UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Ana Paula Zotta Mota

# A study of molecular responses to abiotic and biotic stresses in *Arachis*

Tese de Doutorado

Porto Alegre
2019

Ana Paula Zotta Mota

**A study of molecular responses to abiotic and biotic stresses in *Arachis***

Tese submetida ao Programa de Pós-graduação em Biologia Celular e Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como parte dos requisitos necessários para a obtenção do grau de Doutora em Biologia Celular e Molecular.

Orientadora: Maria Fátima Grossi-de-Sá

Porto Alegre
2019

**BANCA EXAMINADORA**

Prof. Dr. Antonio Costa de Oliveira
Universidade Federal de Pelotas

Prof. Dr. Giancarlo Pasquali
Programa de Pós-graduação em Biologia Celular e Molecular
Universidade Federal do Rio Grande do Sul

Profª. Drª. Maria Helena Bodanese Zanettini
Programa de Pós-graduação em Genética
Universidade Federal do Rio Grande do Sul

Prof. Dr. Luis Fernando Revers
EMBRAPA Uva e Vinho – Bento Gonçalves

## INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado no laboratório de Bioinformática da unidade EMBRAPA Recursos Genéticos e Biotecnologia e na equipe Interactions Plantes-Nématodes no Institut Sophia Agrobiotech – INRA PACA – França. Para a execução do mesmo, contou-se com bolsa de doutorado do projeto "Abordagens inovadoras no controle do nematoide das galhas, *Meloidogyne* spp." da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Para a execução do doutorado sanduíche, contou-se com a bolsa do projeto "Biotechnological strategies for the control of plant-parasitic root-knot nematodes" – Agropolis – CAPES – EMBRAPA.

## DEDICATÓRIA

Dedico esta tese de doutorado à minha querida tia, Conceição Zotta, que me ensinou em todos os momentos da minha caminhada a ser uma pessoa melhor. Saudades.

## AGRADECIMENTOS

Como diz a filosofia Ubuntu: “Eu sou o que sou pelo que nós somos todos”. Eu não poderia ter feito nada deste doutorado sem a insubstituível ajuda de várias pessoas.

Agradeço imensamente à minha orientadora Drª Maria Fátima Grossi-de-Sá por ter me ensinado tantas liçoes essenciais para o meu futuro e sobretudo por estar sempre presente e disposta a me ajudar.

Je voudrais remercier grandement mon directeur de thèse en France, Dr Etienne Danchin, qui m’a accepté de bras ouverts dans son equipe et a été toujours là pour me soutenir me conseiller et m’apprendre.

Impossível descrever minha eterna gratidão à Drª Patricia Messenberg Guimarães, que durante meu doutorado esteve sempre presente, sempre me aconselhando, me ensinando, me ouvindo. Por todos os ensinamentos, por todas as correções, por tudo, para sempre serei grata.

Agradeço à Drª Ana Brasileiro, por ter estado presente todas as vezes que precisei, sempre tão criteriosa, tão ponderada e por ter sido fundamental no meu crescimento como pessoa, científica e pessoal.

Agradeço ao Dr Roberto Togawa que me acolheu desde o primeiro momento em que decidi ir para a bioinformática, que me faz ser melhor a cada dia, que me aconselha, que me ouve, que nunca me deixou desistir, que foi único, essencial e insubstituível, não só no meu doutorado, mas na minha vida.

Às minhas queridas amigas, Thais, Andressa e Larissa, a amizade de vocês sempre foi essencial e sem vocês não teria realizado nem metade do que realizei no doutorado. Obrigada também por cada comida deliciosa que partilhamos e por toparem sempre minha gordices.

À minha amiga Maria Laine Tinoco (*in memoriam*), minha querida amiga, eu não tenho palavras para descrever o quanto você foi importante na minha vida, você foi embora muito cedo, o mundo jamais será o mesmo sem você.

Je remercie mes amies Danila, Laïla, Georgios, Mathilde, Justine, Geoffrey, Lucie, Florence, Joss, vous avez fait de la France le meilleur endroit au monde. Vous me manquez immensément et j’espère vous retrouver bientôt.

Je voudrais remercier à l’equipe SPIBOC, Corine, Martine et Aurelie. Merci beaucoup pour l’aide, pour toutes nos discussions et pour votre amitié.

À todos do laboratório de bioinformatica: Marcos, Priscila, Natalia e Orzenil por sempre estarem dispostos a ajudar e por todas as discussões enriquecedoras que temos.

Aos amigos dos laboratórios LPPIII e LPPI: Mario Saraiva, BrunaM, Andreza, BrunaP, AnaLu, Firmino, Carol, Bruno e todos os outros. Muito obrigada pela ajuda e amizade, sempre ☺

Ao PPGBCM, principalmente a Silvia, que foi sempre solicita e me ajudou muito nesses anos de doutorado.

À minha querida família brasiliense, meu tio Maneco, Leticia, Antonio e Fábio, vocês são meu porto seguro.

Aos meus pais e minha irmã, obrigada por terem feito tudo por mim, por me apoiarem, por serem a melhor família do mundo.

**ABBREVIATIONS**

ABA – Abscisic acid
ABF – ABRE binding factor
AREB – Abscisic acid responsive element binding
BAC – Bacterial artificial chromosome
bHLH – Basic helix-loop-helix
BUSCO – Benchmarking Universal Single-Copy Orthologs
bZIP – Basic Leucine Zipper Domain
CBF/DREB – C-repeat binding factor/DRE binding protein
CBM63 – Carbohydrate binding module 63
cDNA – Complementary DNA
CEGMA – Cora Eukaryotic Genes Mapping Approach
Cq – Cycle threshold
DAI – Days after inoculation
DEG – Differentially expressed gene
DHN – Dehydrins
DPBB – Double psi beta-barrel
ERF – Ethylene responsive factor
ET – Ethylene
FC – Field capacity
FC – Fold change
FDR – False discovery rate
GAPDH – Glyceraldehyde 3-phosphate dehydrogenase
GO – Gene Ontology
IPCC – Intergovernmental Panel on Climate Change
JA – Jasmonic acid
LEA – Late embryogenesis abundant
MAPK – Mitogen-activated protein kinases
MYB – myeloblastosis oncogene
MYC – Myelocytomatosis oncogene
Myr – Million years
NBS-LRR – Nucleotide binding site leucine-rich repeat
OG – Orthogroup
OHR – Orthologs Hit Ratio
ORA59 – Octadecanoid-responsive AP2/ERF 59
PR – Pathogen-related
PYL – Pyrabactin resistance like
PYR – Pyrabactin resistance
qRT-PCR – quantitative Reverse Transcription-Polymerase Chain Reaction
RKN – Root-knot nematode
ROS – Reactive oxygen species
SA – Salicylic acid
SnRK2 – SNF1-related protein kinases
TF – Transcription Factor
ZF-HD – Zinc finger homeodomain

**RESUMO**

Em seu ambiente natural, as plantas são expostas a uma série de estresses abióticos e bióticos. Entre eles, a seca é um grande obstáculo à agricultura e, em associação com patógenos e pragas, pode causar prejuízos de até US$ 38 bilhões de 2005 a 2015, e comprometer a segurança alimentar. Com as mudanças climáticas e aquecimento global, esse cenário tende a agravar-se, aumentando a ocorrência de secas, a amplitude da gama de hospedeiros e virulência de agentes patogênicos. O estudo da resposta molecular à estresses individuais em diferentes plantas assim como a resposta a estresses combinados, constitui um grande avanço, para a compreensão de possíveis trade-offs entre tolerância e suscetibilidade à estresses abióticos e bióticos, quanto o desenvolvimento de cultivares mais resistentes.

Visando um melhor entendimento da resposta moleculares de plantas, o presente estudo empregou diferentes ferramentas, como genômica comparativa e transcritomas. O uso dessas ferramentas visou a identificação de genes comumente regulados por diferentes estresses. No capítulo I, o transcritoma de quatro diferentes espécies de plantas (*Arachis stenosperma*, *Coffea arabica*, *Glycine max* e *Oryza glaberrima*) inoculadas com nematoides das galhas (*Meloidogyne* spp.) foram analisados e combinados com dados de genômica comparativa gerados a partir do proteoma de 22 espécies de plantas. No total, 17 famílias de proteínas ortólogas são comumente reguladas pela inoculação de nematoides das galhas, representando ao todo 364 proteínas. A anotação funcional desses ortogrupos indicou que a maioria dessas proteínas está associada à parede celular, receptores quinases e a estresse, e tendo uma única família de fatores de transcrição representada, ERF a qual desempenha um papel fundamental na resposta de defesa da planta.

No Capítulo II, a análise do transcritoma de *Arachis stenosperma* submetido à combinação de dois estresses, seca e inoculação por nematoides das galhas, foi realizada. Diversos genes de resistência a estresses combinados foram identificados sendo que uma resposta transcricional distinta foi observada quando da imposição de estresses combinados, quando comparada a estresses individuais. Apenas 14 genes foram encontrados em comum entre a seca, inoculação por nematoides e estresses combinados e 209 genes foram exclusivamente regulados no tratamento combinado. Este resultado demonstra que uma reprogramação molecular única é empregada pela planta sob múltiplos estresses, o que não pode ser previsto pela análise de cada estresse aplicado individualmente.

A análise de genômica comparativa e a análise de estresse combinado realizadas no presente estudo permitiram a identificação de famílias de genes essenciais à resistência / tolerância de plantas. A família das expansinas, a qual foi modulada durante a seca, infecção por RKN e UV, revelando uma importante família de genes em resposta a múltiplos estresses e em espécies distintas; e um gene de desidrina, cuja superexpressão em *Arabidopsis* revelou um importante *trade-off* entre estresses abióticos e bióticos.

**ABSTRACT**

In their natural environment plants are exposed to a range of abiotic and biotic stresses. Among them, drought is a major constraint to agriculture and, in association with pathogen and pests, can cause losses of up to $38 billion from 2005 to 2015 and compromise food security. With the advent of climate change and global warming, this scenario tends to become worse, with an increase in drought occurrence and pathogen`s host range and virulence. Breeding programs are doing great efforts to find solutions to improve multiple stress resistance in plants, however, the trade-offs between abiotic stress tolerance and biotic stress susceptibility have hardly been investigated.

Here, a comparative genomics study was conducted, which has the potential to identify putative evolutionarily conserved genes involved in universal defense mechanisms, to identify genes regulated in common by *Meloidogyne* spp infection in four plant species (*Arachis stenosperma*, *Coffea arabica*, *Glycine max*, and *Oryza glaberrima*). In total, 17 orthologs protein families which respond to the inoculation of RKN in the four species, comprising a total of 364 proteins. The functional annotation of these orthogroups indicated that the majority of these genes are associated with the cell wall, receptor kinases and stress related, with the transcription factor ERF playing a pivotal role in these defense responses.

A number of multiple-stress resistant genes was identified by analyzing the transcriptome of a resistant wild *Arachis* species*, A. stenosperma,* under the combination of RKN and drought imposition. An overall distinct transcriptional response was observed for the combinatory stress imposition when compared to each individual stress, with only 14 genes found in common among drought, nematode infection and combined stresses and 209 exclusively regulated genes within the combined treatment. This result demonstrates that a unique molecular reprogramming is employed by the plant under multiple stresses, which cannot be predicted by the analysis of each individual stress alone.

The comparative genomics and the *cross-stress* analysis conducted in this study enabled the identification of genes families which are essential to plant resistance/tolerance. The gene family of the expansins which was modulated during drought, RKN infection and UV, disclosing an important gene family in response to multiple stresses and across species; and a dehydrin gene,

which overexpression in *Arabidopsis* revealed an important trade-off between abiotic and biotic stresses.

# 1. INTRODUCTION

## 1.1. Losses in agriculture due to biotic and abiotic stresses

Plants are sessile organisms that face a range of biotic and abiotic stresses in their own habitat. To overcome these challenges, they evolved a myriad of physiological and molecular defense mechanisms that enable their survival, but are energy consuming and often result in a non-optimal plant growth and development. Biotic and abiotic stresses are known to cause great agricultural losses, and the identification of the molecular apparatus that plants use to respond to different stresses is key to produce more productive and resistant plants.

Drought is one of the most global and devastating abiotic stress in plants, causing damage to agriculture equivalent to $29 billion between 2005 and 2015 (http://www.fao.org/3/I8656EN/i8656en.pdf). These losses are prone to an increment in the next years, once the $CO_2$ levels in the atmosphere have increased, which causes warmer days and consequently, expose plants to heat, drought and other abiotic stresses (Peters et al., 2012; Zandalinas et al., 2018). In addition, climate changes can increase the population of pests and pathogens, with the rise in temperatures leading to an expansion in the host range of pathogens and also to greater virulence (Garrett et al., 2006).

Despite being a global threat, the most affected world regions are developing countries such as Asia, Africa, Latin America and Caribbean, where the lack of more effective programs and population growth will result in an even worst scenario in the next years (IPCC 2014 report at http://www.ipcc.ch). Consequently, studies related to drought and its effects on plant pest resistance are crucial to produce new cultivars more resistant to stresses under these new environmental conditions, without compromising their productivity.

In the actual world scenario, the demand for more adapted crops is growing. The adaptation of crops to various diseases and adverse natural environments is essential, and can be accomplished searching for resistance genes in the wild relatives of crops. These wild relatives are genetically diverse and were evolutionary selected to tolerate/resist to a range of abiotic and biotic stresses.

One example of the use of relative wild species to search for resistance genes is the peanut (*Arachis hypogaea*). *A. hypogaea* is a tetraploid with a narrow genetic base, being susceptible to multiple diseases. The diploid wild relative of peanut (*Arachis* spp.) have a higher genetic diversity and

were evolutionary selected over the years, making them a source of new alleles for introgression, not only in peanut, but also in other important crops (Guimaraes et al., 2015; Leal-Bertioli et al., 2009).

## 1.2. Abiotic stress

### 1.2.1. Drought stress

Drought stress is one of the major causes of agriculture losses in the world, reducing growth and plant yield (Tripathi et al., 2016). The perception of water scarcity by the plant activates a complex response which involves physiological and molecular mechanisms. To avoid water loss, the first plant physiological response is to close the stomata, which occurs through the signaling of the phytohormone ABA (abscisic acid) (Shinozaki and Yamaguchi-Shinozaki, 2007).

In general, drought is perceived by osmosensors in the cell wall, like histidine kinase, receptor-like kinase and other proteins, that trigger a signal transduction cascade which include mitogen-activated protein kinases (MAPKs) (Bartels and Sunkar, 2005; Chaves et al., 2003). This signal perception, on its turn, initiates several transcription changes in the cell, including the activation of transcription factors involved in ABA biosynthesis and genes involved in direct response to osmotic adjustment (Chaves et al., 2003; Shinozaki and Yamaguchi-Shinozaki, 2007).

The phytohormone ABA binds the nucleocytoplasmic receptors PYR/PYL and inhibits the PP2C proteins, which results in the accumulation of SnRK2 proteins (SNF1-related protein kinases) (Finkelstein, 2013). The accumulation of the protein kinases regulates the ABA-responsive transcription factors: AREB (abscisic acid responsive element binding protein)/ABF (ABRE binding factor) and MYC/MYB. Even though these transcription factors respond to ABA regulation, also known as ABA-dependent via, several genes are activated by an ABA-independent pathway. They comprise CBF/DREB and NAC/ZF-HD transcription factors. Additionally, DREB transcription factor is involved in both ABA-dependent and independent pathways, demonstrating a cross-talk between these two pathways, which normally function independently (Lata and Prasad, 2011). Interestingly, most of the genes regulated by ABA possess the same *cis* element, ABRE (Yamaguchi-Shinozaki and Shinozaki, 2006). The schematic representation of this process is show in the Figure 1.

