.30
DN32868_c0_g1	AUX/IAA	DR	-2.34	-0.90
DN25111_c0_g1	AUX/IAA	DR	-2.38	-0.70
DN43651_c0_g1	AUX/IAA	DR	-2.54	0.20
DN8770_c0_g1	AUX/IAA	DR	-2.55	-0.60
DN14473_c0_g1	AUX/IAA	DR	-2.56	-2.00
DN37861_c1_g1	AUX/IAA	DR	-2.58	-0.80
DN38174_c2_g1	AUX/IAA	DR	-2.58	-0.60
DN41780_c0_g1	AUX/IAA	DR	-2.59	-1.00
DN43509_c0_g1	AUX/IAA	DR	-2.62	-1.00
DN43283_c2_g3	AUX/IAA	DR	-2.66	-0.90
DN50382_c0_g1	AUX/IAA	DR	-2.68	-0.10
DN43384_c2_g2	AUX/IAA	DR	-2.81	-0.40
DN43418_c0_g2	AUX/IAA	DR	-2.86	1.00
DN30995_c0_g1	AUX/IAA	DR	-2.92	-1.00
DN49720_c0_g1	AUX/IAA	DR	-2.95	-1.00
DN25971_c0_g1	AUX/IAA	DR	-2.99	-0.30
DN38902_c0_g1	AUX/IAA	DR	-3.03	0.20
DN32987_c0_g1	AUX/IAA	DR	-3.06	-0.30
DN37861_c2_g1	AUX/IAA	DR	-3.06	-0.80
DN41598_c0_g1	AUX/IAA	DR	-3.1	-1.00
DN42816_c0_g1	AUX/IAA	DR	-3.19	-1.00
DN42915_c1_g2	AUX/IAA	DR	-3.53	-0.80
DN43578_c1_g1	AUX/IAA	DR	-3.68	0.60
DN11351_c0_g3	AUX/IAA	DR	-4.27	0.01
DN43384_c3_g1	AUX/IAA	DR	-4.85	0.50
DN42482_c0_g1	AUX/IAA	DR	-5.05	-2.00
DN11145_c0_g2	AUX/IAA	DR	-8.61	-0.60
DN85164_c0_g5	AUX/IAA	DR	-8.96	-0.90
DN25853_c0_g1	AUX/IAA	DR	-9.32	0.50
DN34126_c1_g6	AUX/IAA	DR	-11.47	-10.00
DN14473_c0_g2	AUX/IAA	DR	-11.76	-2.00
DN27263_c0_g1	DDT	DR	-2.88	-0.40
DN41315_c0_g1	GNAT	UR	1.51	1.00
DN6455_c0_g1	GNAT	DR	-2.2	-0.10
DN43631_c0_g1	GNAT	DR	-3.25	-0.60
DN40018_c0_g1	IWS1	DR	-2.22	-0.04
DN43464_c4_g1	IWS1	DR	-1.84	-0.20
DN26119_c0_g1	IWS1	DR	-3.27	-0.80
DN98280_c0_g1	Jumonji	UR	2.42	0.10

Continuação

<b>Transcrito</b>	<b>Família CoREG</b>	<b>Expressão</b>	<b>Log<sub>2</sub>FC Jc171</b>	<b>Log<sub>2</sub>FC Jc183</b>
DN43558_c0_g1	Jumonji	DR	-2.75	-0.10
DN32843_c0_g1	Jumonji	DR	-2.8	-0.50
DN41234_c0_g2	Jumonji	DR	-3.04	-0.10
DN38633_c1_g2	Jumonji	DR	-3.12	-0.40
DN36764_c0_g2	Jumonji	DR	-3.19	-0.30
DN40616_c0_g1	Jumonji	DR	-3.21	-0.20
DN43138_c0_g1	Jumonji	DR	-3.38	0.10
DN42406_c1_g1	Jumonji	DR	-3.99	-0.06
DN27416_c0_g2	Jumonji	UR	7.16	6.00
DN41972_c0_g1	Jumonji	UR	4.82	1.00
DN12540_c0_g1	Jumonji	UR	2.54	0.60
DN43103_c2_g1	Jumonji	DR	-2.09	-0.80
DN43429_c1_g1	Jumonji	DR	-2.4	0.10
DN41771_c0_g1	Jumonji	DR	-2.52	0.01
DN42484_c0_g1	Jumonji	DR	-2.54	-0.30
DN41834_c0_g1	Jumonji	DR	-2.6	-0.30
DN33207_c0_g2	Jumonji	DR	-2.67	0.30
DN39021_c0_g2	Jumonji	DR	-2.68	-0.40
DN43141_c0_g1	Jumonji	DR	-2.68	-0.20
DN42406_c2_g1	Jumonji	DR	-2.7	-0.50
DN42921_c0_g2	Jumonji	DR	-2.7	-0.30
DN74634_c0_g1	Jumonji	DR	-2.71	-0.40
DN23662_c0_g2	Jumonji	DR	-2.73	-0.20
DN43329_c0_g1	Jumonji	DR	-2.75	-0.09
DN43550_c1_g1	Jumonji	DR	-2.75	0.01
DN43288_c0_g1	Jumonji	DR	-2.79	-0.50
DN40546_c3_g4	Jumonji	DR	-2.94	-0.50
DN85482_c0_g1	Jumonji	DR	-3	0.50
DN40162_c0_g1	Jumonji	DR	-3.06	-0.10
DN36390_c0_g2	Jumonji	DR	-3.09	-0.50
DN37202_c0_g1	Jumonji	DR	-3.19	-0.40
DN26132_c0_g1	Jumonji	DR	-3.26	-0.30
DN39172_c0_g1	Jumonji	DR	-3.36	0.04
DN15239_c0_g2	Jumonji	DR	-3.38	-0.20
DN8929_c0_g1	Jumonji	DR	-3.39	-0.20
DN36766_c0_g1	Jumonji	DR	-3.4	1.00
DN37558_c0_g1	Jumonji	DR	-3.46	-0.20
DN40716_c0_g1	Jumonji	DR	-3.54	-1.00
DN50702_c0_g1	Jumonji	DR	-3.63	-0.80
DN42166_c0_g1	Jumonji	DR	-3.79	-0.50
DN42522_c0_g1	Jumonji	DR	-3.97	0.01
DN42445_c0_g1	Jumonji	DR	-4.13	-0.70
DN37276_c0_g2	Jumonji	DR	-4.24	0.40
DN29733_c0_g3	Jumonji	DR	-4.34	1.00