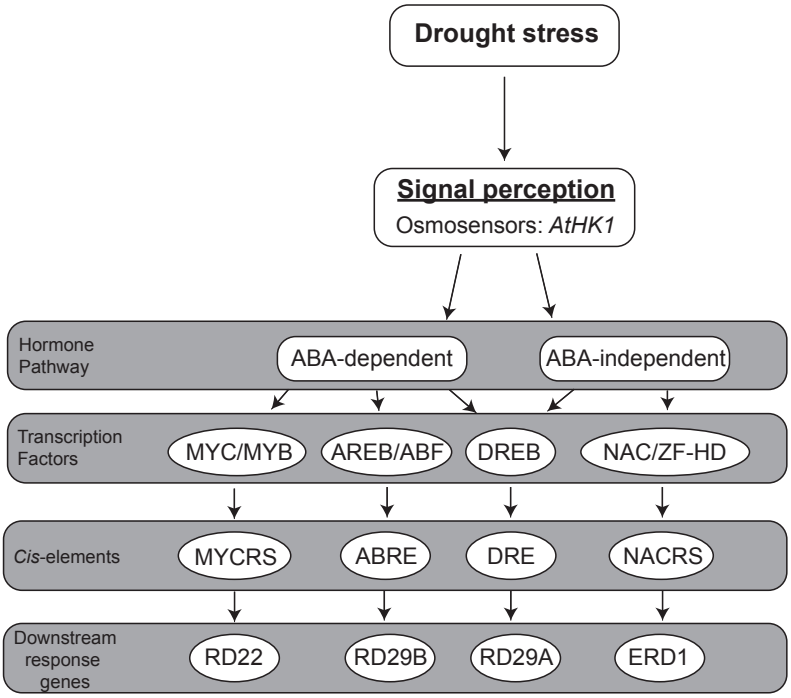



**Figure 1:** Schematic representation of plant response to drought. Adapted from Lata and Prasad, (2011).

In addition, drought induces the increase of reactive oxygen species (ROS), which damage the membranes and macromolecules of the plant (Bartels and Sunkar, 2005). As a defense response, the plant produces different antioxidants [e.g. SOD (superoxide dismutase), APX (ascorbate peroxidase), CAT (catalase)] that protect itself from severe damage(Choudhury et al., 2017), and increase plant tolerance to drought.

### 1.2.2. Dehydrin gene family

Among several genes activated by transcription factors in response to drought, genes belonging to LEA (late embryogenesis abundant) family are the most studied. These proteins act as chaperones, and normally are not expressed in vegetative tissues, except in cases of dehydration. They are thought to protect other proteins, enabling the cells to overcome the drought stress (Bartels and Sunkar, 2005). LEA family comprises several subgroups, with dehydrins (DHNs) belonging to group II. The DHN proteins are one of the most important proteins from LEA family, and their

involvement in different abiotic stresses are well-known (Hara et al., 2005; Kovacs et al., 2008). DHN are defined by a conserved domain, knowns as the K-segment. The K-segment, together with other two complementary segments, S- and Y-segments, classify the DHNs into 5 different subclasses: $Y_nSK_n$; $K_n$; $SK_n$; $K_nS$; and $Y_nK_n$ (Close, 1996). DHNs are often associated with abiotic stresses, and the subclasses to which they belong can interfere on how they respond to specific stresses (Abedini et al., 2017; Graether and Boddington, 2014).

Previous studies associated the $SK_n$-type DHNs mainly to tolerance against cold stress. However, recent studies with tobacco overexpressing $SK_n$-type DHNs demonstrated that this protein can increase the plant tolerance to multiple stresses (cold, drought, salt and osmotic stresses) (Bao et al., 2017; Hill et al., 2016). Additionally, DHNs respond to the wounding and exogenous hormones ABA, JA, SA and ET, which orchestrate plant response to biotic stresses (Hanin et al., 2011; Rosales et al., 2014; Shen et al., 2004). To date, despite its important role in abiotic stress tolerance, few studies have examined DHNs involvement in biotic stresses, with few tests demonstrating its effect against fungal infection (Turco et al., 2004; Yang et al., 2012). The characterization of DHNs gene family, and other gene families which respond to different stresses, is an important step toward the development of stress-tolerant crop varieties.

### 1.2.3. Expansin gene family

Expansins are proteins known to have a cell-wall loosening activity, and are characterized by two conserved domains (DPBB and CBM63). The superfamily of expansins is phylogenetically divided into four subfamilies: α-expansin (EXPA), β-expansins (EXPB), expansin-like A (EXLA) and expansin-like B (EXLB). The subfamilies EXPA and EXPB have been previously described, as acting in the cell wall extension in processes like fruit ripening, pollen tube penetration, root and shoot elongation and abscission (Cosgrove, 2000). In addition, EXPA and EXPB are also known to be involved in the response to several abiotic and biotic stresses (Marowa et al., 2016). A number of studies show that the induction of expansin genes leads to increased tolerance to abiotic stress (Cosgrove, 2015; Sasidharan et al., 2011). Expansins are also involved in the oxidative stress response (Baxter et al., 2014). Despite extensive studies on EXPA and EXPB, the role of the two other expansin subfamilies (EXLA and EXLB) is not yet clear. Our previous studies on *Arachis* spp. root transcriptome subjected to drought and root-knot nematode (Brasileiro et al.,

2015; Guimaraes et al., 2015), revealed a gene from the EXLB subfamily which responds to both stresses. Further studies are therefore necessary on the functional role of EXLB to clarify its role in both abiotic and biotic stresses.

### 1.3. Biotic stress

Pathogens are divided in two major classes: biotrophs and necrotrophs. Biotroph pathogens are those which obtain their nutrients from living cells and necrotrophs those which derive from dead cells (Glazebrook, 2005). Nematodes are biotrophs who use their stylets to inject effectors into the plant cell directly (Dangl et al., 2013). This activates two types of defense responses: PTI (Pathogen/microbial-associated molecular patterns (PAMP/MAMP) triggered immunity) and ETI (Effector-triggered immunity). The PTI is first activated by the perception of pathogen signals, through the PPR (pattern recognition receptors). The ETI is activated by the recognition of some effectors by the resistance genes (R genes), which in most of the cases belong to the NBS-LRR family (Figure 2).

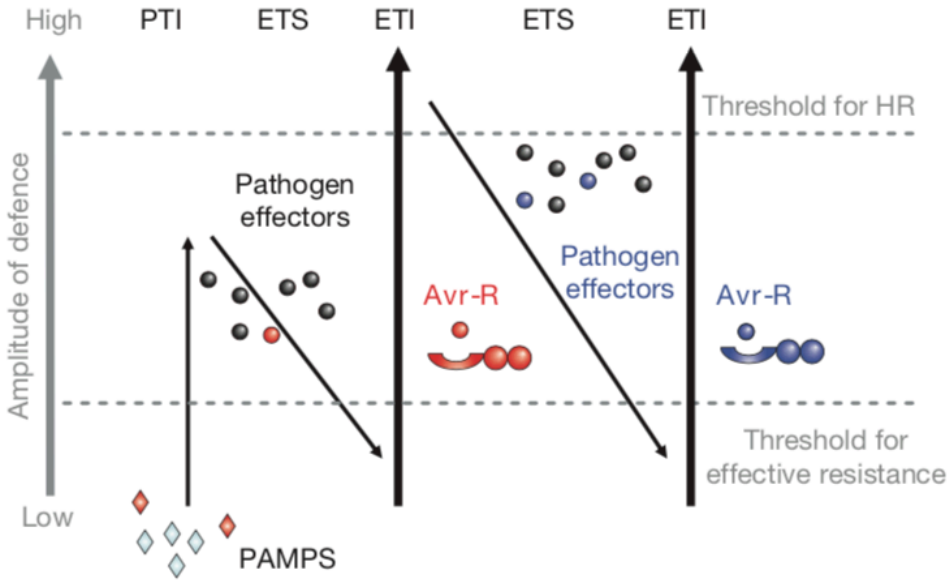



**Figure 2: Zigzag model from Jones and Dangl (2006).** PAMPs or MAMPs activate the PTI response of the plant. The ETS (effector-triggered susceptibility) is a result of the success of the pathogen. The effectors of the pathogen activate the ETI response of the plant, normally through NBS-LRR genes. The effector which were recognized by NBS-LRR genes are them called Avr. The ETI response lead to a Hypersensitive

Response (HR). The pathogen can acquire new effectors, suppressing the ETI response. New NBS-LRR alleles can be acquired by selection, leading again to ETI response.

After pathogen recognition, plants undergo a response mediated by hormone signaling. Three major hormone pathways are known to respond to pathogen infection: Salicylic acid (SA), Jasmonic acid (JA) and Ethylene (ET).

The accumulation of these hormones in the plant lead to the activation of different transcription factors (e.g. AP2/ERF, NAC, MYB, MYC/bHLH, TGA/bZIP and WRKY), which lead to the activation of resistance genes. In the case of SA, the activation of pathogenesis-related (PR) genes is mediated by TGA (Bostock et al., 2014; Pieterse et al., 2012). On the other hand, JA activate different downstream genes, PDF1.2 and VSP2, for example. These genes are activated through MYC2 and ERF1 transcription factors. These two hormone pathways often act antagonistically, once MYC2 can suppress the response of SA and TGA can suppress the activation of JA (Pieterse et al., 2012) (Figure 3).

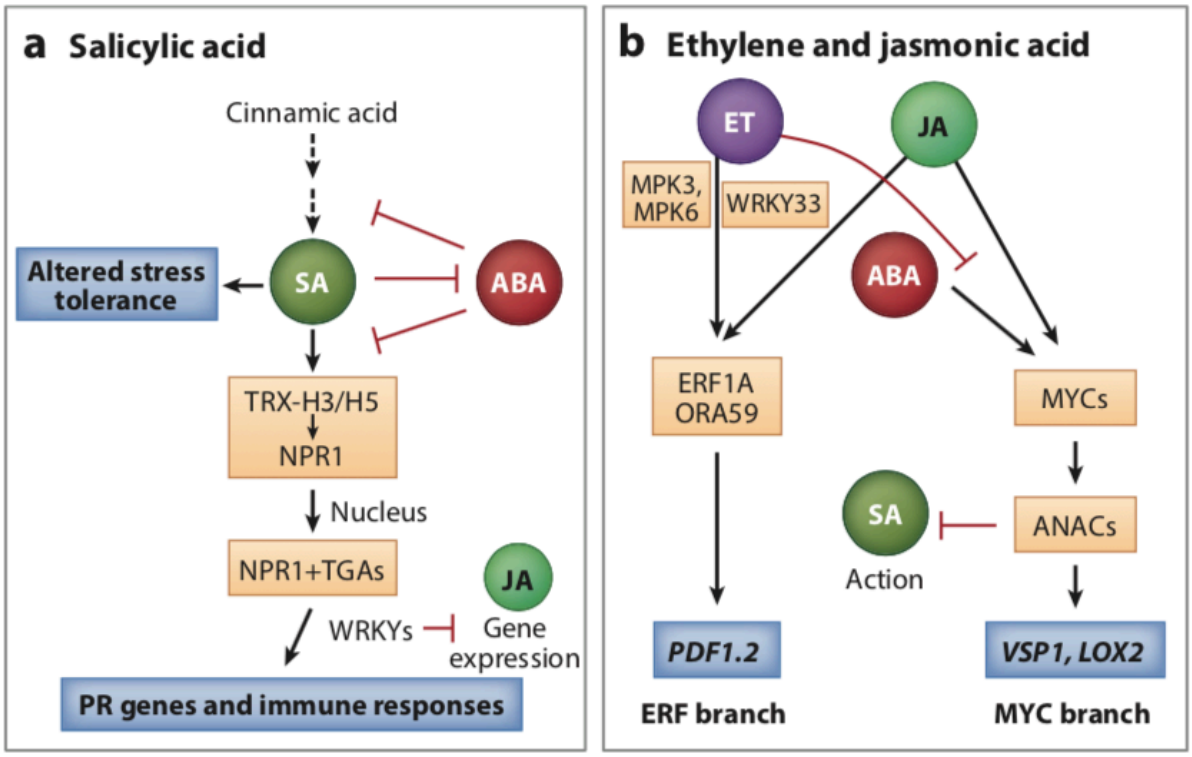



**Figure 3: Phytohormone interactions in abiotic and biotic stress. Adapted from Bostock et al., (2014).** (a) Salicylic acid pathway and Abscisic acid *crosstalk*. (b) Ethylene, Jasmonic acid and Abscisic acid *crosstalk*. The black arrows show the positive interactions and the red arrows the inhibitions.

### 1.3.1. Root-knot nematode infection

In their natural environment, plants are not only exposed to abiotic stresses, but are constantly facing different pathogens, being attacked by bacteria, fungi, viruses and nematodes (Glazebrook, 2005). *Meloidogyne* spp., known as root-knot nematodes (RKN), are among the most harmful pathogens worldwide, causing an estimate of $150 billion/year of losses in agriculture (Abad et al., 2008). They are well adapted to tropical and temperate climates, and can infect almost all vascular plants (Jones et al., 2013).

The four most harmful species from the genus are *M. incognita*, *M. arenaria*, *M. javanica* and *M. hapla,* that together affect most important food crops, followed by *M. graminicola* which is a parasite of rice (*Oryza* spp.). The name root-knot nematode, is a reference to the galls formed in the plants roots, which can be observed after the infection of most species of *Meloidogyne.*

RKN have two distinct life phases, exophytic and endophytic, and during their life cycle evolve in five different stages, from eggs (J1), J2 juveniles, J3, J4 to a final differentiated adult stage (male or female). The first stage of RKN life is inside the egg (J1). The eggs hatch and juveniles (J2) migrate in the soil and invade plant roots, generally through the root elongation zone. Once inside the plant, they migrate along the roots to the protoxylem zone and using their stylets, puncture the plant cell to establish feeding sites. When the feeding site is established, nematodes induce the plants to divide the nucleus of the target cell without cytokinesis and undergo the other three stages, J3, J4 and female. Only the female form remains in the roots, in a sedentary life form, while the males leave the roots (Figure 4).

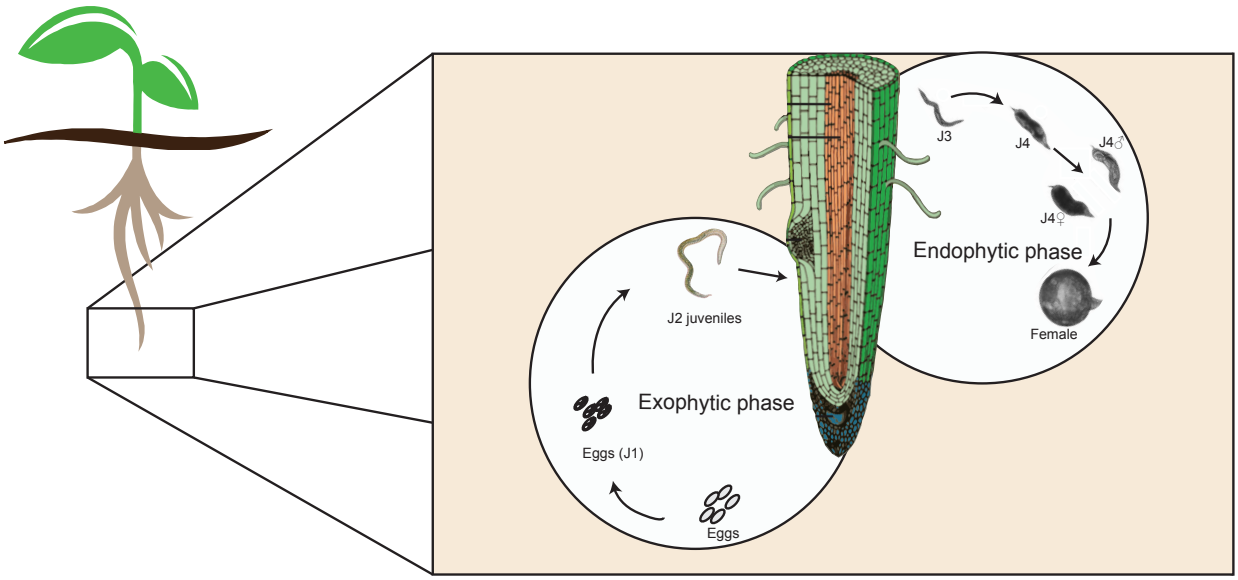



**Figure 4**: Life cycle of a root-knot nematode. The exophytic and endophytic phases are demonstrated in the white circles.

The multinucleated cell is used by the nematode as a feeding site. Each nematode can form more than one feeding site along the root. The J2 moult to the J3 and J4 stages, which do not have a stylet, therefore do not feed, before moulting to the adult stage. The male nematodes leave the roots while the female nematode acquire a pear-shaped and lay eggs in the called galls.

### 1.3.2. Nucleotide-binding site (NBS) gene family

Most R genes activated by plants in response to pathogens belong to the NBS-LRR gene family (Jones and Dangl, 2006). The NBS-LRR gene family is composed by two main domains, the nucleotide-binding site and the leucine rich repeat. In addition to these two domains, NBS-LRR genes can contain a TIR-domain or a Coiled-coil (CC) domain, being therefore classified in different subfamilies TIR-NBS-LRR and CC-NBS-LRR (Meyers et al., 1999). Several studies have used NBS-LRR genes in transgenic plants to increase plant resistance to pathogens, including an overexpression of a NBS-LRR of *Arachis hypogaea* against *Ralstonia solanacearum in Nicotiana benthamiana* (Zhang et al., 2016) *and Arabidopsis thaliana* and *Oryza sativa* plants overexpressing a maize NBS-LRR, to increase resistance to *Pseudomonas syringae*, (Xu et al., 2018). Nonetheless, few studies have used this class of R genes to increase RKN resistance

(Barbary et al., 2016), with only one report on the use of tobacco plants overexpressing a *Prunus sogdiana* NBS-LRR conferring resistance to *Meloidogyne incognita* (Zhu et al., 2017).

In *Arachis*, surveys of NBS-LRR genes were performed using two different approaches, degenerate primers and BAC sequencing (Bertioli et al., 2003; Wang et al., 2012), however, only with the complete sequencing of the *A. duranensis* genome, a more robust genome-wide identification of NBS-LRR gene family was possible (Bertioli et al., 2016; Song et al., 2017). The latter study found 365 NBS-LRRs in *A. duranensis*, however a more detailed characterization of the family and the disclosure of the functional role of each of these genes is still needed.
*A. stenosperma* is a wild species that harbors important resistances to many biotic and abiotic stresses (Guimaraes et al., 2015; Leal-Bertioli et al., 2009). This species is highly resistant to root-knot nematode, displaying a strong HR (Guimaraes et al., 2015; Proite et al., 2008). The study of NBS-LRR gene family in this species has the potential to identify resistant genes linked to RKN resistance and other diseases, which will be of great importance for crop improvement.

### 1.4. Combined abiotic and biotic stresses

In nature, plants are exposed to multiple stresses simultaneously. However, a small number of studies deal with more than one stress at a time, with few reports showing that the plant response to individual stresses is not the same as to the imposition of two or more stresses simultaneously (Atkinson and Urwin, 2012; Pandey et al., 2015; Suzuki et al., 2014). Therefore, the identification of the molecular mechanisms involved in the response to combined stress is important to produce better crops and increase productivity.
Whole transcriptome analysis of plants submitted to individual and combined stresses revealed that genes found in the intersection of individual treatments differ significantly from those found in the combined stresses (Atkinson et al., 2015). Also, the tolerance to individual stress imposition, do not implicate in the tolerance to both stresses applied simultaneously (Ramegowda and Senthil-kumar, 2015). Therefore, considering the predicted impacts of climate change and global warming in agricultural yield (Garrett et al., 2006), the analysis of plant responses to combined stresses is crucial for the development of more adapted and resistant crops.

Plants can react differently when submitted to simultaneous abiotic and biotic stresses with synergic, antagonistic or non-interactive responses (Mittler, 2006; Ramegowda and Senthil-kumar, 2015; Suzuki et al., 2014). Both abiotic and biotic stresses can activate common signals and genes (Chan, 2012), with different responses mostly governed by the action of phytohormones, such as salicylic acid (SA), jasmonic acid (JA), ethylene (ET) and abscisic acid (ABA). Whereas SA is mostly known by its action against biotrophic and JA necrotrophic pathogens, many reports describe a major JA role in different plant-pathogen interactions, including RKN (Lee et al., 2018; Mota et al., 2018). Whereas the convergence of these two hormonal signaling pathways (*crosstalk*) is often reported as antagonistic (Sendon et al., 2011), cooperative exceptions might occur (Brossa et al., 2011).

Several studies describe ABA as the main phytohormone controlling responses to abiotic stresses, and more recently as playing a central role in plant response to both abiotic and biotic stresses (Bostock et al., 2014). The phythormones JA and ET, as well as SA, might interact with ABA in positive and/or negative ways, which are stablished depending on the plant background (Fan et al., 2009; Ollas and Dodd, 2016). Fan et al., (2009), showed that the overexpression of NCED5 gene, one of the precursors of ABA pathway, in *Arabidopsis thaliana* increased its susceptibility to bacteria. On the other hand, Adie et al., (2007) demonstrated that ABA is essential to activated the plant response to necrotrophic pathogens, although this result cannot be extrapolated to all necrotrophic pathogens nor to all the responses activated by JA.

Despite SA pathway being known to respond to biotrophic pathogens, JA is strongly activated in response to the infection of root-knot nematodes (*Meloidogyne* spp.), as previously described in *Ipomea batatas* (Lee et al., 2018). The accumulation of JA might, in turn, activate two different branches, ERF and MYC, which are points of interaction with ET and ABA pathways, respectively (Figure 3).

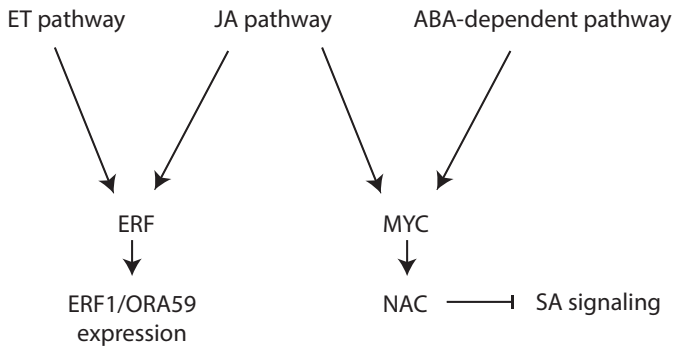



**Figure 3:** Schematic representation of the crosstalk between ET, JA and ABA pathways. Jasmonic acid (JA) activates the ERF and the MYC branches. The ERF branch interacts with the ethylene (ET) pathway activating the ERF transcription factor and subsequently the expression of ERF1 and ORA59. The MYC branch interacts with the abscisic acid (ABA)-dependent pathway, activating the NAC transcription factors, which blocks the signaling through salicylic acid (SA) pathway.

## 1.5. Comparative genomics

The use of comparative genomics to analyze several plants simultaneously enables the identification of putative evolutionarily conserved genes involved in universal defense mechanisms. The functional characterization of these genes as stress responsive can drive biotechnological strategies to enhance crop yield and fitness. Different software packages were developed to generate sets of orthologous proteins, which are defined as genes products that have emerged as a result of an evolutionary speciation event (Koonin, 2005). Basically, these software packages use reciprocal best similarity relationships (e.g. by BLAST) to cluster proteins in groups of orthologs. Nowadays, more than 150 plant genomes, from different orders and families, are available in databases to comparative analysis.

Nonetheless, despite the same origin, sometimes genes display function divergence among plant species (Remm et al., 2001). The combination of comparative analysis and transcriptome data enables the identification of genes which are activated by the same type of stress in different plant species. Also, it makes possible to infer that these genes have kept the same function, despite their evolutionary distance.

Studies using transcriptomes of rice and poplar allowed identification of transcription factors associated with leaf development in both species (Street et al., 2008). Another study using

transcripts of different species of *Sorghum* found multiple stress responsive-genes, to drought, salinity, cold, heat, and oxidative-stress (Woldesemayat et al., 2018).

The application of comparative methodologies to identify gene function among different plant species, including wild *Arachis*, has the potential to identify candidate genes for biotic and abiotic stress resistance tolerance to be used in a wide range of crops.

## 1.6. OVERVIEW

Predicted climate changes require the development of more adapted cultivars which will have the ability to respond to multiple challenges, such as pathogen attacks under water scarcity conditions. New sources of biotic and abiotic resistances are therefore needed, in order to provide new alleles to breeding programs and biotechnology.

The present study aimed to identify genes highly conserved among species responding to RKN infection and also genes responsive to RKN under drought stress. For that, two different approaches were used: a multi-species comparison and the imposition of combined stresses. The first approach aimed to find genes that are evolutionary conserved in response to inoculation of root-knot nematodes in four plant species. For that, different species of root-knot nematode, from the *Meloidogyne* genus were inoculated in four plant species, *Arachis stenosperma*, *Coffea arabica*, *Glycine max* and *Oryza glaberrima* and their root transcriptome analyzed for differential gene expression. To allow this comparison, a database containing orthologs groups of proteins from 22 plants species was produced, which can be further exploited to identify conserved genes for other plant traits.

As in nature, abiotic and biotic stresses occur simultaneously, the study of mechanisms and molecular pathways that govern these individual and combined responses is essential for the development of more adapted cultivars. To study different molecular responses to combined and individual stresses, the root transcriptome of *A. stenosperma* plants subjected to drought, root-knot nematode and the combination of both stresses was analyzed, in order to identify genes responsive to individual and combined stresses.

Finally, as an important outcome of this study two candidate genes, dehydrins and expansins, were identified for the development of stress-tolerant crop varieties, and a detailed genome wide study of these important gene families conducted.

# 2. AIMS

## 2.1. General aims

- To compare the response of different plants species to a single stress, nematode infection or drought
- To compare the response of *A. stenosperma* to combined nematode infection and drought stresses
- To characterize important gene families which respond to biotic and abiotic stress, separately or combined

## 2.2. Specific aims

- To compare plants from different families in order to stablish a phylogenetic and orthology relation among them
- To produce a comprehensive transcriptome of *A. stenosperma*
- To compare the transcriptome data from different plants to the infection of *Meloidogyne* spp.
- To obtain RNA from roots subjected to drought stress, *M. arenaria* infection, and combined stress and analyze the transcriptome data
- To identify genes of *A. stenosperma* regulated by drought stress, *M. arenaria* infection, and combined stress
- To select candidate genes from the previous analyses
- To characterize the expansin gene family in *A. duranensis* and *A. ipaënsis*
- To characterize the dehydrin gene family in legumes species
- To characterize the response of *A. thaliana* overexpressing a AdDHN1 gene to the imposition of drought and root-knot nematode *M. incognita*

# Chapter I

## Comparative genomics: The common response of plants to stress

**Evolutionarily conserved plant genes responsive to the root-knot nematode infection by comparative genomics and transcriptomics**

Ana Paula Zotta Mota, Fabricio BM Arraes, Bruno Paes de Melo, Priscila Grynberg, Anne-Sophie Petitot, Diana Fernandez, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro, Erika VSA Freire, Maria Fatima Grossi-de-Sa, Etienne GJ Danchin

Artigo em preparação para submissão

## 1. ABSTRACT

Root-knot nematodes (RKN) belong to the *Meloidogyne* genus and can parasitize a large number of plants worldwide. They constitute one of the most important causes of damage in agriculture. Advances in comparative genomics allow the identification of evolutionarily conserved plant genes responsive to pathogen infection and constitute a valuable tool for a better understanding of the genetic determinants involved in plant-pathogen interactions. In the present study, the comparison of the genomes from 22 plant species, including agronomically important crops, allowed the identification of core evolutionarily conserved genes and groups of orthologs and in-paralogs across 214 Myr of plant evolution. Overall, 35,238 orthogroups were identified, in which 6,132 were evolutionarily conserved and universal to the 22 species studied here. In parallel, RNA-Seq data previously obtained from four plant species that are resistant to RKN (*Arachis stenosperma*, *Coffea arabica*, *Glycine max* and *Oryza glaberrima*) were used to identify genes significantly regulated during the early stages of RKN infection. Several genes were found as differentially expressed upon RKN infection for *O. glaberrima*, *C. arabica*, *A. stenosperma* and *G. max* (2,597; 743; 665 and 653, respectively). By crossing this information with the previous comparative genomics analysis, 17 orthogroups were identified as conserved among the four RKN-resistant species and containing at least one nematode-responsive gene from each plant species. Despite their phylogenetic distance, these orthogroups had representatives from all the 22 plant species. These results indicated that the genes presented in these 17 orthogroups could be involved in ancestrally conserved pathways triggered following parasite infection. This comparative genomics approach represents a promising predictive tool towards the identification of plant defense-related genes that could be applied to different plant-pathogens or abiotic stresses interactions.

## 2. INTRODUCTION

Plants are sessile organisms that face a range of stresses in their natural habitat that are overcome using sophisticated and intricate perception and responsive mechanisms. Under intensive agricultural production areas, environmental stress conditions can drastically affect plant growth and development leading to a severe decrease in crop yields. Identifying the mechanisms by which plants respond and overcome the different types of stress is an essential step towards development of resistant plants for sustainable crops production under stressful environments (Nobori and Tsuda, 2019).

In the last years, from 2005 to 2015 the effects of climate change have hugely affected the agriculture, causing around $100 billion dollars of losses. These losses include the attack of plants by pests and pathogens, with losses of around $10 billion dollars (FAO report 2017).

Among the most important pathogens are the root-knot nematodes (RKN). Obligate endoparasites from the *Meloidogyne* genus, with approximately 100 species described (Decraemer and Hunt, 2006), that can parasite almost all cultivated plant species in the tropical and temperate regions. Consequently, they have been ranked number one in terms of nematode damages to agriculture. The four most common species from the genus are *Meloidogyne incognita*, *M. arenaria*, *M. javanica* and *M. hapla,* which cause the majority of the yield losses in crops (Agrios, 2005). Another important *Meloidogyne* species is *M. graminicola*, which can parasitize both upland and lowland *Oryza* spp. (De Waele and Elsen, 2007). RKN infect the plant roots by establishing feeding sites and forming multinucleate giant cells. The response of the plants to the RKN infection depends on plant species and genotype (Jones et al., 2013) with few species showing high levels of RKN resistance, mediated by the Hypersensitive Response (HR), triggered in the early stages of the infection.

Transcriptome analyses of plans submitted to abiotic and biotic stresses have been largely used to identify genes responsive to stress, however, this method is often specie-specific. However, with the grow of the number of complete plant genomes, new approaches to identify genes across species were developed. One of these approaches is the comparative genomics, which allow the simultaneous analysis of several plants to identify putative evolutionarily conserved genes involved in plant defense mechanisms (Emms and Kelly, 2015).

Different methods for inference of gene homology across taxa have been developed to identify, using sets of proteins, orthologs and in-paralogs groups. Orthologs are defined as gene products originated from an evolutionary speciation event, and in-paralogs are genes originated from a duplication event in the same species. Despite having one common ancestral, the orthologous genes could display function divergence between plant species (Remm et al., 2001). The subsequent functional characterization of these genes, using the available transcriptome data from multiple species, can address specific orthologous groups as stress-responsive and can drive further biotechnological strategies to enhance crop yield and fitness. These methods of inference are based in the use of the reciprocal best similarity relationships (e.g. by BLAST) to cluster proteins in groups of orthologs and in-paralogs. Nowadays, the great availability of plant genomes (more than 150 plant genomes), from different orders and families, allow a more accurate inference of orthology relationships. In the present study, the genome of nine monocots and 13 dicots plant species were exploited aiming to identify a core of evolutionarily conserved genes. These species were selected based on their genome assembly sequence quality, economic relevance and responsiveness to different types of stress. An outgroup species, *Amborella trichopoda,* which display a good genome assembly and phylogenetic proximity with Angiosperms (Albert et al., 2013) was also included in the analysis.

Comparative genome analysis has been applied in a number of different plant species to assign gene and protein functions by integrating sequence data arising from functional and structural genomics. Previous studies allowed the identification of transcription factors associated with leaf development in *Populus* sp. or genes responsive to multiple stress, as drought, salinity, cold, heat, and oxidative stress in *Sorghum* spp., using transcriptome data (Street et al., 2008; Woldesemayat et al., 2018).

The present study aimed to identify gene families that are commonly responsive to the infection of different *Meloidogyne* species and also evolutionarily conserved in several plants species with distinct degrees of resistance. To accomplish the identification of commonly responsive gene families, transcriptome data from different plant species, in association with homology inference analyses were used. The combination of three different strategies, structural, transcriptional and comparative genomics, will improve the prediction of gene and protein function and help to understand general plant defense molecular mechanisms against root-knot nematodes.

Previous studies conducted by our group have generated transcriptome data from four RKN-resistant plant species (*Arachis stenosperma*, *Coffea arabica*, *Glycine max* and *Oryza glaberrima*) infected with three different RKN species (*M. arenaria*, *M. incognita* and *M. graminicola*) (Beneventi et al., 2013; Guimaraes et al., 2015; Petitot et al., 2017; Valéria et al., 2010). These transcriptome data were further combined with the homology database generated in the present study, aiming to identify clusters of proteins commonly regulated during the response to RKN infection in the four pathosystems studied.

The combination of orthogroups and the transcriptome data from different species, will allow to find conserved proteins responding to RKN infection, in different species. In addition, the generation of homology groups (orthologs and in-paralogs) of proteins will constitute a robust database of evolutionarily conserved plant genes, and a repository of general interest to be shared with the scientific community.