Continuação

<b>Transcrito</b>	<b>Família CoREG</b>	<b>Expressão</b>	<b>Log<sub>2</sub>FC Jc171</b>	<b>Log<sub>2</sub>FC Jc183</b>
DN36849_c0_g1	Jumonji	DR	-4.6	-0.30
DN38510_c0_g1	Jumonji	DR	-5.17	-1.00
DN9364_c0_g1	Jumonji	DR	-7.43	-0.60
DN27262_c0_g1	Jumonji	DR	-7.66	0.50
DN27416_c0_g1	Jumonji	DR	-7.86	-1.00
DN14177_c0_g1	Jumonji	DR	-8.38	1.00
DN267_c0_g1	Jumonji	DR	-11.18	-0.70
DN42993_c0_g1	LUG	UR	3.64	3.00
DN37406_c0_g1	LUG	DR	-2	-0.60
DN8711_c0_g1	LUG	DR	-2.12	NA
DN42878_c0_g1	LUG	DR	-2.36	-0.30
DN39981_c0_g1	LUG	DR	-2.51	-0.50
DN62740_c0_g1	LUG	DR	-2.78	0.30
DN36197_c0_g1	LUG	DR	-3.37	-1.00
DN52256_c0_g1	LUG	DR	-4.9	-1.00
DN53612_c0_g1	MBF1	UR	6.11	0.60
DN97713_c0_g1	mBF1	UR	4.06	0.60
DN85308_c0_g1	MBF1	UR	1.69	0.40
DN43954_c0_g1	mTERF	DR	-2.19	-0.70
DN41924_c0_g2	mTERF	DR	-2.38	-0.10
DN43507_c1_g1	mTERF	DR	-2.83	-0.90
DN8742_c0_g1	mTERF	DR	-2.85	-0.80
DN35083_c1_g1	mTERF	DR	-2.93	-0.40
DN40814_c0_g2	mTERF	DR	-4.28	-0.07
DN10241_c0_g1	mTERF	DR	-4.48	-2.00
DN16475_c0_g1	Others	UR	7.76	-0.30
DN9185_c0_g1	Others	UR	6.55	2.00
DN61057_c1_g1	Others	UR	1.65	0.09
DN49875_c0_g1	Others	DR	-2.68	-0.60
DN25072_c0_g1	Others	DR	-2.95	-0.30
DN33300_c0_g1	Others	DR	-2.35	-0.70
DN8990_c0_g1	Others	DR	-2.35	-0.02
DN37295_c0_g1	Others	DR	-2.49	0.00
DN36253_c0_g1	Others	DR	-2.67	-1.00
DN32791_c0_g1	Others	DR	-2.68	-0.30
DN38834_c0_g1	Others	DR	-2.9	-0.30
DN34181_c0_g2	Others	DR	-3.11	-0.70
DN37792_c1_g1	Others	DR	-3.33	-0.80
DN6442_c0_g1	Others	DR	-3.42	-1.00
DN17175_c0_g1	Others	UR	6.74	0.10
DN100970_c0_g1	Others	UR	2.72	0.30
DN1101_c0_g1	Others	UR	2.5	1.00
DN61235_c0_g1	Others	UR	2.25	-0.09
DN33043_c1_g3	Others	UR	1.87	0.40

Continuação

Transcrito	Família CoREG	Expressão	Log <sub>2</sub> FC Jc171	Log <sub>2</sub> FC Jc183
DN35570_c0_g1	PHD	DR	-2.16	-0.40
DN25851_c0_g1	PHD	DR	-2.23	-0.30
DN35870_c0_g1	PHD	DR	-2.32	-0.40
DN42396_c0_g1	PHD	DR	-2.39	-0.40
DN24918_c0_g1	PHD	DR	-2.45	-0.10
DN37912_c0_g1	PHD	DR	-2.49	-0.04
DN41870_c0_g1	PHD	DR	-2.59	-0.40
DN40865_c1_g1	PHD	DR	-2.65	-0.10
DN6525_c0_g1	PHD	DR	-2.7	-0.70
DN36436_c0_g1	PHD	DR	-2.72	-0.30
DN38001_c1_g1	PHD	DR	-2.77	-0.20
DN43335_c0_g1	PHD	DR	-2.77	-0.10
DN35033_c0_g1	PHD	DR	-2.82	-0.70
DN36416_c1_g1	PHD	DR	-2.82	-0.40
DN26519_c0_g1	PHD	DR	-2.85	-0.20
DN40275_c0_g1	PHD	DR	-2.87	-0.90
DN74365_c0_g1	PHD	DR	-2.88	-0.10
DN42530_c0_g1	PHD	DR	-2.91	-0.70
DN38412_c0_g1	PHD	DR	-2.93	-0.30
DN38534_c0_g1	PHD	DR	-2.96	-0.03
DN40251_c0_g1	PHD	DR	-3.15	-1.00
DN33798_c0_g1	PHD	DR	-3.21	-0.50
DN41921_c0_g1	PHD	DR	-3.22	-0.20
DN36397_c0_g1	PHD	DR	-3.26	-0.90
DN40865_c0_g1	PHD	DR	-3.69	-1.00
DN43335_c0_g2	PHD	DR	-4.96	0.80
DN41339_c0_g1	PHD	DR	-6.64	0.60
DN6345_c0_g1	PHD	DR	-7.34	-0.60
DN86030_c0_g1	Pseudo ARR-B	DR	-4.42	-0.50
DN43355_c1_g1	RB	DR	-2.63	-0.50
DN27469_c0_g1	SET	UR	1.48	-0.08
DN30645_c0_g2	SET	DR	-2.23	-0.20
DN12076_c0_g2	SET	DR	-2.46	-0.01
DN26419_c0_g1	SET	DR	-2.63	-0.60
DN27423_c0_g2	SET	DR	-5.43	-0.30
DN50086_c0_g1	SNF2	DR	-2.06	-0.30
DN36373_c0_g1	SNF2	DR	-2.38	-0.30
DN43259_c0_g2	SNF2	DR	-2.5	-0.50
DN27418_c0_g1	SNF2	DR	-2.57	-0.60
DN38606_c0_g1	SNF2	DR	-2.7	-0.60
DN40579_c0_g2	SNF2	DR	-2.75	-0.60
DN43635_c1_g1	SNF2	DR	-2.76	-0.60
DN62351_c0_g1	SNF2	DR	-2.79	-0.40
DN34178_c0_g1	SNF2	DR	-2.93	-0.30

Continuação

<b>Transcrito</b>	<b>Família CoREG</b>	<b>Expressão</b>	<b>Log<sub>2</sub>FC Jc171</b>	<b>Log<sub>2</sub>FC Jc183</b>
DN61267_c0_g2	SNF2	DR	-3.01	-0.60
DN28188_c0_g1	SNF2	DR	-3.73	-0.30
DN34357_c0_g2	SNF2	DR	-10.72	-0.60
DN31255_c0_g2	SNF2	DR	-11.62	-2.00
DN38574_c0_g1	SWI/SNF-BAF60b	UR	2.44	0.08
DN23170_c0_g1	SWI/SNF-BAF60b	UR	1.99	0.30
DN8859_c0_g1	SWI/SNF-BAF60b	DR	-2.4	-0.30
DN32942_c0_g1	SWI/SNF-BAF60b	DR	-2.59	-0.20
DN25832_c0_g1	SWI/SNF-SWI3	DR	-2.37	-1.00
DN41946_c0_g1	SWI/SNF-SWI3	DR	-3.02	-0.70
DN73525_c0_g1	TAZ	DR	-2.53	-0.50
DN38623_c0_g1	TAZ	DR	-3.1	-0.60
DN61294_c0_g1	TRAF	UR	2.19	0.30
DN97665_c0_g1	TRAF	DR	-1.81	-0.50
DN39096_c0_g1	TRAF	DR	-2.2	-0.10
DN40414_c0_g1	TRAF	DR	-2.46	-0.20
DN49388_c0_g1	TRAF	DR	-2.49	-0.80
DN37234_c0_g1	TRAF	DR	-2.54	-0.60
DN9067_c0_g1	TRAF	DR	-2.59	-0.70
DN40414_c3_g1	TRAF	DR	-4.07	-0.80

## APÊNDICE C

**Apêndice C.** Termos GO enriquecidos para CoREGs UR, DR e não DEs ( $p$ -value 0.001) evidenciando os termos comuns e exclusivos de acordo com o diagrama de Venn (Figura 9). Identificadores dos termos GO presente na primeira coluna da esquerda para direita. Descrição dos termos, quantidade de anotações nos bancos do PlantRegMap, presença na em número de CoREGs diferentes e aspecto (P: processo biológico; C: Componente celular; F: Função molecular) Há quantidades maiores de termos na tabela devido as análises terem sido feitas três análises, assim um mesmo termo GO possui estatísticas diferentes para cada uma das análises. 55 termos GO para o agrupamento “Comum entre todos”, sem duplicatas, resultam em 15 termos GO originais.

GO.ID	Agrupamento	Termo	Anotações	Presença	$p$ -value	$q$ -value	Aspecto
GO:0009755	Comum entre todos	Hormone-mediated signaling pathway	244	4	1.20E-05	1.66E-02	P
GO:0009734	Comum entre todos	Auxin-activated signaling pathway	79	3	1.50E-05	1.66E-02	P
GO:0032870	Comum entre todos	Cellular response to hormone stimulus	279	4	2.00E-05	1.66E-02	P
GO:0071495	Comum entre todos	Cellular response to endogenous stimulus	283	4	2.10E-05	1.66E-02	P
GO:0071365	Comum entre todos	Cellular response to auxin stimulus	90	3	2.20E-05	1.66E-02	P
GO:0007165	Comum entre todos	Signal transduction	651	5	2.70E-05	1.66E-02	P
GO:0044700	Comum entre todos	Single organism signaling	655	5	2.80E-05	1.66E-02	P
GO:0023052	Comum entre todos	Signaling	656	5	2.80E-05	1.66E-02	P
GO:0071310	Comum entre todos	Cellular response to organic substance	360	4	5.30E-05	2.80E-02	P
GO:0009725	Comum entre todos	Response to hormone	817	5	8.00E-05	3.54E-02	P
GO:0007154	Comum entre todos	Cell communication	820	5	8.20E-05	3.54E-02	P
GO:0070887	Comum entre todos	Cellular response to chemical stimulus	429	4	0.00011	4.02E-02	P
GO:0009719	Comum entre todos	Response to endogenous stimulus	880	5	0.00011	4.02E-02	P
GO:0010033	Comum entre todos	Response to organic substance	1026	5	0.00024	8.14E-02	P
GO:0051716	Comum entre todos	Cellular response to stimulus	1162	5	0.00043	1.36E-01	P
GO:0009733	Comum entre todos	Response to auxin	264	3	0.00053	1.57E-01	P
GO:0005634	Comum entre todos	Nucleus	2201	8	0.00029	1.97E-01	C
GO:0005515	Comum entre todos	Protein binding	4177	11	1.30E-05	1.51E-02	F
GO:0007165	Comum entre todos	Signal transduction	651	23	2.20E-09	2.82E-07	P
GO:0044700	Comum entre todos	Single organism signaling	655	23	2.50E-09	3.12E-07	P
GO:0023052	Comum entre todos	Signaling	656	23	2.60E-09	3.16E-07	P
GO:0051716	Comum entre todos	Cellular response to stimulus	1162	30	9.30E-09	8.49E-07	P