## 3. MATERIAL AND METHODS

### 3.1. Identification of evolutionarily conserved genes in plants

*Selection of plant genomes and transcriptomes in a phylogenetic context*

The predicted proteomes of 21 species were retrieved from their respective whole genome sequence available on public databases (Table 1). These species were chosen based on their phylogenetic distribution, the quality of the whole genome sequence and their economic interest. BUSCO analysis was conducted to assess the predicted proteome completeness using the plants database in the proteome mode (Simão et al., 2015). Because our main goal was to identify widely conserved nematode-responsive genes, we included as much as possible, plant species which have been subject of transcriptomic analysis upon nematode infection. For *Coffea*, *Glycine* and *Oryza*, reference genomes and corresponding predicted proteomes were already publicly available. However, for *A. stenosperma* no reference genome and predicted proteome were available and we first had to assemble a transcriptome and predict the coding sequences (see below), before adding this 22$^{nd}$ species to the comparative analysis. To put the selected species in an evolutionary

framework, we reconstructed a dated phylogeny of the 22 species based on Zanne et al., (2013), using TreeGraph (Stöver and Müller, 2010).

**Table 1:** Number of predicted proteins of each of the 22 plant species here studied, number of proteins present in orthogroups and the BUSCO score of completeness.

| Species* | Number of proteins | Number of proteins in orthogroups | BUSCO score (%) |
|---|---|---|---|
| *Arachis duranensis*[a] | 36,734 | 33,539 | 86.9 |
| *Arachis ipaënsis*[a] | 41,840 | 36,898 | 84.5 |
| *Arachis stenosperma*[b] | 50,693 | 29,218 | 90.1 |
| *Arabidopsis thaliana*[c] | 27,416 | 24,010 | 99.3 |
| *Amborella trichopoda*[c] | 27,313 | 19,533 | 85.2 |
| *Coffea canephora*[c] | 25,574 | 22,672 | 93.4 |
| *Gossypium arboreum*[c] | 40,134 | 35,121 | 94.4 |
| *Glycine max*[c] | 56,044 | 50,262 | 97.4 |
| *Gossypium raimondii* [c] | 37,505 | 34,930 | 97.4 |
| *Musa acuminata*[c] | 45,856 | 41,028 | 96.5 |
| *Medicago sativa*[d] | 59,831 | 46,679 | 89.2 |
| *Medicago truncatula*[c] | 50,894 | 41,103 | 93.6 |
| *Oryza glaberrima*[c] | 33,164 | 31,721 | 92.7 |
| *Oryza sativa*[c] | 42,189 | 36,651 | 95.6 |
| *Oropetium thomaeum*[c] | 28,446 | 22,404 | 70.3 |
| *Phaseolus vulgaris*[c] | 27,197 | 26,293 | 95.9 |
| *Sorghum bicolor*[c] | 34,211 | 28,762 | 98.3 |
| *Setaria italica*[c] | 34,584 | 30,460 | 98.4 |
| *Solanum lycopersicum*[c] | 34,727 | 29,480 | 95.5 |
| *Solanum tuberosum*[c] | 35,119 | 30,552 | 84.1 |
| *Triticum aestivum*[c] | 100,344 | 84,862 | 90.7 |
| *Zea mays*[c] | 63,480 | 37,753 | 92.2 |
| Total | 933,295 | 773,931 | - |

*The letters indicated the database used to retrieve de genome sequence of each species: [a]http://peanutbase.org, [b] transcriptome assembly from the present study; and [c] https://phytozome.jgi.doe.gov v11; [d] transcriptome assembly from (Postnikova et al., 2015).

*Genome-guided transcriptome assembly and proteome prediction of* Arachis stenosperma

A comprehensive transcriptome of *A. stenosperma* was constructed since there was no available genome for this species. A total of 28 transcriptome libraries of the wild *A. stenosperma* obtained by our group from root and leaf samples subjected to drought (Vinson et al., 2018), *M. arenaria* inoculation (Guimaraes et al., 2015), Ultraviolet (UV) exposure and combined stresses (unpublished results) were used to the construction of the transcriptome of *A. stenosperma*. All the libraries were sequenced with the Illumina HiSeq technology, using the TruSeq protocol for the library construction. The raw reads from the 28 libraries were analyzed with FastQC (Andrews, 2010). According to their quality based in the phred 33, the reads were trimmed by Trimmomatic (Bolger et al., 2014) using the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 and MINLEN:36. Trimmomatic was then used to eliminate Illumina adapters. The quality of the cleaned reads was check again with FastQC and only those with good quality, according to the FastQC criteria (base sequence content, GC content, sequence duplication, overrepresented sequences and Kmer content) were used to perform the transcriptome assembly. All the reads were aligned to the reference genome of *A. duranensis,* a diploid wild *Arachis* species closely related to *A. stenosperma* (Bertioli et al., 2016) by the STAR software (Dobin et al., 2013), using the Two Pass parameter. The coordinated file generated by STAR was used as input to Trinity Genome-guided software (Grabherr et al., 2011). The completeness of the contigs generated by Trinity was assessed by two different software: The Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007) and Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015). CEGMA have a different number of core genes (248), conserved in eukaryote genomes than BUSCO (1,440), giving a percentage score according to the number of complete genes found in the input sequence. After the assembly of reads, the redundant contigs were retrieved and eliminated by EviGene software (http://arthropods.eugenes.org/EvidentialGene/).

The completeness of the resulting non-redundant contigs were assessed by CEGMA and BUSCO, as described above. Additionally, to check if the contigs were assembled to their true full length, according to the reference, an Orthologs Hit Ratio (OHR) was performed against the transcripts from the reference genome *of A. duranensis*. The OHR analysis check whether the assembled

contig have a smaller (OHR < 1), equal (OHR = 1.0) or bigger (OHR > 1.0) sequence, compared to the reference genome.

The final set of assembled contigs was used to prediction of protein sequences using the *ab initio* method, using TransDecoder software. The first step of TransDecoder (Grabherr et al., 2011) predicts the most likely longest open read frame from each contig, based on the nucleotide sequences and the second step consists in a Blast search, against the UniProt database (version from January 2017). After the Blast search, sequences with high similarity with non-plant organisms (e.g. bacteria, virus or nematodes) were eliminated. The resulting contig sequences were functionally annotated to their conserved PFAM domains. The final step was the translation of the nucleotide sequences of the generated *A. stenosperma* contigs into amino acid sequences, by the TransDecoder.Predict parameters of TransDecoder software. The final set of deduced amino acid sequences were used in the further analyses.

*Inference of gene homology*

The whole predicted proteomes of all of the 22 plant species were submitted to OrthoFinder software (Emms and Kelly, 2015) for inference of their orthologs and in-paralogs groups (orthogroups or in-paralogs groups). The first step of OrthoFinder analysis consists of a Blastp all-against-all to find the best reciprocal hit for each protein. The blast results were processed by OrthoFinder using the default inflation values. The resulting file was used as input to the Family Companion software (Cottret et al., 2018), which generates the statistical matrices of abundancy of genes per orthogroups, presence and absence of orthogroups per species and the percentage of proteins present in orthogroups. The Kinfin software (Laetsch and Blaxter, 2017) was then used to visualize the rarefaction curve of the species.

*Functional annotation of plant proteomes*

The PfamScan (version 1.5) (Mistry et al., 2007) was used to predict the conserved domains in each of the predicted proteins of the 22 plant species. The resulting file was used as input for the prediction of Gene Ontology terms with Pfam2Go association file (http://geneontology.org/external2go/pfam2go) and a perl script.

### 3.2. Identification of genes responsive to nematode infection

*Bioassays, RNA isolation, and Illumina sequencing*

*A. stenosperma* inoculated with *M. arenaria*: Plants of *A. stenosperma* grown in greenhouse were inoculated with *M. arenaria* infective juveniles (J2) and root samples collected before inoculation (non-inoculated control) and at three time points (3, 6 and 9 days after inoculation; DAI), in three biological replicates according to Morgante et al., (2013). The isolation of the total RNA from samples, the cDNA synthesis and the Illumina transcript sequences were produced by Guimaraes et al., (2015).

*O. glaberrima* inoculated with *M. graminicola*: Plants of *O. glaberrima* grown in greenhouse were inoculated with *M. graminicola* J2 and root samples collected before inoculation (non-inoculated control) and at three time points (2, 4 and 8 DAI) in three biological replicates, RNA extracted, cDNA synthesized and the Illumina transcript sequences produced by Petitot et al., (2017).

*G. max* inoculated with *M. incognita:* A total of 120 four-week-old plants of *G. max* (cultivar 'PI595099') grown in greenhouse were challenged with 1,350 *M. incognita* J2 and the root samples collected at zero (non-inoculated control), 4, 8 and 12 DAI in three biological replicates, as previously described in Beneventi et al., (2013). The total RNA was extracted from whole roots using the kit ReliaPrep™ (Promega©, Wisconsin, USA) according to the manufacturer's instructions. The TruSeq ™ SBS v5 protocol was used for the library construction (Illumina, San Diego, CA) and the 12 resulting libraries were sequenced in Illumina HiSeq 4000 at University of Illinois (USA).

*C. arabica inoculated with M. incognita*: A total of 15 six-week-old plants of *C. arabica* (cv. UFV 408-28) grown in greenhouse were inoculated with 5,000 J2 of *M. incognita* and root samples collected at zero (non-inoculated control) and 6 DAI in three biological replicates, as previously described in Albuquerque et al., (2010). The total RNA was extracted from the root samples using the RNAeasy mini kit (Qiagen, Hilden, Germany) and the cDNA produced using the SuperScript II enzyme and oligo (dT) 20 primer (Invitrogen, Carlsbad, USA). The TruSeq ™ SBS v5 protocol (Illumina, San Diego, CA) was used for the library construction (Illumina, San Diego, CA) and six resulting libraries were sequenced in Illumina HiSeq 2500 at University of Illinois (USA).

*Pre-processing of the reads and statistical analysis*

The Illumina transcript sequences from the libraries of the four plant species were analyzed separately (12 libraries for *A. stenosperma*, *O. glaberrima* and *G. max* each and six libraries for *C. arabica*), using the same pipeline described below. The raw reads of each species were trimmed as described in the section 3.1 of this Chapter. The trimming and the removal of TruSeq Adaptor from Illumina sequencing was conducted by Trimmomatic software using the parameters of LEADING, TRAILING, SLIDINGWINDOW and MINLEN (Bolger et al., 2014).

The cleaned reads of *G. max* were quantified into the transcripts generated from the reference genome (*G. max* Wm82.a2.v1; http://phytozome.jgi.doe.gov/). For *O. glaberrima* and *C. arabica*, the cleaned reads were quantified into the transcripts from the closest available genome (*O. sativa* and *C. canephora*, respectively; http://phytozome.jgi.doe.gov/). The cleaned reads of *A. stenosperma* were mapped into the transcriptome assembly produced as described in the section 3.1 of this Chapter. The mapped reads were quantified into the corresponding predicted transcriptomes, using Kallisto (Bray et al., 2016) with the default settings to obtain transcripts counts and abundances.

The transcript counts obtained by Kallisto were used as inputs in differential expression methods EdgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2013). Genes were considered as Differentially Expressed Genes (DEGs) when their relative gene expression levels had an adjusted p-value (false discovery rate; FDR) $< 0.05$ and at least 4-fold change (FC) value between inoculated and control samples ($Log2FC > 2$ or $< -2$). Only the genes identified as DEGs in both DESeq and EdgeR statistical analyses were considered for further analyses.

*Functional annotation of DEGs using Mercator*

The MapMan ontology (Thimm et al., 2004) was inferred using the online tool Mercator (http://www.plabipd.de/portal/mercator-sequence-annotation). Mercator assigned MapMan BINs (functional categories/sub-categories) to the transcript sequences of each transcript identified as DEGs by the statistical analyses. Mercator incorporates BLAST and CDD searches as well as an optional InterProScan annotation.

# 4. RESULTS

## 4.1. Transcriptome assembly of *A. stenosperma*

The public available genomes of *A. duranensis* and *A. ipaënsis*, the wild parental of the cultivated peanut (*A. hypogaea*) (Bertioli et al., 2016) allowed its exploitation for genomic studies of closely related wild species, without a genome sequence, such as *A. stenosperma,* which have the same A-type genome as *A. duranensis*. Therefore, taking *A. duranensis* as reference genome, a comprehensive transcriptome of *A. stenosperma* was built using the raw data from 28 Illumina-sequenced cDNA libraries previously obtained by our group (Guimaraes et al., 2015; Vinson et al., 2018). These libraries were constructed from roots and leaves of *A. stenosperma* plants subjected to different types of stresses, such as drought, nematode infection, UV exposure and combined stresses. A total of 81.06% of the raw reads from the 28 cDNA libraries were uniquely mapped to the reference genome of *A. duranensis*, with an average length of reads of 334 bp. The *A. duranensis* genome allowed the mapping of *A. stenosperma* transcriptome reads, using the genome-guided strategy, using Trinity software, that generated a total of 224,703 contigs. To check the completeness of the transcriptome assembled by Trinity, the contigs were firstly submitted to BUSCO and CEGMA software packages. From the total of the core genes from the database of BUSCO (1,440 genes) and of CEGMA (248 genes), 94% and 100%, respectively, were found in the assembly generated by Trinity. In a subsequent step, the redundancy of the 224,703 assembled contigs were eliminated by EviGene software, resulting in a reduction of 86.65% with a final number of 30,003 contigs. After eliminating of the redundancy, BUSCO assessment was performed in the 30,003 resulting contigs to verify whether the coverage was reduced. The completeness of the resulting contigs was of 90.1% and 91.53% for the BUSCO and CEGMA databases, respectively. This high reduction after EviGene removal is expected and could be caused by alternative spliced variants generated during the process of genome-guided of Trinity, due to the high number of reads used as input. The integrity of the 30,003 contigs showed a high peak around 1.0 (Suppl. Figure 1), which means that contigs had the true full length, according to the reference. The N50 median of the non-redundant contigs was 2,363 bp. These quality analyses demonstrate that the strategy of genome-guided, using the *A. duranensis* genome as reference was accurate for the assembly of a comprehensive transcriptome of *A. stenosperma*.

The final set of 30,003 contigs were used subsequently to predicted the proteome of *A. stenosperma* by TransDecoder software followed by Blast and a PFAM analyses. The final set of proteins, based in the reconstructed non-redundant contigs, was of 50,693 and used in the further comparative analyses with other proteomes from 21 plant species.

### 4.2. Identification of evolutionarily conserved plant genes

A dated phylogeny tree was reconstructed based in Zanne et al., (2013) using the 22 species chosen in this study. A total of four clades had representatives in this study: Fabidae, Malvidae, Asteridae and Poaceae. To infer a set of homology genes in plants, these 22 species were chosen to represent the species of interest for different biotic and abiotic stresses and also plant-models (e.g. *A. thaliana*), also genome quality parameters as high completeness and good gene annotation were considered. The species were distributed as follow: 13 species from the Eudicotyledoneae and eight species from the Monocotyledoneae, together with an outgroup (*A. trichopoda*), spanning more than 214 Myr of plant evolution. Four genera had two or more representatives, (*Oryza* spp., *Solanum* spp., *Gossypium* spp., *Arachis* spp., and *Medicago* spp.) because of the great economical interest of the genus (Figure 1).

Interestingly, the phylogenomic tree of the species generated by the alignment of the proteins from the orthogroups of OrthoFinder (Suppl. Figure 2), was compared to the dated tree of species constructed in the study (Figure 1), resulting in a very similar relation among the species in both trees. This confirms that the use of multi-copy gene families for the construction of a phylogenomic tree by is a suitable method.

Overall, the predicted proteomes of the 22 plant species represented a total of 933,295 proteins (Table 1). The number of predicted proteins in each species varied from 25,574 (*C. canephora*) to 100,344 (*T. aestivum*) genes and according to BUSCO analysis, they had a completeness score ranging from 70% for *O. thomaeum* to 99% for *A. thaliana.* The high BUSCO score for *A. thaliana* was expected as it is a plant-model, while species less studied or which the transcriptome was used (e.g. *A. stenosperma*), presented a lower BUSCO score (Table 1). The subsequent OrthoFinder comparative analysis allowed assigning the majority of these proteins (773,931; 83%) to 35,238 orthogroups shared by at least two species (non singletons) (Table 1; Figure 1). The percentage of proteins assigned into orthogroups varied according to the species, being the highest rates for *O.*

*glaberrima* and *G. raimondii*, (95% and 93% respectively), while the lowest rates were 57% and 59% for *A. stenosperma* and *Z. mays*, respectively (Table 1; Figure 1).