Continuação

GO.ID	Agrupamento	Termo	Anotações	Presença	p-value	q-value	Aspecto
GO:0007154	Comum entre todos	Cell communication	820	24	3.70E-08	2.83E-06	P
GO:0071310	Comum entre todos	Cellular response to organic substance	360	15	2.20E-07	1.32E-05	P
GO:0070887	Comum entre todos	Cellular response to chemical stimulus	429	16	3.80E-07	2.23E-05	P
GO:0032870	Comum entre todos	Cellular response to hormone stimulus	279	13	4.30E-07	2.46E-05	P
GO:0071495	Comum entre todos	Cellular response to endogenous stimulus	283	13	5.00E-07	2.79E-05	P
GO:0009755	Comum entre todos	Hormone-mediated signaling pathway	244	11	4.80E-06	2.11E-04	P
GO:0010033	Comum entre todos	Response to organic substance	1026	21	7.40E-05	2.55E-03	P
GO:0009719	Comum entre todos	Response to endogenous stimulus	880	19	8.50E-05	2.86E-03	P
GO:0009725	Comum entre todos	Response to hormone	817	18	1.00E-04	3.30E-03	P
GO:0009733	Comum entre todos	Response to auxin	264	9	3.00E-04	8.96E-03	P
GO:0009734	Comum entre todos	Auxin-activated signaling pathway	79	5	0.00045	1.23E-02	P
GO:0071365	Comum entre todos	Cellular response to auxin stimulus	90	5	0.00082	2.05E-02	P
GO:0005634	Comum entre todos	Nucleus	2201	38	9.30E-05	3.16E-02	C
GO:0005515	Comum entre todos	Protein binding	4177	147	1.00E-30	2.32E-27	F
GO:0044700	Comum entre todos	Single organism signaling	655	67	3.30E-21	5.40E-19	P
GO:0023052	Comum entre todos	Signaling	656	67	3.60E-21	5.70E-19	P
GO:0007165	Comum entre todos	Signal transduction	651	66	1.10E-20	1.49E-18	P
GO:0007154	Comum entre todos	Cell communication	820	70	8.70E-18	7.37E-16	P
GO:0009755	Comum entre todos	Hormone-mediated signaling pathway	244	37	1.50E-17	1.25E-15	P
GO:0032870	Comum entre todos	Cellular response to hormone stimulus	279	38	2.10E-16	1.69E-14	P
GO:0070887	Comum entre todos	Cellular response to chemical stimulus	429	47	3.20E-16	2.49E-14	P
GO:0071495	Comum entre todos	Cellular response to endogenous stimulus	283	38	3.50E-16	2.68E-14	P
GO:0051716	Comum entre todos	Cellular response to stimulus	1162	81	2.40E-15	1.56E-13	P
GO:0071310	Comum entre todos	Cellular response to organic substance	360	41	7.60E-15	4.75E-13	P
GO:0009734	Comum entre todos	Auxin-activated signaling pathway	79	20	2.00E-14	1.20E-12	P
GO:0071365	Comum entre todos	Cellular response to auxin stimulus	90	20	3.00E-13	1.62E-11	P
GO:0009719	Comum entre todos	Response to endogenous stimulus	880	60	4.30E-11	2.10E-09	P
GO:0009725	Comum entre todos	Response to hormone	817	57	5.80E-11	2.78E-09	P