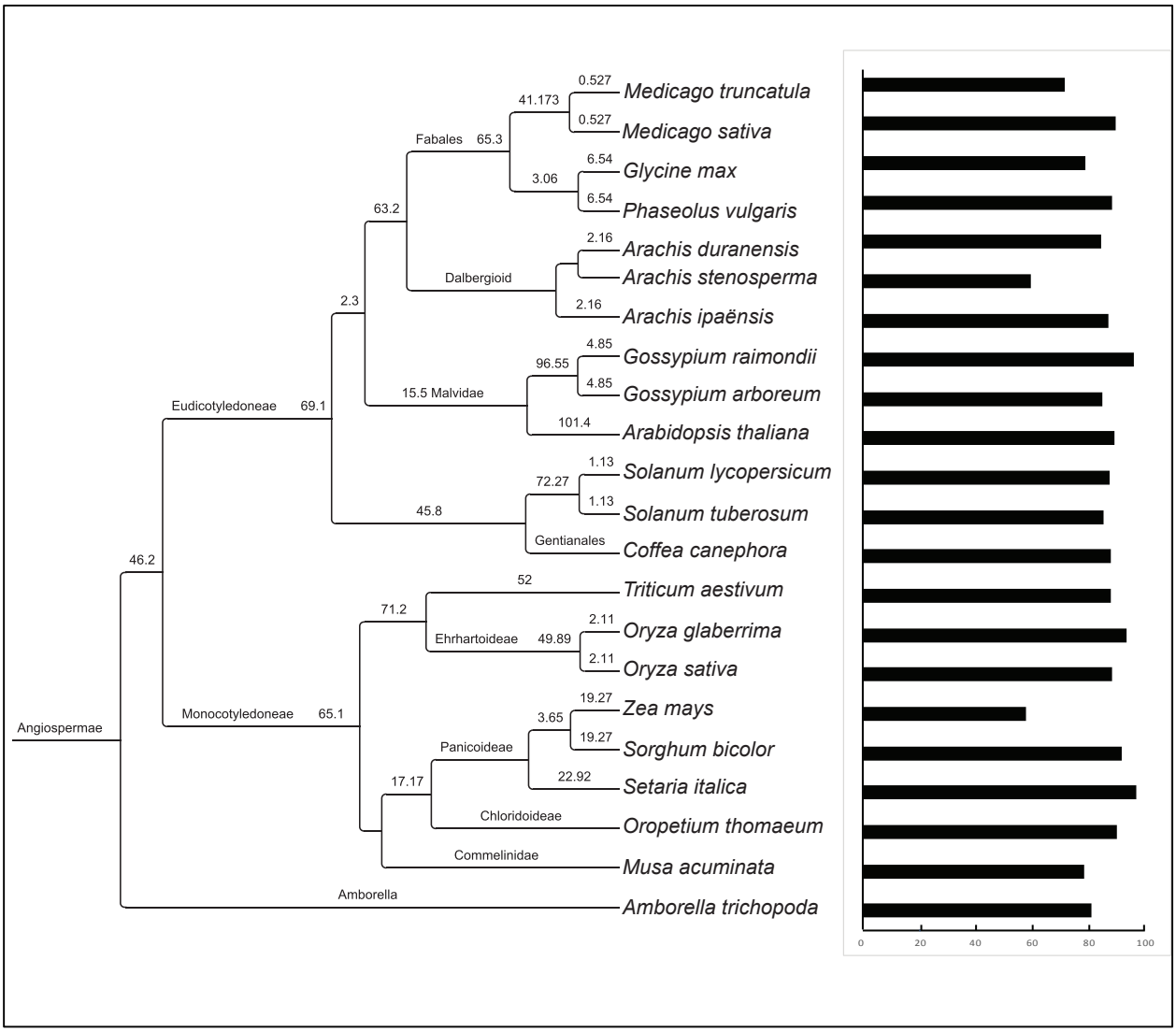



**Figure 1**: Dated phylogenetic tree of the 22 species and percentage of proteins of each species present in orthogroups. The branches of the phylogenetic tree are represented in Mya. Each bar represents the percentage of the total number of proteins for each species which belong to one of the orthogroups.

To verify whether the set of plant species chosen for this study could produce a complete group of homology genes, a rarefaction curve was produced (Figure 2). The rarefaction curve showed that a plateau was reached at around 20 species, indicating that the addition of a 21th plant species to the homology inference does not provide any substantial novelty in the number of groups. It

confirms that the selection of 22 plant species constitutes a good representation of the diversity of Angiosperm proteins.

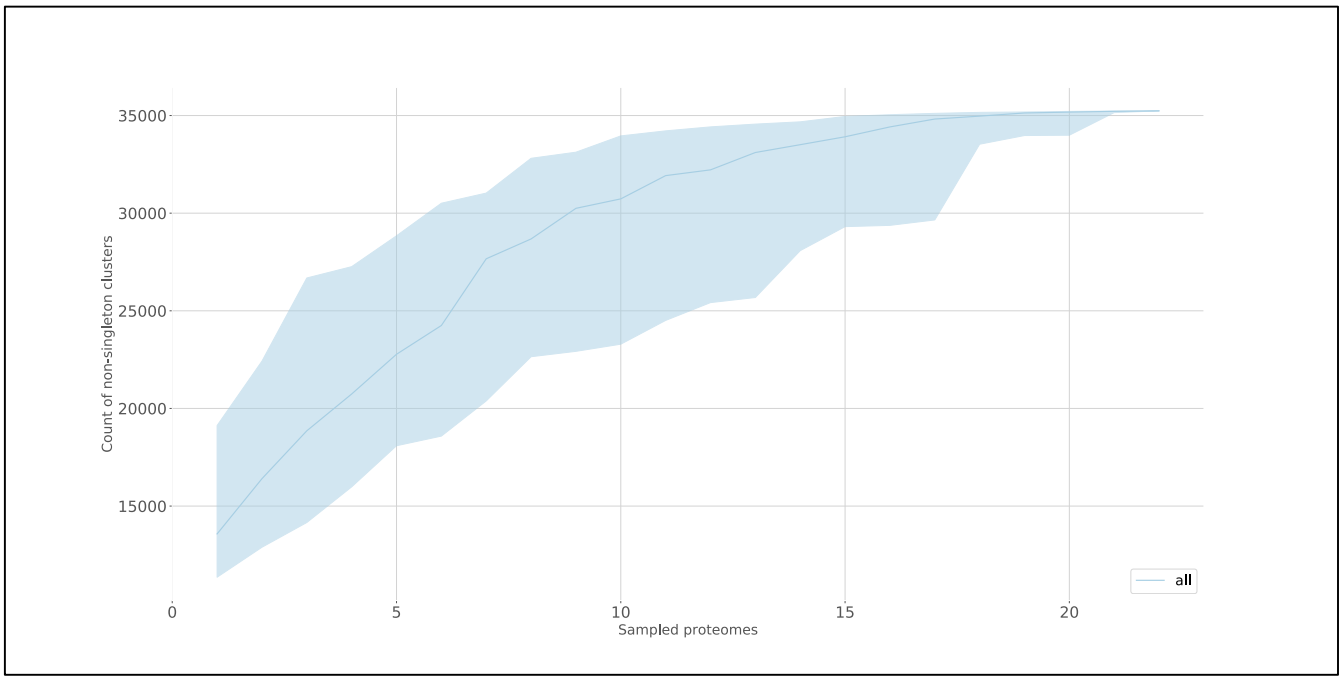



**Figure 2**: Number of clusters per number of predicted proteomes used in the OrthoFinder analysis. The blue line represents the number of orthogroups obtained when using different numbers of proteomes. The axis y show the number of non-singleton clusters and the axis x the number of sampled proteomes.

Aiming to identify the evolutionarily conserved genes among the 22 plant species, all the analysis conducted hereafter considered the 35,238 non-singletons orthogroups. From these orthogroups, only 692 were in-paralogs, in other words, had genes generated from a duplication event in one species, without an orthologous gene in other species. From the total of in-paralogs groups, 12.86% belong to *Triticum aestivum* (89 orthogroups). This is expected, since *T. aestivum* have undergo several events of duplication (Pont et al., 2011). This result indicates that the great majority of the groups (98%) are conserved between at least two species, since all the species have a common ancestral.

Moreover, 6,132 orthogroups were conserved among all the 22 plants species (Suppl. Figure 3). These orthogroups represent a core of conversed genes, and probably are involved on basal biochemical process of plants.

As expected, plant species belonging to the same genus had more orthogroups in common, with *Oryza* spp. sharing 5,275 orthogroups, followed by *Arachis* spp. with 1,543 orthogroups, *Gossypium* spp. 1,279 orthogroups; *Medicago* spp. 1,181 orthogroups; *Solanum* spp. 989 orthogroups (Suppl. Figure 3). Interestingly, the highest number of common orthogroups is shared among all the 22 species (6,132) configuring a robust dataset towards identification of evolutionarily core genes.

The functional annotation using the PFAM conserved domains of the 933,295 proteins predicted in the 22 proteomes identified 4,637 common domains, with 59% shared among all the species (2,753 domains) and 7% were species-specific with *M. truncatula*, *S. lycopersicum* and *M. sativa* with the highest number of exclusive domains with 131, 51 and 38, respectively (Figure 3).

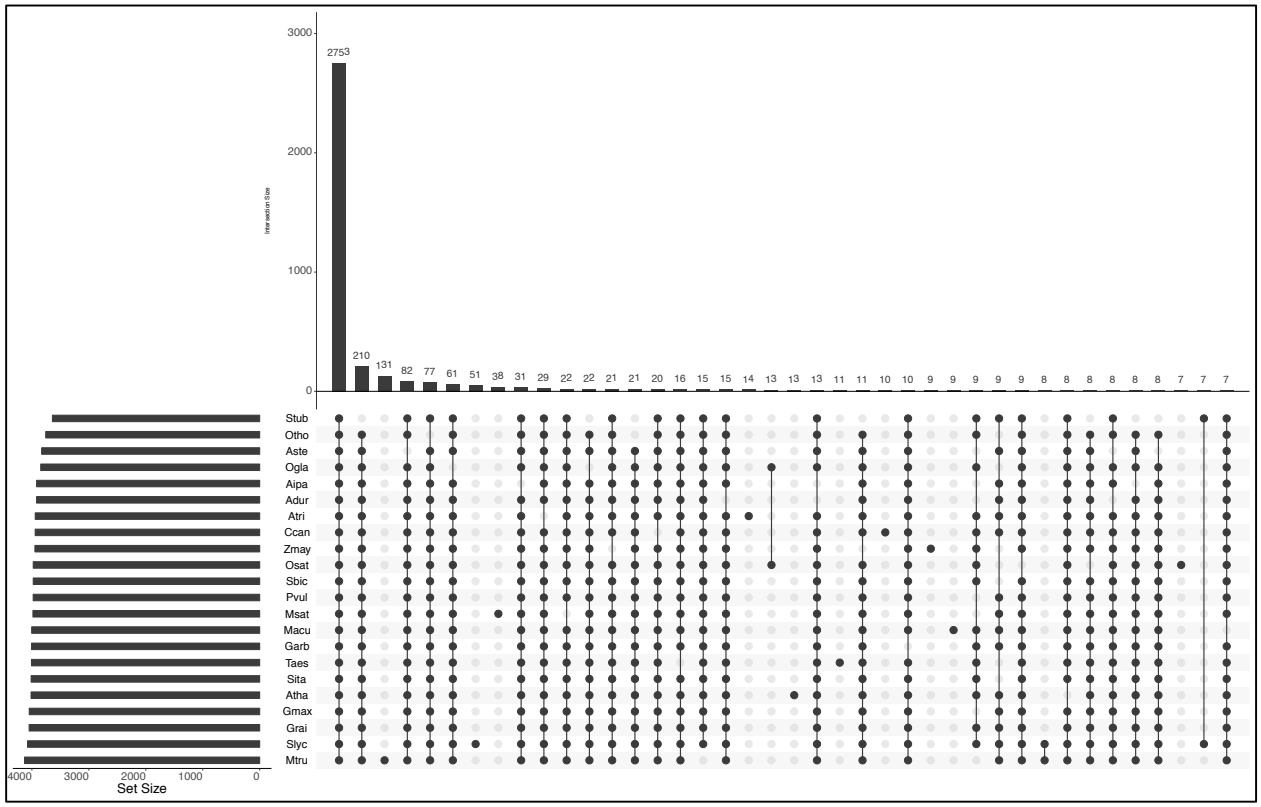



**Figure 3**: Intersection of the conserved PFAM domains of the 22 predicted proteomes. The vertical bars represent the number of orthogroups. The horizontal bars represent the number of orthogroups of each species. The black dots indicate the intersections among the four species.

### 4.3. Transcriptome and Orthology integrative analysis

*Common orthogroups among the four RKN-resistant species*

Aiming to identify common nematode-responsive genes, the subsequent analyses had been directed to the four genotypes (*A. stenosperma*, *C. arabica*, *G. max* and *O. glaberrima*) known to be resistant to RKN infection (Beneventi et al., 2013; Petitot et al., 2017; Proite et al., 2008; Valéria et al., 2010). Once that *C. arabica* does not have a reference genome available, nor sufficient transcriptome data to generate a comprehensive transcriptome database, the closer species from the same genus, *Coffea canephora*, was used as reference in the further analyses. Similarly, the transcriptome data from *O. glaberrima* was quantified using the reference genome of *Oryza sativa*. Therefore, the further results have used the references genomes of *C. canephora* for *C. arabica* and *O. sativa* for *O. glaberrima*.

The root transcriptome surveys of these genotypes were exploited to identify genes differentially regulated under nematode infection compared to non-infected controls.

Based on the previous inferred orthogroups for the 22 plant species (Table 1; Figure 1), a total of 9,323 orthogroups, encompassing a total of 98,502 proteins are common to the four RKN-resistant genotypes (Figure 4). In addition, 8,457 exclusive in-paralogs groups were identified in *O. sativa*, 2,170 in *A. stenosperma*, 1,261 in *G. max* and 610 in *C. canephora* (Figure 4). The difference in the number of in-paralogs groups could be due to duplication events occurred after the differentiation of the species. As expected, the two legume species (*A. stenosperma* and *G. max*) shared a high number of common orthogroups (736), once they had the smallest evolutionary distance (Figure 1).

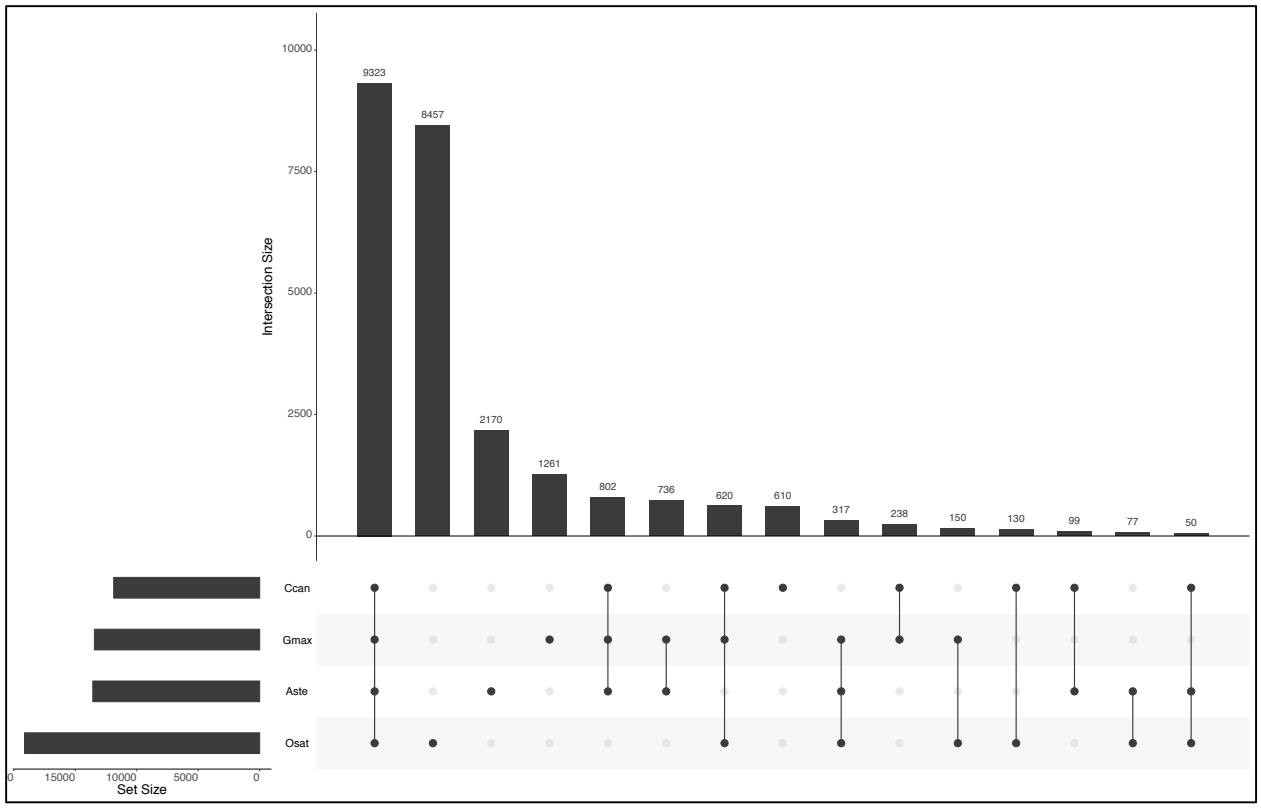



**Figure 4**: Intersection of the orthogroups of *A. stenosperma* (Aste), *C. canephora* (Ccan), *G. max* (Gmax) and *O. sativa* (Osat). The vertical bars represent the number of orthogroups. The horizontal bars represent the number of orthogroups per species. The black dots indicate the intersections among the four species.