Continuação

GO.ID	Agrupamento	Termo	Anotações	Presença	p-value	q-value	Aspecto
GO:0009733	Comum entre todos	Response to auxin	264	24	2.70E-07	1.03E-05	P
GO:0005634	Comum entre todos	Nucleus	2201	127	4.60E-09	3.12E-06	C
GO:0005515	Comum entre todos	Protein binding	4177	638	1.00E-30	1.16E-27	F
GO:0080117	Exclusivo DR	Secondary growth	6	4	6.10E-08	4.20E-06	P
GO:0080190	Exclusivo DR	Lateral growth	6	4	6.10E-08	4.20E-06	P
GO:0048569	Exclusivo DR	Post-embryonic organ development	215	12	1.80E-07	1.12E-05	P
GO:0003002	Exclusivo DR	Regionalization	117	9	4.50E-07	2.54E-05	P
GO:0009653	Exclusivo DR	Anatomical structure morphogenesis	528	17	1.20E-06	5.93E-05	P
GO:0007389	Exclusivo DR	Pattern specification process	136	9	1.60E-06	7.67E-05	P
GO:0009888	Exclusivo DR	Tissue development	378	14	2.30E-06	1.09E-04	P
GO:0040007	Exclusivo DR	Growth	341	13	4.00E-06	1.79E-04	P
GO:0045892	Exclusivo DR	Negative regulation of transcription, DNA-templated	115	8	4.30E-06	1.91E-04	P
GO:1902679	Exclusivo DR	Negative regulation of RNA biosynthetic process	119	8	5.50E-06	2.37E-04	P
GO:1903507	Exclusivo DR	Negative regulation of nucleic acid-templated transcription	119	8	5.50E-06	2.37E-04	P
GO:0051253	Exclusivo DR	Negative regulation of RNA metabolic process	121	8	6.30E-06	2.69E-04	P
GO:0045814	Exclusivo DR	Negative regulation of gene expression, epigenetic	33	5	6.40E-06	2.71E-04	P
GO:0045934	Exclusivo DR	Negative regulation of nucleobase-containing compound metabolic process	136	8	1.50E-05	6.19E-04	P
GO:0048646	Exclusivo DR	Anatomical structure formation involved in morphogenesis	136	8	1.50E-05	6.19E-04	P
GO:0097306	Exclusivo DR	Cellular response to alcohol	100	7	1.70E-05	6.96E-04	P
GO:0009892	Exclusivo DR	Negative regulation of metabolic process	396	13	2.00E-05	8.04E-04	P
GO:2000113	Exclusivo DR	Negative regulation of cellular macromolecule biosynthetic process	142	8	2.00E-05	8.04E-04	P
GO:0048437	Exclusivo DR	Floral organ development	143	8	2.10E-05	8.38E-04	P
GO:0010558	Exclusivo DR	Negative regulation of macromolecule biosynthetic process	146	8	2.50E-05	9.89E-04	P
GO:0016458	Exclusivo DR	Gene silencing	110	7	3.10E-05	1.22E-03	P
GO:0048645	Exclusivo DR	Organ formation	76	6	3.50E-05	1.35E-03	P
GO:0031327	Exclusivo DR	Negative regulation of cellular biosynthetic process	154	8	3.60E-05	1.38E-03	P
GO:0009890	Exclusivo DR	Negative regulation of biosynthetic process	157	8	4.20E-05	1.58E-03	P
GO:0051172	Exclusivo DR	Negative regulation of nitrogen compound metabolic process	157	8	4.20E-05	1.58E-03	P

## Continuação

GO.ID	Agrupamento	Termo	Anotações	Presença	p-value	q-value	Aspecto
GO:0010271	Exclusivo DR	Regulation of chlorophyll catabolic process	2	2	6.50E-05	2.25E-03	P
GO:0034757	Exclusivo DR	Negative regulation of iron ion transport	2	2	6.50E-05	2.25E-03	P
GO:0040010	Exclusivo DR	Positive regulation of growth rate	2	2	6.50E-05	2.25E-03	P
GO:1901404	Exclusivo DR	Regulation of tetrapyrrole catabolic process	2	2	6.50E-05	2.25E-03	P
GO:0006342	Exclusivo DR	Chromatin silencing	29	4	8.40E-05	2.86E-03	P
GO:0031324	Exclusivo DR	Negative regulation of cellular metabolic process	223	9	8.50E-05	2.86E-03	P
GO:0097305	Exclusivo DR	Response to alcohol	341	11	1.00E-04	3.30E-03	P
GO:0045927	Exclusivo DR	Positive regulation of growth	12	3	0.00011	3.58E-03	P
GO:0071496	Exclusivo DR	Cellular response to external stimulus	94	6	0.00011	3.58E-03	P
GO:0048827	Exclusivo DR	Phyllome development	298	10	0.00016	5.10E-03	P
GO:0010272	Exclusivo DR	Response to silver ion	3	2	0.00019	5.93E-03	P
GO:0034756	Exclusivo DR	Regulation of iron ion transport	3	2	0.00019	5.93E-03	P
GO:0071217	Exclusivo DR	Cellular response to external biotic stimulus	3	2	0.00019	5.93E-03	P
GO:0008152	Exclusivo DR	Metabolic process	12173	114	0.00026	7.91E-03	P
GO:0016036	Exclusivo DR	Cellular response to phosphate starvation	40	4	3.00E-04	8.96E-03	P
GO:0009886	Exclusivo DR	Post-embryonic morphogenesis	114	6	0.00033	9.67E-03	P
GO:0042594	Exclusivo DR	Response to starvation	74	5	0.00033	9.67E-03	P
GO:0009584	Exclusivo DR	Detection of visible light	4	2	0.00039	1.09E-02	P
GO:0040009	Exclusivo DR	Regulation of growth rate	4	2	0.00039	1.09E-02	P
GO:0043966	Exclusivo DR	Histone H3 acetylation	4	2	0.00039	1.09E-02	P
GO:0046898	Exclusivo DR	Response to cycloheximide	4	2	0.00039	1.09E-02	P
GO:0080001	Exclusivo DR	Mucilage extrusion from seed coat	4	2	0.00039	1.09E-02	P
GO:1901654	Exclusivo DR	Response to ketone	4	2	0.00039	1.09E-02	P
GO:0071215	Exclusivo DR	Cellular response to abscisic acid stimulus	79	5	0.00045	1.23E-02	P
GO:0071229	Exclusivo DR	Cellular response to acid chemical	171	7	0.00049	1.33E-02	P
GO:0010087	Exclusivo DR	Phloem or xylem histogenesis	47	4	0.00056	1.51E-02	P
GO:0010029	Exclusivo DR	Regulation of seed germination	48	4	0.00061	1.63E-02	P
GO:0043271	Exclusivo DR	Negative regulation of ion transport	5	2	0.00064	1.65E-02	P

Continuação

GO.ID	Agrupamento	Termo	Anotações	Presença	p-value	q-value	Aspecto
GO:1902074	Exclusivo DR	Response to salt	5	2	0.00064	1.65E-02	P
GO:2000069	Exclusivo DR	Regulation of post-embryonic root development	5	2	0.00064	1.65E-02	P
GO:0051049	Exclusivo DR	Regulation of transport	49	4	0.00066	1.68E-02	P
GO:1900140	Exclusivo DR	Regulation of seedling development	49	4	0.00066	1.68E-02	P
GO:0010073	Exclusivo DR	Meristem maintenance	86	5	0.00067	1.70E-02	P
GO:0060968	Exclusivo DR	Regulation of gene silencing	23	3	0.00082	2.05E-02	P
GO:0017006	Exclusivo DR	Protein-tetrapyrrole linkage	6	2	0.00096	2.36E-02	P
GO:0032776	Exclusivo DR	DNA methylation on cytosine	6	2	0.00096	2.36E-02	P
GO:0031667	Exclusivo DR	Response to nutrient levels	94	5	0.001	2.45E-02	P
GO:0005886	Exclusivo DR	Plasma membrane	1155	25	7.30E-05	3.16E-02	C
GO:0071944	Exclusivo DR	Cell periphery	1488	28	0.00028	6.34E-02	C
GO:0043168	Exclusivo DR	Anion binding	2965	67	6.30E-14	7.31E-12	F
GO:0000166	Exclusivo DR	Nucleotide binding	3115	67	6.70E-13	6.48E-11	F
GO:1901265	Exclusivo DR	Nucleoside phosphate binding	3115	67	6.70E-13	6.48E-11	F
GO:0036094	Exclusivo DR	Small molecule binding	3181	67	1.80E-12	1.61E-10	F
GO:0004888	Exclusivo DR	Transmembrane signaling receptor activity	24	7	1.20E-09	8.99E-08	F
GO:0005034	Exclusivo DR	Osmosensor activity	4	4	6.00E-09	4.22E-07	F
GO:0019199	Exclusivo DR	Transmembrane receptor protein kinase activity	4	3	2.70E-06	1.59E-04	F
GO:0004675	Exclusivo DR	Transmembrane receptor protein serine/threonine kinase activity	2	2	7.90E-05	3.99E-03	F
GO:0043424	Exclusivo DR	Protein histidine kinase binding	12	3	0.00014	6.50E-03	F
GO:0019900	Exclusivo DR	Kinase binding	36	4	0.00028	1.25E-02	F
GO:0042802	Exclusivo DR	Identical protein binding	163	7	0.00065	2.69E-02	F
GO:0060090	Exclusivo UR	Binding, bridging	6	2	7.30E-06	1.51E-02	F
GO:0003713	Comum UR e não DE	Transcription coactivator activity	10	2	2.20E-05	1.70E-02	F
GO:0003712	Comum UR e não DE	Transcription cofactor activity	43	2	0.00043	2.09E-01	F
GO:0000989	Comum UR e não DE	Transcription factor activity, transcription factor binding	44	2	0.00045	2.09E-01	F
GO:0000988	Comum UR e não DE	Transcription factor activity, protein binding	51	2	0.00061	2.09E-01	F
GO:0003712	Comum UR e não DE	Transcription cofactor activity	43	9	4.40E-05	2.22E-03	F

Continuação

<b>GO.ID</b>	<b>Agrupamento</b>	<b>Termo</b>	<b>Anotações</b>	<b>Presença</b>	<b>p-value</b>	<b>q-value</b>	<b>Aspecto</b>
GO:0000988	Comum UR e não DE	Transcription factor activity, protein binding	51	9	0.00018	8.19E-03	F
GO:0003713	Comum UR e não DE	Transcription coactivator activity	10	4	0.00045	1.87E-02	F

## APÊNDICE D

**Apêndice D.** Análise com genes modelo associados aos Unigenes UR totais das bibliotecas RNA-Seq para FTs enriquecidos (88) em conjunto de 1660 alvos pesquisados. *p-value* 0.001. Background indica a quantidade total de genes no banco de dados.

FT ID.	Background	Interações previstas	Interações	<i>p_value</i>	<i>q_value</i>
Jcr4S00146.190	57437	977	129	9.10E-48	2.75E-45
Jcr4S14409.20	57437	1007	127	7.51E-45	1.13E-42
Jcr4S00022.240	57437	1041	124	3.36E-41	3.39E-39
Jcr4S05882.20	57437	964	118	1.89E-40	1.43E-38
Jcr4S03153.40	57437	991	119	6.10E-40	3.68E-38
Jcr4S00909.50	57437	1165	105	5.88E-25	2.96E-23
Jcr4S00969.70	57437	756	77	6.40E-22	2.76E-20
Jcr4S04936.40	57437	974	88	2.97E-21	1.12E-19
Jcr4S13202.30	57437	9095	401	4.39E-19	1.47E-17
Jcr4S00027.270	57437	907	79	2.51E-18	7.59E-17
Jcr4S00945.40	57437	918	79	5.06E-18	1.39E-16
Jcr4S00361.90	57437	964	81	8.07E-18	2.03E-16
Jcr4S01040.60	57437	12026	493	1.24E-17	2.88E-16
Jcr4S09941.10	57437	942	79	2.24E-17	4.83E-16
Jcr4S00524.90	57437	1063	82	7.30E-16	1.47E-14
Jcr4S02641.40	57437	872	72	1.20E-15	2.27E-14
Jcr4S04332.40	57437	594	57	1.43E-15	2.54E-14
Jcr4S00028.40	57437	881	72	2.04E-15	3.42E-14
Jcr4S04809.10	57437	14149	550	2.68E-15	4.26E-14
Jcr4S01086.80	57437	816	68	4.46E-15	6.74E-14
Jcr4S00150.190	57437	1024	76	5.82E-14	8.37E-13
Jcr4S00242.50	57437	999	74	1.34E-13	1.84E-12
Jcr4S01323.60	57437	1132	80	1.79E-13	2.26E-12
Jcr4S28273.10	57437	1132	80	1.79E-13	2.26E-12
Jcr4S01276.40	57437	671	57	2.72E-13	3.29E-12
Jcr4S04680.20	57437	11897	465	5.57E-13	6.47E-12
Jcr4S03743.20	57437	961	68	9.63E-12	1.08E-10
Jcr4S00197.100	57437	473	43	1.85E-11	2.00E-10
Jcr4S00021.40	57437	676	53	3.62E-11	3.77E-10
Jcr4S22404.10	57437	784	57	1.39E-10	1.40E-09
Jcr4S00197.90	57437	926	63	2.57E-10	2.51E-09
Jcr4S05150.10	57437	721	53	3.83E-10	3.61E-09
Jcr4S00783.40	57437	677	49	2.69E-09	2.46E-08
Jcr4S00433.80	57437	7830	309	5.59E-09	4.97E-08
Jcr4S00253.40	57437	620	44	2.75E-08	2.38E-07
Jcr4S02081.30	57437	738	48	1.07E-07	8.95E-07
Jcr4S05308.30	57437	4474	188	1.35E-07	1.10E-06
Jcr4S11123.10	57437	4912	200	4.85E-07	3.85E-06
Jcr4S05725.30	57437	4402	182	6.18E-07	4.78E-06
Jcr4S05837.30	57437	5316	213	6.34E-07	4.79E-06
Jcr4S00717.70	57437	739	46	6.51E-07	4.80E-06
Jcr4S11652.10	57437	525	36	9.23E-07	6.64E-06