*Gene expression profile of root transcriptome*

RNA-Seq data available from the four species was analyzed using the same pipeline and only the genes identified as DEGs (FDR < 0.05 and $Log_2FC > 2$ or < -2) in both statistical softwares (EdgeR and DESeq) were considered for further analyses. This strategy maximizes the precision of the differential expression prediction, once that both statistical programs use different algorithms and approaches, producing the most reliable set of DEGs. The results showed a total of 4,658 genes differentially expressed (DEGs) in response to RKN inoculation, being 2,597 in *O. glaberrima*; 743 in *C. arabica*; 665 in *A. stenosperma* and 653 in *G. max* (Figure 5). The four-times higher number of DEGs observed in *O. glaberrima* may be due to *M. graminicola* infection mechanism (Petitot et al., 2017). Most of these DEGs were regulated positively in *A. stenosperma*, *G. max* and

*O. glaberrima* (486, 484 and 2,318 DEGs, respectively), while in *C. arabica*, most of the DEGs (598) were negatively regulated (Figure 5).

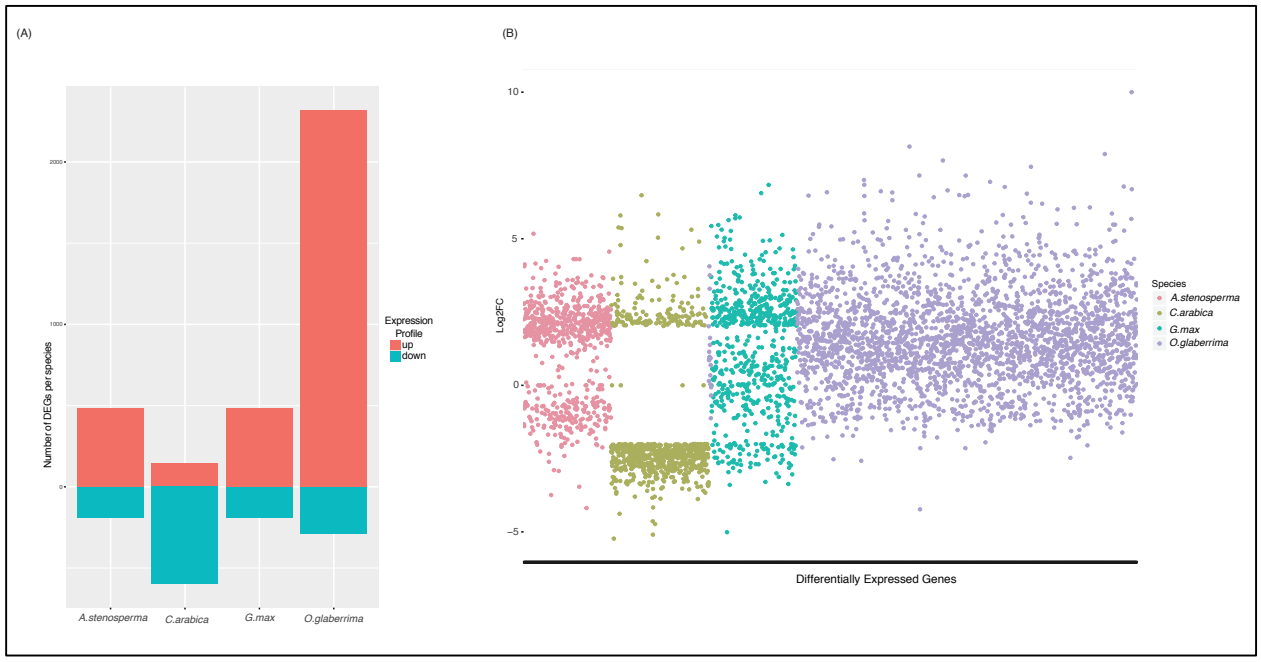



**Figure 5**: Differentially expressed genes (DEGs) in response to RKN inoculation in *A. stenosperma*, *C. arabica*, *G. max* and *O. glaberrima*. (A) Barplot represent the total number of DEG for the four species. The red bars indicate the upregulated genes and the blue bars the downregulated. (B) Expression value of the DEGs of the four species in all the time points. Each species is represented by a color and the y-axis is in Log2FC values obtained from DESeq.

*Integration of Orthology and Transcriptome data*

Aiming to conduct a comprehensive analysis of common RKN-resistance mechanisms, the data obtained in the homology inference of genes was integrated with these obtained in the transcriptome analysis of the four RKN-resistant genotypes. The association of the 9,323 orthogroups common to the four genotypes with the 4,658 DEGs from identified as RKN-responsive revealed that 17 orthogroups contained at least one representative DEG from each plant genotype. These 17 orthogroups encompass a total of 18,422 proteins, from the 22 species, were 3,661 proteins belong to the four target species and 364 proteins were identified as DEGs (Figure 6). *O. sativa* is the species with the highest number of in-paralogs groups (1,369; Figure 6),

followed by *A. stenosperma* (404), *C. canephora* (350) and *G. max* (226). The high number of in-paralogs groups of *O. sativa* is expected, once it had the highest number of DEGs from the four species. In addition, *O. sativa* is the species which shared the highest number of orthogroups in bilateral relations (only two species) with 95 orthogroups shared with *C. canephora*, 81 with *A. stenosperma* and 80 with *G. max* (Figure 6).

The *A. stenosperma* high number of in-paralogs groups could be due to the use of a comprehensive transcriptome in place of a genome.

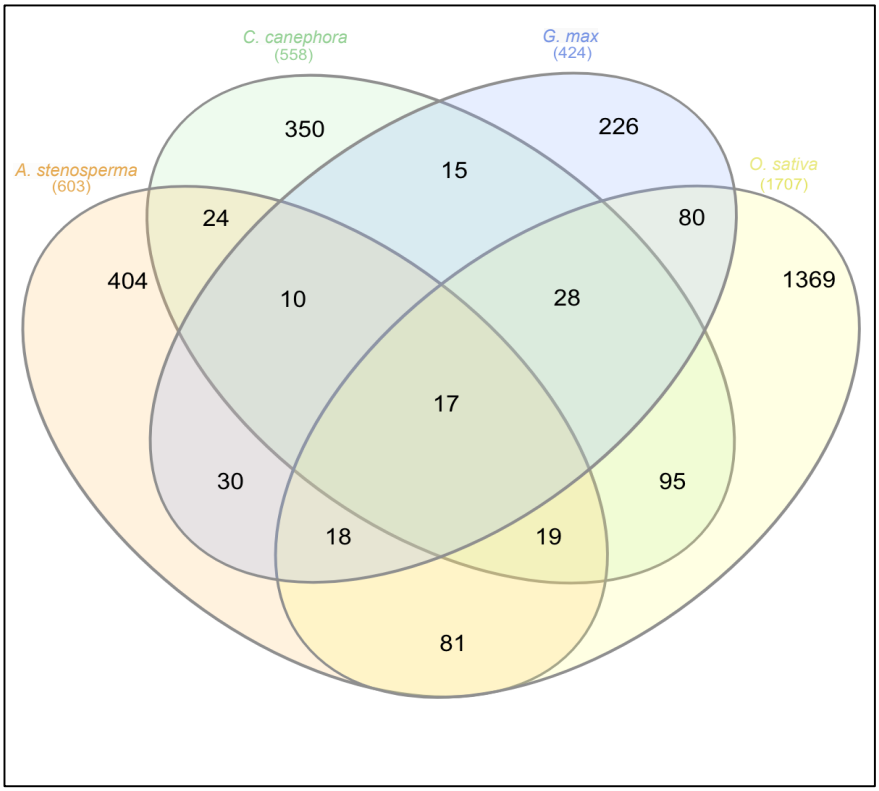



**Figure 6**: Venn Diagram to show the association of the orthogroups and in-paralogs groups containing DEG of the four species. The number displayed in the Venn Diagram correspond to the number of orthogroups identified. The superimposed regions of all circles show the number of orthogroups shared among the species.

The analysis of the expression profiles of the 364 RKN-responsive DEGs belonging to the 17 common orthogroups revealed that most of them (80%) were upregulated in *A. stenosperma*, *G. max* and *O. sativa.* Conversely, in *C. canephora* 73% of the DEGs were downregulated in response

to RKN infection. The expression values varied from -5 (*G. max*) fold to 7.5 (*O. sativa*), but most of the genes had their expression around 2 to 3-fold (Figure 7).

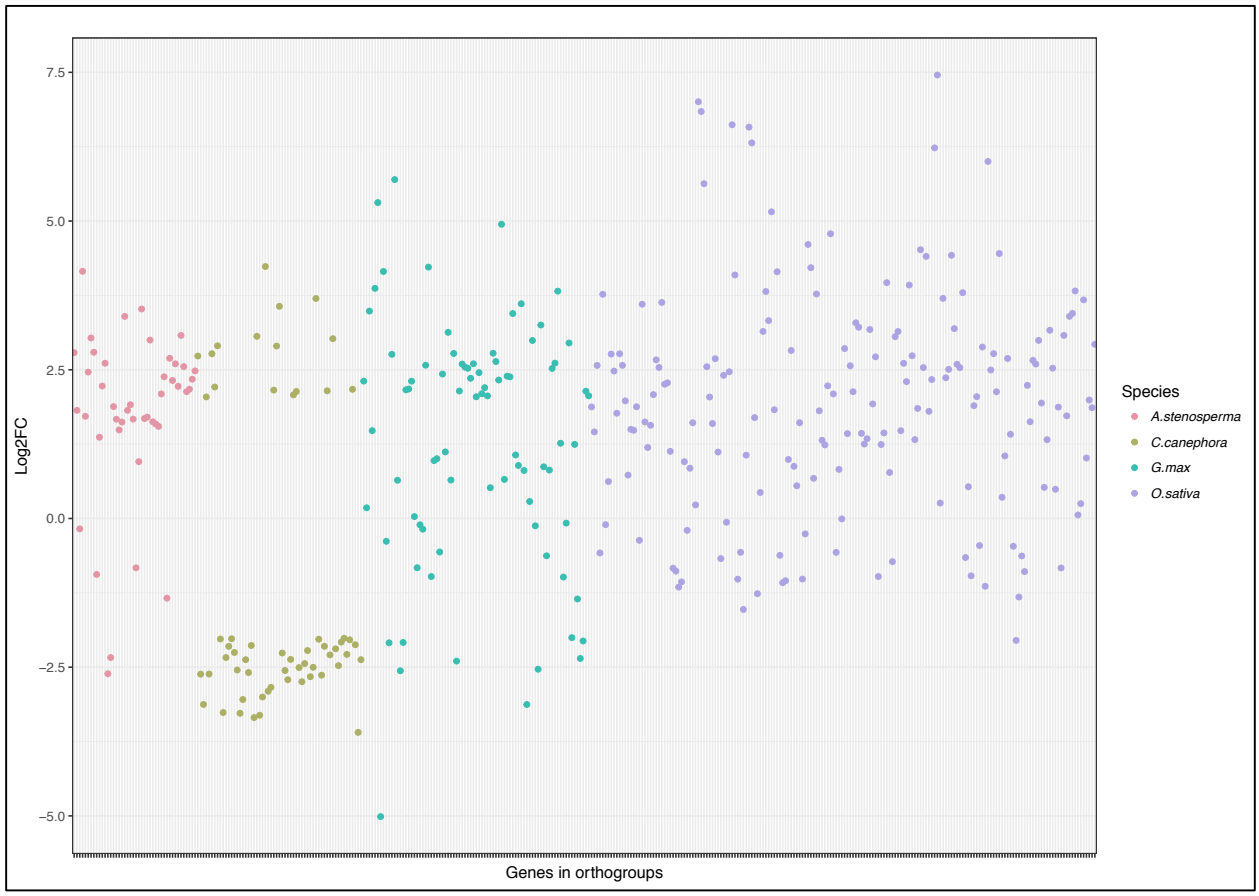



**Figure 7:** The expression profile of the 364 DEGs from the four species (*A.stenosperma*, *C. canephora*, *G. max* and *O. sativa*) belonging to the 17 common orthogroups. The different colors indicate each of the four species. Each dot represents a different gene. The expression values are in $Log_2FC$ obtained from DESeq.

The distribution of the proteins of the four species in orthogroups was uneven. From the total of DEGs of each species in orthogroups, *O. sativa* showed the lowest percentage (6%) of the 2,597 DEGs belonging to the 17 common orthogroups. The highest percentage (12%) was observed in *G. max*, followed by *C. canephora* (7.9%) and *A. stenosperma* (6.6%) of DEGs belonging to one of the 17 orthogroups (Table 2).

**Table 2:** The 17 orthogroups commonly regulated in response to root-knot nematode. The first column represents the orthogroups code generated by OrthoFinder analysis. The number of differentially expressed genes for each species present in groups, and the PFAM domains annotated for each group.

| Protein Family | Total number of proteins | *A. stenosperma* | | *C. canephora* | | *G. max* | | *O. sativa* | | PFAM domains |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Total | DEGs | Total | DEGs | Total | DEGs | Total | DEGs | |
| OG0000001 | 3661 | 207 | 5 | 103 | 2 | 251 | 10 | 145 | 34 | Kinase (PF07714) |
| OG0000003 | 3554 | 114 | 1 | 340 | 7 | 192 | 2 | 172 | 49 | LRR (PF08263; PF13855; PF00069) |
| OG0000004 | 3508 | 152 | 7 | 198 | 10 | 193 | 24 | 137 | 26 | Cytochrome p450 (PF00067) |
| OG0000014 | 1285 | 50 | 1 | 52 | 7 | 73 | 3 | 58 | 11 | PTR2 (PF00854) |
| OG0000019 | 948 | 53 | 7 | 38 | 7 | 85 | 2 | 23 | 3 | Pectinesterase (PF04043; PF01095) |
| OG0000033 | 656 | 28 | 1 | 37 | 2 | 34 | 3 | 28 | 3 | Methyltransferase (PF08100; PF00891) |
| OG0000035 | 649 | 23 | 2 | 13 | 6 | 52 | 3 | 33 | 9 | Expansin (PF03330; PF01357) |
| OG0000036 | 647 | 27 | 8 | 13 | 1 | 54 | 3 | 26 | 3 | AP2 (PF00847) |
| OG0000050 | 531 | 20 | 1 | 19 | 1 | 43 | 1 | 21 | 5 | Heat shock factor (PF00447) |
| OG0000053 | 522 | 6 | 1 | 12 | 1 | 40 | 2 | 29 | 11 | Hydrophobic seed (PF14547) |
| OG0000056 | 514 | 9 | 1 | 31 | 1 | 27 | 2 | 19 | 6 | UDPGT (PF00201) |
| OG0000057 | 507 | 5 | 1 | 17 | 2 | 28 | 1 | 19 | 9 | Cupin (PF00190) |
| OG0000058 | 502 | 20 | 3 | 34 | 4 | 47 | 1 | 12 | 4 | FAD (PF01565; PF08031) |
| OG0000073 | 415 | 12 | 1 | 15 | 2 | 34 | 1 | 9 | 3 | GST (PF02798; PF00043) |
| OG0000179 | 263 | 10 | 2 | 20 | 4 | 19 | 2 | 6 | 2 | Dirigent (PF03018) |
| OG0000309 | 197 | 11 | 1 | 3 | 1 | 37 | 20 | 6 | 1 | Fasciclin (PF02469) |
| OG0001906 | 71 | 3 | 1 | 3 | 1 | 9 | 1 | 2 | 1 | DXP (PF13292; PF02779; PF02780) |
| Total | 18,430 | 750 | 44 | 948 | 59 | 1,218 | 81 | 745 | 180 | |

*Functional analyses of the common RKN-responsive DEGs*

The 364 DEGs identified as common RKN-responsive in the 17 orthogroups were distributed into 12 functional categories (Figure 8), with most of them (32%) assigned to signaling categories, followed by cytochrome P450 (18%) and stress and secondary metabolism (11%). These was expected once the two orthogroups with the highest number of genes of this study were assigned to the kinase PFAM and cytochrome P450 PFAM. The signaling category encompass proteins involved in first response of plant to stress and pathogen perception, as kinases LRR genes, and glycoprotein-like. Proteins assigned to stress, cytochrome P450 and secondary metabolism categories, could also be involved in common resistance response to pathogens. In the Transcription factor (TF) category, only genes coding for ERF TF family were identified as commonly induced among the four plant species, showing a general upregulation. Genes associated with metabolism of the phytohormone cytokinin were also identified as commonly regulated in all the species.