Continuação

<b>FT ID.</b>	<b>Background</b>	<b>Interações previstas</b>	<b>Interações</b>	<b>p_value</b>	<b>q_value</b>
Jcr4S00700.20	57437	9865	360	1.02E-06	6.98E-06
Jcr4S13035.30	57437	10464	379	1.02E-06	6.98E-06
Jcr4S05114.10	57437	10022	362	2.50E-06	1.68E-05
Jcr4S08087.10	57437	490	33	3.51E-06	2.31E-05
Jcr4S01332.10	57437	494	33	4.18E-06	2.69E-05
Jcr4S00344.40	57437	5450	212	5.00E-06	3.14E-05
Jcr4S00911.80	57437	235	20	6.09E-06	3.75E-05
Jcr4S01477.90	57437	505	33	6.69E-06	4.04E-05
Jcr4S00420.40	57437	3857	157	8.54E-06	5.05E-05
Jcr4S00002.200	57437	222	19	8.90E-06	5.17E-05
Jcr4S03542.40	57437	7438	275	9.67E-06	5.51E-05
Jcr4S05581.10	57437	10112	360	9.99E-06	5.59E-05
Jcr4S02384.70	57437	518	33	1.14E-05	6.25E-05
Jcr4S09760.20	57437	4640	182	1.49E-05	8.04E-05
Jcr4S01064.30	57437	1211	61	1.56E-05	8.27E-05
Jcr4S02861.50	57437	1011	53	1.74E-05	8.84E-05
Jcr4S00137.180	57437	6446	241	1.76E-05	8.84E-05
Jcr4S01985.20	57437	6446	241	1.76E-05	8.84E-05
Jcr4S03169.20	57437	4005	160	1.90E-05	9.41E-05
Jcr4S00037.60	57437	5002	193	2.17E-05	1.06E-04
Jcr4S01276.110	57437	3752	151	2.24E-05	1.07E-04
Jcr4S03656.30	57437	1293	63	3.16E-05	1.48E-04
Jcr4S05684.30	57437	345	24	3.22E-05	1.48E-04
Jcr4S02864.30	57437	3310	135	3.23E-05	1.48E-04
Jcr4S00044.40	57437	4527	175	4.74E-05	2.14E-04
Jcr4S00968.20	57437	6057	225	5.07E-05	2.25E-04
Jcr4S00001.70	57437	5741	214	6.37E-05	2.79E-04
Jcr4S00053.270	57437	4584	175	9.07E-05	3.91E-04
Jcr4S00401.40	57437	673	37	9.55E-05	4.06E-04
Jcr4S00118.70	57437	1816	79	1.76E-04	7.38E-04
Jcr4S11106.10	57437	673	36	1.98E-04	8.18E-04
Jcr4S06462.20	57437	652	35	2.21E-04	9.00E-04
Jcr4S04046.10	57437	9441	327	2.23E-04	9.00E-04
Jcr4S01288.20	57437	3680	142	2.55E-04	1.01E-03
Jcr4S00009.60	57437	5550	202	3.65E-04	1.43E-03
Jcr4S00112.130	57437	474	27	3.69E-04	1.43E-03
Jcr4S00117.150	57437	187	14	4.01E-04	1.53E-03
Jcr4S01308.20	57437	2682	107	4.30E-04	1.62E-03
Jcr4S00244.90	57437	460	26	5.12E-04	1.91E-03
Jcr4S13209.20	57437	440	25	5.83E-04	2.15E-03
Jcr4S00012.190	57437	872	42	6.32E-04	2.30E-03
Jcr4S07369.20	57437	646	33	7.51E-04	2.70E-03
Jcr4S01221.80	57437	378	22	7.94E-04	2.82E-03
Jcr4S00335.50	57437	1601	68	8.46E-04	2.93E-03
Jcr4S01955.50	57437	201	14	8.49E-04	2.93E-03
Jcr4S00483.40	57437	1630	69	8.55E-04	2.93E-03

## APÊNDICE E

**Apêndice E.** Análise com genes modelo associados aos Unigenes DR para FTs enriquecidos (21) em conjunto de 1283 alvos pesquisados. *p-value* 0.001. Background indica a quantidade total de genes no banco de dados.

<b>FT ID.</b>	<b>Background</b>	<b>Interações previstas</b>	<b>Interações</b>	<b><i>p_value</i></b>	<b><i>q_value</i></b>
Jcr4S01040.60	57437	12026	475	2.11E-34	6.38E-32
Jcr4S04680.20	57437	11897	458	3.11E-30	4.69E-28
Jcr4S13202.30	57437	9095	376	7.58E-30	7.63E-28
Jcr4S04809.10	57437	14149	500	1.19E-24	9.00E-23
Jcr4S13035.30	57437	10464	358	1.71E-14	1.03E-12
Jcr4S04046.10	57437	9441	327	7.65E-14	3.85E-12
Jcr4S00433.80	57437	7830	280	2.30E-13	9.93E-12
Jcr4S00700.20	57437	9865	336	2.82E-13	1.06E-11
Jcr4S03542.40	57437	7438	267	6.38E-13	2.14E-11
Jcr4S05114.10	57437	10022	338	8.44E-13	2.55E-11
Jcr4S05581.10	57437	10112	336	6.71E-12	1.84E-10
Jcr4S00044.40	57437	4527	174	7.69E-11	1.94E-09
Jcr4S00778.70	57437	3407	138	2.82E-10	6.55E-09
Jcr4S06393.50	57437	849	46	9.70E-08	2.09E-06
Jcr4S01308.20	57437	2682	103	6.26E-07	1.26E-05
Jcr4S01525.20	57437	1132	48	4.58E-05	8.65E-04
Jcr4S01107.10	57437	976	42	8.86E-05	1.49E-03
Jcr4S11783.10	57437	976	42	8.86E-05	1.49E-03
Jcr4S01270.90	57437	2096	76	1.05E-04	1.66E-03
Jcr4S00823.40	57437	658	30	2.64E-04	3.99E-03
Jcr4S02818.20	57437	712	31	4.78E-04	6.87E-03

## APÊNDICE F

**Apêndice AF.** Análise de genes modelo associados aos Unigenes CoREGs DR para FTs enriquecidos (13) em conjunto de 196 alvos pesquisados. *p-value* 0.001. Background indica a quantidade total de genes no banco de dados.

<b>FT ID.</b>	<b>Background</b>	<b>Interações previstas</b>	<b>Interações</b>	<b><i>p_value</i></b>	<b><i>q_value</i></b>
Jcr4S04680.20	57437	11897	80	4.68E-10	1.22E-07
Jcr4S01040.60	57437	12026	77	1.19E-08	1.55E-06
Jcr4S13202.30	57437	9095	60	3.31E-07	2.87E-05
Jcr4S04046.10	57437	9441	61	5.34E-07	3.47E-05
Jcr4S13035.30	57437	10464	65	8.43E-07	4.39E-05
Jcr4S04809.10	57437	14149	80	1.41E-06	6.09E-05
Jcr4S00433.80	57437	7830	48	4.52E-05	1.68E-03
Jcr4S03542.40	57437	7438	45	1.10E-04	3.56E-03
Jcr4S05114.10	57437	10022	56	1.34E-04	3.86E-03
Jcr4S00700.20	57437	9865	55	1.65E-04	4.29E-03
Jcr4S00778.70	57437	3407	25	2.04E-04	4.83E-03
Jcr4S05581.10	57437	10112	54	5.72E-04	1.24E-02
Jcr4S06393.50	57437	849	9	9.99E-04	2.00E-02

## APÊNDICE G

**Apêndice G.** Análise de genes modelo associados aos Unigenes CoREGs não-DEs para FTs enriquecidos (20) em conjunto de 906 alvos pesquisados. *p-value* 0.001. Background indica a quantidade total de genes no banco de dados.

<b>FT ID.</b>	<b>Background</b>	<b>Interações previstas</b>	<b>Interações</b>	<b><i>p_value</i></b>	<b><i>q_value</i></b>
Jcr4S01040.60	57437	12026	305	2.03E-15	6.14E-13
Jcr4S13202.30	57437	9095	245	5.54E-15	8.37E-13
Jcr4S06393.50	57437	849	48	1.36E-13	1.37E-11
Jcr4S04680.20	57437	11897	280	2.81E-10	2.12E-08
Jcr4S00433.80	57437	7830	199	6.66E-10	4.02E-08
Jcr4S05581.10	57437	10112	235	4.08E-08	2.05E-06
Jcr4S03542.40	57437	7438	179	2.68E-07	1.16E-05
Jcr4S00700.20	57437	9865	220	2.74E-06	1.03E-04
Jcr4S04809.10	57437	14149	297	5.42E-06	1.82E-04
Jcr4S02818.20	57437	712	27	3.87E-05	1.17E-03
Jcr4S04651.20	57437	801	28	1.20E-04	3.25E-03
Jcr4S01308.20	57437	2682	70	1.29E-04	3.25E-03
Jcr4S00911.80	57437	235	12	1.97E-04	4.29E-03
Jcr4S03153.40	57437	991	32	1.99E-04	4.29E-03
Jcr4S00030.120	57437	1517	44	2.14E-04	4.31E-03
Jcr4S13035.30	57437	10464	217	3.14E-04	5.92E-03
Jcr4S00700.100	57437	186	10	3.40E-04	6.04E-03
Jcr4S01086.80	57437	816	27	3.67E-04	6.16E-03
Jcr4S03589.10	57437	190	10	4.07E-04	6.47E-03
Jcr4S01852.10	57437	276	12	9.00E-04	1.36E-02