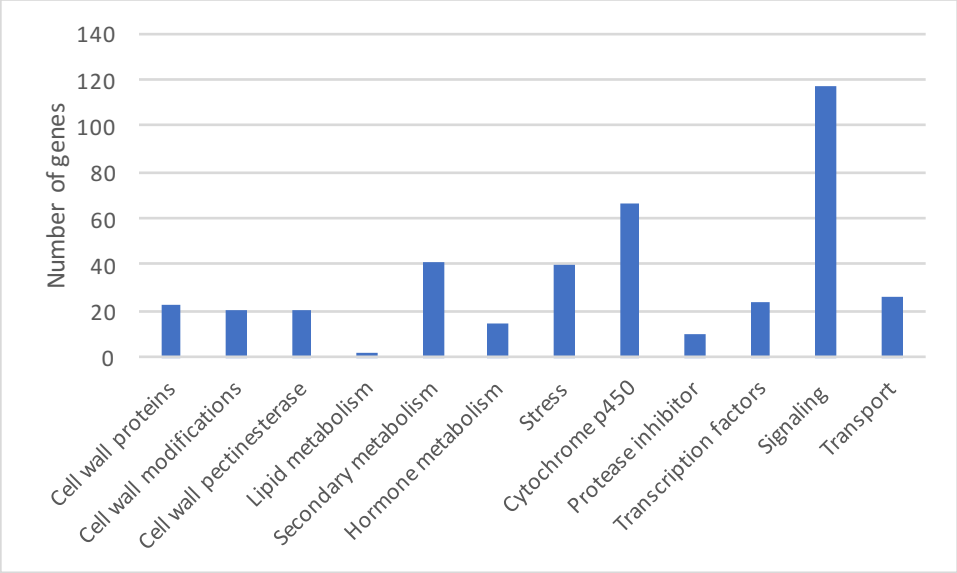



**Figure 8:** Graphical representation of functional categories. The functional annotation of all the DEGs from the four species. The categories are based on the Mercator annotation.

The 17 orthologous identified in the present study were functionally annotated for their conserved PFAM domains (Table 2). The number of DEGs from each species presented in the orthogroups were uneven, with *G. max* and *O. sativa* with the greatest number of genes (81 and 180,

respectively), compared to *A. stenosperma* and *C. canephora* (44 and 59, respectively). The first three orthogroups (OG0000001; OG0000003 and OG0000004) encompassed the majority (48.2%) of the 364 DEGs (Table 2). These orthogroups are associated with several PFAM domains, therefore, define a precise annotation for these orthogroups. Most of the domain architecture found in the groups did not had a gene function associated or previous studies describing their function. Except by the genes from the OG0000035, which are expansin from the sub-group of α-expansins, which were described before to be associated to several stresses response (Marowa et al., 2016) and the orthogroup OG0000036, that is the transcription factor from the ERF family, AP2, which is associated to abiotic stress response (Xie et al., 2019). Interestingly, *G. max* presented an augmented number of genes in the orthogroup OG0000309 (Fasciclin), which is a gene associated with abortion of fertilized ovaries in *Arabidopsis* (Cagnola et al., 2018).

## 5. DISCUSSION

The use of comparative genomics to identify core evolutionarily conserved genes and groups of orthologs and in-paralogs, based on protein sequences from distant plant species, proved to be a robust method. This method allowed the identification of more than 35,000 non-singleton orthogroups, clustered with high confidence, of which 6,132 were universal to all the plant genomes compared. Other studies have used this same strategy to determine orthogroups, however, this was the first time that a high number of plant species spanning a distant evolutionary diversity was analyzed (Woldesemayat et al., 2018; Yoshida et al., 2017). The choice of the species was based on their genome quality, economical interest and their unique response to different biotic and abiotic stresses. This dataset of core evolutionarily conserved genes in a large panel of plant of interest will constitute a valuable resource for several different studies in the future.

### *5.1. Evolutionary distant plant species have genes with similar functions*

Despite spanning more than 210 Million years of evolution, this analysis allowed the identification of 6,132 orthogroups encompassing 3,939,809 genes universally conserved between monocots and dicots. These evolutionary conserved genes in the chosen Angiosperma probably correspond to a core of plant processes and functions. However, previous studies have demonstrated that, only the sequence similarity alone is not enough to transfer the functionality to a gene (Das et al., 2016).

The present study suggests that complementing orthology information with expression data from the species is instrumental in the assignment of a function.

Besides universally-conserved genes, the in-paralogs groups observed in all the species can be due to differential loss, retention or expansion of ancestral gene copies as adaptive responses to different challenged such as defense response to environmental conditions that each species face (Das et al., 2016).

Here, together with the inference of orthologous groups, the use of transcriptional data from different species was applied. Previous studies using the same approach have identified commonly regulated genes in species of *Sorghum* genus, and constructed a functional genomics database (Tian et al., 2016) and more recently, comparing the *Sorghum* species with related model species to find commonly regulated genes to drought, salt, cold, heat and oxidative stresses (Woldesemayat et al., 2018).

Here, to determine besides the sequence similarity, the conservation of functional protein domains, was used the PFAM annotation, based in the curated database of InterPro, which is a reliable source for functional annotation. The high conservation of PFAM domains in the 22 species confirms that the plants have inherited most of genes from a common ancestor, and most likely have conserved its function.

### *5.2. Number of differentially expressed genes for each species show the specialized responses*

The differential gene expression analyses of the four plants showed a diverse number of DEGs for each species. These differences can be associated to the evolutionary distance among the studied species, once they span more than 234 Myr of divergence in extreme cases. For the closer species, *A. stenosperma* and *G. max*, the number of differentially expressed genes was more similar. Also, the use of different RKN species can activate specialized responses, and different plants can show differences in response to a same stress as well (Jones et al., 2013).

The combination of the comparative genomics with the differential expression data from all the four species was able to find 17 orthogroups commonly responding to nematode infection. These results showed that, despite the evolutionary distance among the species, and the difference in the root-knot nematode species, plants have kept a set of commonly regulated genes to respond to

stress imposition. This was also previously predicted by computational approaches, comparing public databases from multiple stresses (Breitling et al., 2005; Shaik and Ramakrishna, 2014).

Most of the differentially expressed genes commonly regulated for the four species where related with signaling and cell wall indicating that the first response, which seems to be common to all the plant species, occurs in the first barrier of the plant, and did not present a species-specific response nor a specificity to the nematode species. They can also be associated to other stresses impositions, such as drought, and could be used as multiple-stress and cross-species genes for biotechnology crop improvement.

In the present study, only one transcription factor family was identified, the ERF transcription family, which was commonly regulated for all the species. The ERF transcription factor can be associated to two different hormone pathways, the ethylene pathway and the jasmonic acid pathway. Both pathways are known to respond to biotic stresses, however, only few studies associate this transcription factor to nematode infection (Lee et al., 2018).

Interestingly, no pathogen-related (PR) genes were commonly identified in this analyses, showing that, despite the great number of orthologous groups found in common, genes may develop different functions (Snel et al., 2002).

A group annotated to the fasciclin gene family was found in all the four species, with a high number of genes from *G. max* being regulated upon nematode infection (20). The expression profile of this groups showed a contrasting expression between *G. max* and the other three species, with *G. max* showing an upregulated profile. This gene family is often associated with ovaries abortion in *Arabidopsis*, however, a study with *Populus* had associated this gene with salt response (Zang et al., 2015).

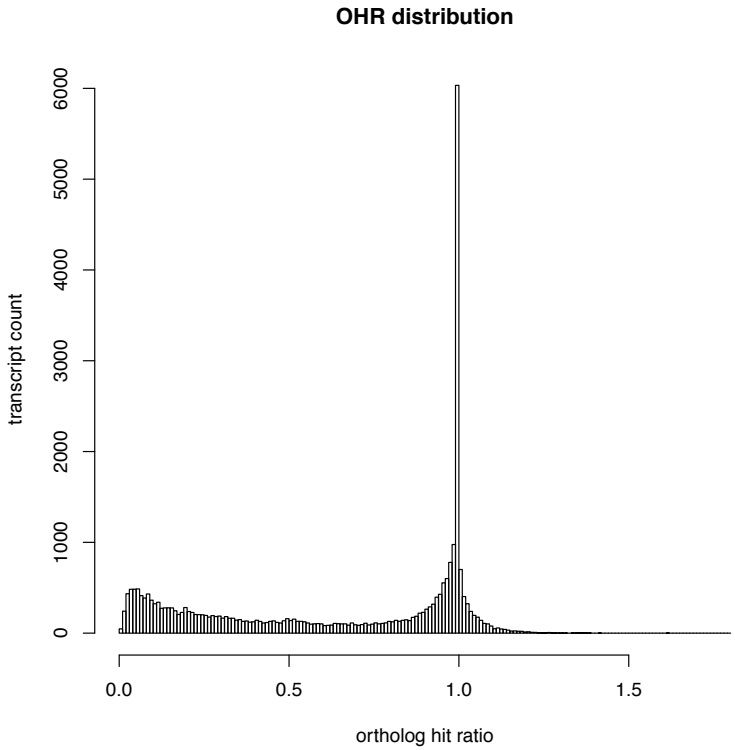
Another gene family associated to stress response found in this study was the expansin gene family. The expansin gene family is a cell wall loosening protein and known to respond to different stresses (Marowa et al., 2016). Recently, a gene from the expansin-like b subfamily, was found as regulated by root-knot nematode in *A. stenosperma* (Guimaraes et al., 2015).

*Cross-species nematode-responsive genes are candidates for crop improvement*

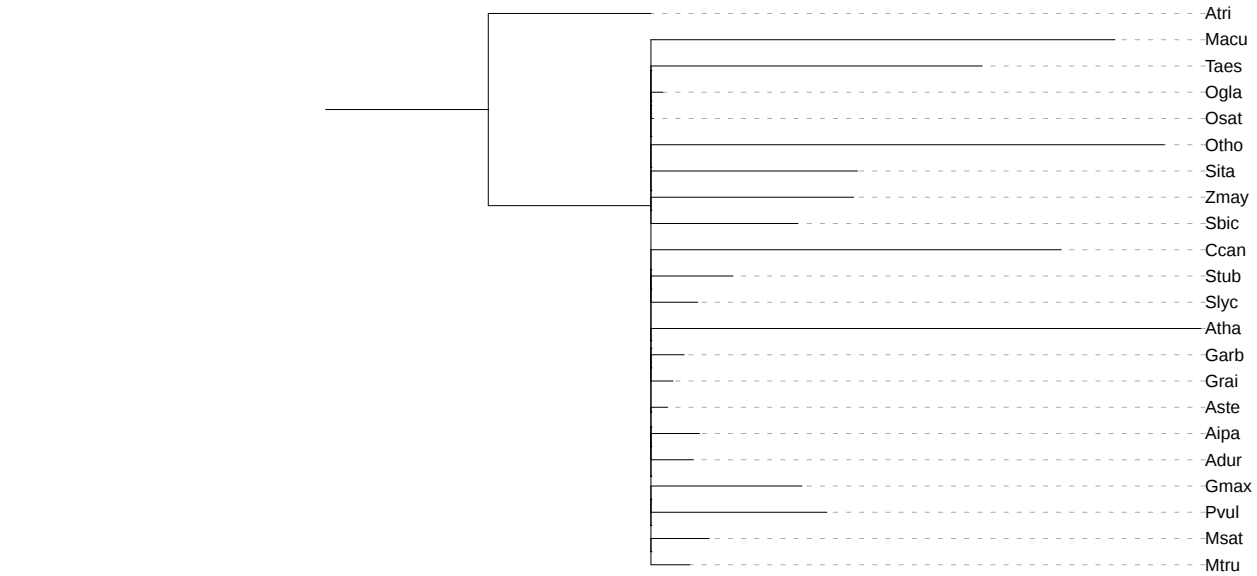
The present study has showed that the comparative genomics approach, comparing distant plant species could find commonly regulated genes responding to root-knot nematodes, independently

of their species. Find genes from several species infected by RKN, which respond commonly is an important step through the selection of candidate genes for crop improvement and evolutionary determination of gene function, once we could determine that orthologous genes were responding to the same stimulus. Besides being an important tool for biotechnological improvement of plants, the development of a database with orthologous groups from 22 plants species will allow future studies to use the same strategy and determine commonly regulated genes for several other stresses conditions.

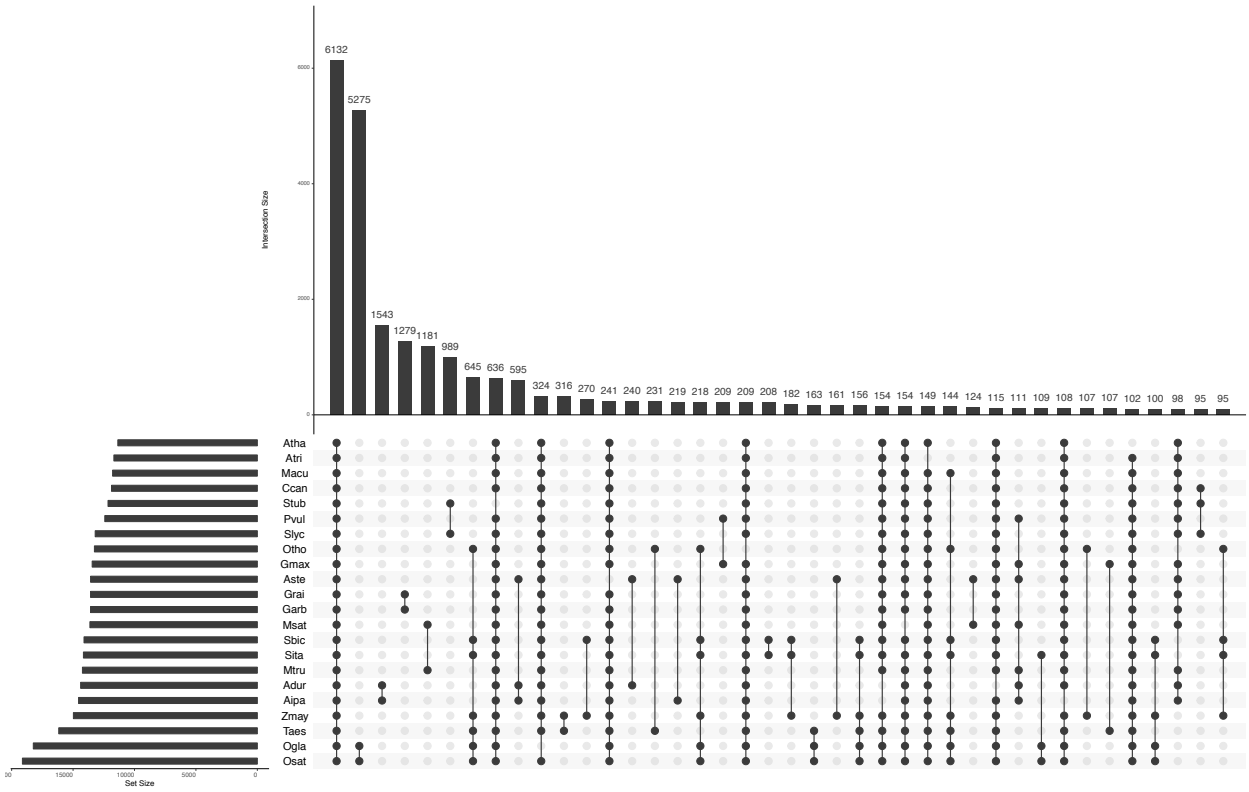
## 6. SUPPLEMENTARY MATERIAL




**Supplementary Figure 1:** Ortholog hit ratio of the non-redundant contigs of *A. stenosperma*. The OHR scores the contigs for length smaller (OHR < 1), equal (OHR = 1.0) or bigger (OHR > 1.0) than the reference. The reference genome is *A. duranensis*.




**Supplementary Figure 2:** Phylogenomic tree of the species generated by the alignment of the proteins from the 35,238 orthogroups of OrthoFinder.




**Supplementary Figure 3:** Intersection of the orthogroups of the 22 species. The vertical bars represent the number of orthogroups. The horizontal bars represent the number of orthogroups per species. The black dots indicate the intersections among the four species.

# Chapter II

Cross-stress: The response of *Arachis stenosperma* to combined abiotic and biotic stresses

**Crosstalk between biotic and abiotic stress in wild *Arachis*: the whole is not equal to the sum of its parts**

Ana Paula Zotta Mota, Bruna Vidigal, Ana Claudia Guerra de Araujo, Etienne Danchin, Maria Fatima Grossi de Sá, Ana Cristina Miranda Brasileiro, Patricia Messenberg Guimaraes

Artigo em preparação para publicação

## 1. ABSTRACT

Drought, heat, salinity and pathogens are some of the stresses to which plants are continuously, and often simultaneously, exposed in nature. Most studies have focused in plant responses to single stresses, however, it was recently revealed that the responses to combinatory biotic and abiotic stresses are different to those prompted by each stress separately. Root-knot nematodes and drought cause major losses in agriculture, but plant defense mechanisms to both impairments simultaneously have not yet been studied. In this study, the transcriptional changes and the stress-regulatory crosstalk in a wild *Arachis* species, resistant to root-knot nematode and drought, was analysed upon biotic, abiotic and both stresses combined. Illumina Hi-Seq reads from *A. stenosperma* roots submitted to drought produced the highest number of exclusive differentially expressed genes (DEGs), followed by combined stresses and nematode infection treatments. Genes identified in common among the three treatments included those associated to the Ethylene (ET) and Salicylic Acid (SA) hormonal pathways, while genes exclusive to the combined stress treatment were mostly associated to abscisic acid (ABA). Functional annotation indicated that the responses to biotic and abiotic stresses were mostly specialized, with no GO terms enriched in common, with and only one transcription factor ERF (ethylene responsive factor) identified in common among all treatments. The arrangement and expression trends of these DEGs into metabolic pathways highlights the prevalence ABA as the major phytohormone regulating plant stress responses and interacting with jasmonic acid, salicylic acid and ethylene to promote a concerted defense response against both, drought and RKN infection.

## 2. INTRODUCTION

Plants are sessile organisms which are constantly exposed to different environmental stresses. These include hazards as drought, heat, salinity and biotic stresses caused by fungi, bacteria, virus and nematodes. Over the course of evolution, plants have developed different networks to overcome these stresses (Mickelbart et al., 2015; Roux et al., 2014). To date, a great deal of understanding of plant responses to individual stresses under experimental conditions has been amassed (Pandey et al., 2015). However, in the field plants are exposed to a combination of abiotic and biotic stresses, and their responses towards combined stresses is still largely unknown. Recent studies revealed that plant responses to combinatory stresses are different from those due to the imposition of individual stresses, and activate different signaling pathways (Ramegowda and Senthil-kumar, 2015; Suzuki et al., 2014). In addition, in the case of simultaneous occurrence of biotic and abiotic stresses, plants can prioritize the response to one stress over another, causing a decrease in the tolerance to other types of stress (Ramegowda and Senthil-kumar, 2015).

The roles of phytohormones such as abscisic acid (ABA), salicylic acid (SA) and jasmonic acid (JA) in activating the defense response against abiotic and biotic stresses is well known (Pieterse et al., 2012; Shinozaki and Yamaguchi-Shinozaki, 2007). However, few studies address the various regulatory mechanisms and signaling events mediated by these hormones in response to simultaneous stresses or how their pathways overlap and interact with one another (Atkinson et al., 2013; Berens et al., 2018; Kissoudis et al., 2016; Rizhsky et al., 2004; Visser et al., 2018).

In many stress responses, abscisic acid (ABA) is the main regulator, playing an important role in drought and plant-microbe interactions with systemic effects on resistance and susceptibility (Bostock et al., 2014). Although in most cases ABA interacts negatively with SA and JA acids causing a decrease in plant tolerance to pathogens (Berens et al., 2018; Ku et al., 2018), there are also several examples in which ABA enhances plant disease resistance (Cao et al., 2011; Ton et al., 2009).

For most legume crops, including peanut (*Arachis hypogaea*), the greatest hazard is drought, which tends to increase over the years due to the impact of climate changes (Zandalinas et al., 2018). In addition to water scarcity, the cultivated species is affected by several pathogens, including the root-knot nematode (RKN) *Meloidogyne arenaria,* which causes significant losses in tropical areas

(Jones et al., 2013). Previous studies investigated the effects of drought and nematode infection in wild *Arachis* species, and highlighted the different pathways involved in response to each of these stresses (Brasileiro et al., 2015; Guimaraes et al., 2015; Mota et al., 2018; Vinson et al., 2018). Overall, the hypersensitive response (HR) displayed by *A. stenosperma* against *M. arenaria* could be linked to the induction of the jasmonic acid pathway (JA), whilst an ABA-independent response was identified leading to adaptive physiological changes in the drought tolerant *A. duranensis* (Guimaraes et al., 2017; Vinson et al., 2018).

In this study, we analysed the global gene differential expression of an RKN and drought resistant species *A. stenosperma*, under combined and individual stresses, in order to assess the integrated, overlapping and also nonlinear plant responses to these challenges. The better understanding of the stress-regulatory signaling and crosstalk that plants apply to separate and common challenges is crucial to optimize plant performance in the face of diverse challenges.

## 3. MATERIAL AND METHODS

### 3.1.Plant material

A total of 80 plants of *A. stenosperma* (accession V10309) were cultivated in soil (2:1 soil:sand mixture) in green-house. After six-weeks, three sets of plants were subjected to different stresses, whilst control plants were kept in normal conditions.

***Nematode bioassay****:* Twenty plants of *A. stenosperma* were inoculated in the roots with 10,000 juveniles (J2) of *M. arenaria*. The plants were kept well-watered for seven days after the inoculation. Roots were collected at the $7^{th}$ day after inoculation (DAI), and divided into three biological replicates.

***Drought assay****:* Twenty plants were subjected to a gradual water deficit (dry-down) during seven days, or until 25% of the field capacity (FC) was reached. Roots of the stressed plants were collected for RNA extraction, and divided into three biological replicates.

***Combined stresses assay**:* The third set of 20 plants was subjected to the stresses above simultaneously. Six-week-old plants at 70% of field capacity were subjected to water withdrawal at day one followed by a inoculation of 10,000 juveniles (J2) of *M. arenaria.* Plants were kept with a gradual water deficit for seven days in green-house. At the $7^{th}$ day, roots were collected for RNA extraction and divided in three biological replicates.

### 3.2.Transcriptome sequencing

*Arachis stenosperma* roots RNA was extracted for each treatment using a modified lithium chloride protocol (Morgante et al., 2011) and purified with Invisorb Plant RNA Mini Kit (Invitek, Berlin, Germany). The cDNA was produced using the Super Script II enzyme and oligo (dT) 20 primer (Invitrogen). Three independent biological replicates were produced, generating 12 libraries which were sequenced in HiSeq4000 at Macrogen Korea. The library construction was conducted using the TruSeq ™ Stranded Total RNA LT Sample Prep Kit. The libraries were paired-end sequenced with an average read of 100 bp.

### 3.3.Gene expression analysis

The quality of the raw reads was analysed by FastQC (Andrews, 2010). According to the quality, each library was individually trimmed by Trimmomatic (Bolger et al., 2014) using the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Only reads with a quality superior to Q>30 were kept for further analyses. The adapters from TruSeq were also retrieved and eliminated when present. The cleaned reads were quantified, into the *A. duranensis* reference genome, by Kallisto (Bray et al., 2016). The number of read counts were normalized and used as input to EdgeR (Robinson et al., 2010). The genes with a value of FDR < 0.05 and a log2FC > 2 or < -2 were considered as differentially expressed genes (DEG).

### 3.4.Functional analysis

The functional annotation and classification of DEGs was carried out using two different methods. The enrichment of Gene Ontology (GO) terms was also conducted by the hypergeometrical test, from func package (Prüfer et al., 2007). The enrichment was made by the comparison between the

mapped genes and the differentially expressed genes from each treatment, with only the GO terms with a FDR < 0.05 considered for further analyses. The DEGs were searched for Transcription Factor genes using the Transcription factor database (http://planttfdb.cbi.pku.edu.cn).

### 3.5.Relative expression by qRT-PCR

The relative expression of selected genes found in the previous analysis was conducted by qRT-PCR. Total RNA from the four treatments: control, drought, combined stress and nematode infection, was extracted in biological triplicates and used for cDNA synthesis.

The qRT-PCR analysis was conducted using three technical replicates for each sample, using the Platinum® SYBR® Green qPCR Super Mix- UDG w/ROX kit (Invitrogen), in the StepOne Plus Real-time PCR system (Applied-Biosystem). The primers design and the reaction settings were conducted as described by Morgante et al., (2013). Average cycle threshold (Cq) values were estimated using the online real-time PCR Miner tool (Zhao et al., 2005) and normalized with two reference genes (GAPDH and 60S), as previously established by Morgante et al., (2011). The REST software (Pfaffl et al., 2002) was used to determine the expression ratios of the transcripts from the treated libraries against the control.

## 4. RESULTS

*A. stenosperma* roots submitted to individual and combined biotic (RKN infection) and abiotic (drought) stresses were used to produce four cDNA libraries: control, drought, nematode infection and combined stresses, that were sequenced by Illumina HiSeq 4000 technology. Each library produced over 30 million raw reads, with an average of 10% reduction on the number of reads after the removal of adapters and poor quality sequences (Table 1). Trimmed reads of *A. stenosperma* were quantified into the available reference genome of *A. duranensis* (Bertioli et al., 2016), with an average of 60% of reads successfully quantified in the reference genome, as previously observed (Mota et al., 2018). An average of 25,000 genes could be quantified per library, representing approximately 70% of the total number of genes identified in *A. duranensis* (36,743 gene models) genome (Table 1).

**Table 1**: Illumina HiSeq 4000 sequencing data from *A. stenosperma* quantified in the *A. duranensis reference* genome.

| | | Nematode | Drought | Combined | Control |
|---|---|---|---|---|---|
| Replicate 1 | Number of raw reads | 34,173,582 | 35,004,558 | 37,568,042 | 29,324,662 |
| | Number of reads after trimming | 30,028,088 | 31,823,728 | 33,512,716 | 26,572,592 |
| | Number of ıtified reads into the reference | 10,631,810 | 9,189,171 | 10,385,619 | 8,469,816 |
| | Total number of gene models | 25,446 | 24,848 | 25,386 | 25,943 |
| Replicate 2 | Number of raw reads | 34,236,660 | 37,275,858 | 33,190,158 | 36,116,360 |
| | Number of reads after trimming | 31,699,592 | 33,061,206 | 30,115,720 | 31,430,734 |
| | Number of ıtified reads into the reference | 11,653,431 | 9,679,977 | 8,651,041 | 9,593,936 |
| | Total number of gene models | 25,715 | 25,266 | 25,227 | 24,825 |
| Replicate 3 | Number of raw reads | 31,200,332 | 40,506,646 | 36,428,454 | 34,885,968 |
| | Number of reads after trimming | 29,136,428 | 37,119,422 | 31,889,266 | 30,148,320 |
| | Number of ıtified reads into the reference | 10,650,358 | 10,935,791 | 9,641,333 | 8,735,584 |
| | Total number of gene models | 25,330 | 25,695 | 24,986 | 24,998 |

| | | | | | |
|---|---|---|---|---|---|
| DEG * | Number of genes with FDR < 0.05 | 1,252 | 6,103 | 3,346 | N/A |
| | Number of genes Log2FC >2 or < -2 | 193 | 1,034 | 570 | N/A |

* Differential gene expression analyses of the three treatments against the control

A total of 7,543 significantly expressed (FDR < 0.05) genes were identified in at least one treatment compared to the control condition. From these 1,410 genes were differentially expressed ($Log_2FC$ 2> or <-2) or DEGs distributed in the drought (1,034), combined stresses (570) and nematode infection (193) libraries (Figure 1A; Table 1). Overall, a reduction in gene expression was observed in all treatments, with an average of 60% of the genes being downregulated (Figure 1B). The highest number of exclusive genes (674) was identified in the drought , followed by combined stresses (209) and nematode (154) treatments. Most genes in common to two or more treatments were identified between drought/combined stress (348), while combined stress/nematode and drought/nematode shared only 27 and 26 genes respectively. Among all genes in common, only 14 were shared by all three treatments (Figure 1A).

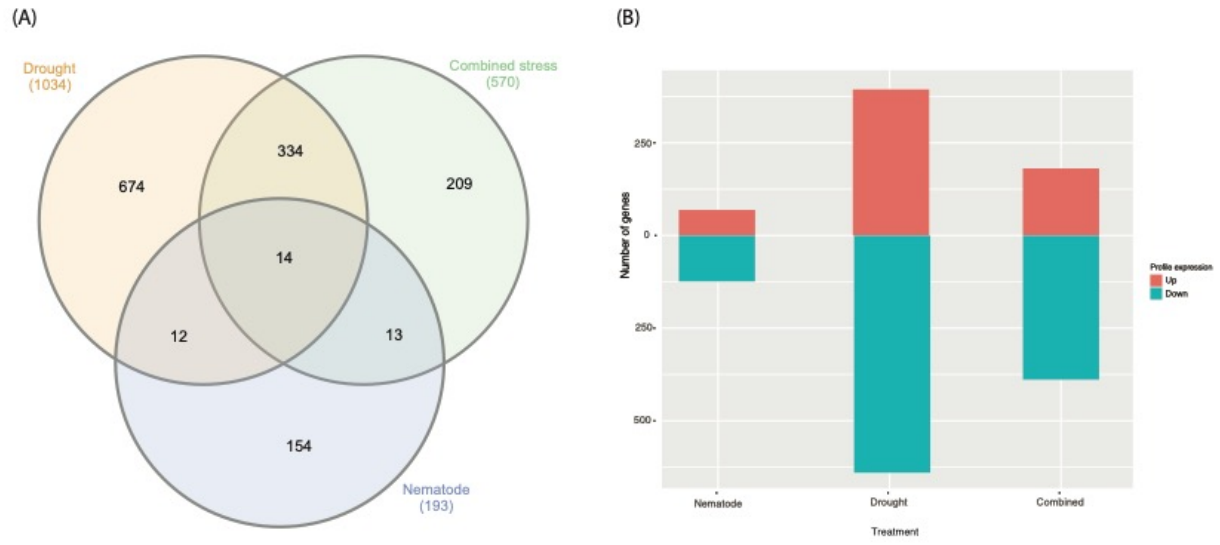



**Figure 1:** Differentially Expressed Genes (DEGs) in *A. stenosperma* subjected to drought, nematode and combined stresses compared to the control. (A) Venn diagram showing the number of exclusive and shared DEGs among treatments. (B) Number of genes per treatment and their expression trend. Red bars indicate upregulated genes and the blue bars indicates downregulated genes. DEGs are differentially expressed genes at Log2FC > 2 or < -2 and FDR < 0.05.

DEGs identified under drought and combined stresses treatments showed the greatest accordance on their expression profiles (up or downregulation) with 98% to 74% of genes displaying the same trend (Figure 2). On the other hand, only 24% of the upregulated and 49% of the downregulated DEGs identified in the nematode treatment showed an expression trend compatible with the other

treatments. In addition, the expression magnitude of DEGs in the drought and combined treatments was also higher than in the nematode treatment (Figure 2).

This data shows that in *A. stenosperma*, biotic and abiotic stress activate different sets of responsive genes, which are also different from those triggered when the stress occurs in combination.

This small number of genes displaying the same expression profile in response to drought, nematode and combined treatments suggests that, in addition to using different armoury in response to biotic and abiotic stress, plants also respond differently to stresses applied individually or in combination. This specialized response was previously observed in *Arabidopsis* plants (Rasmussen et al., 2013).

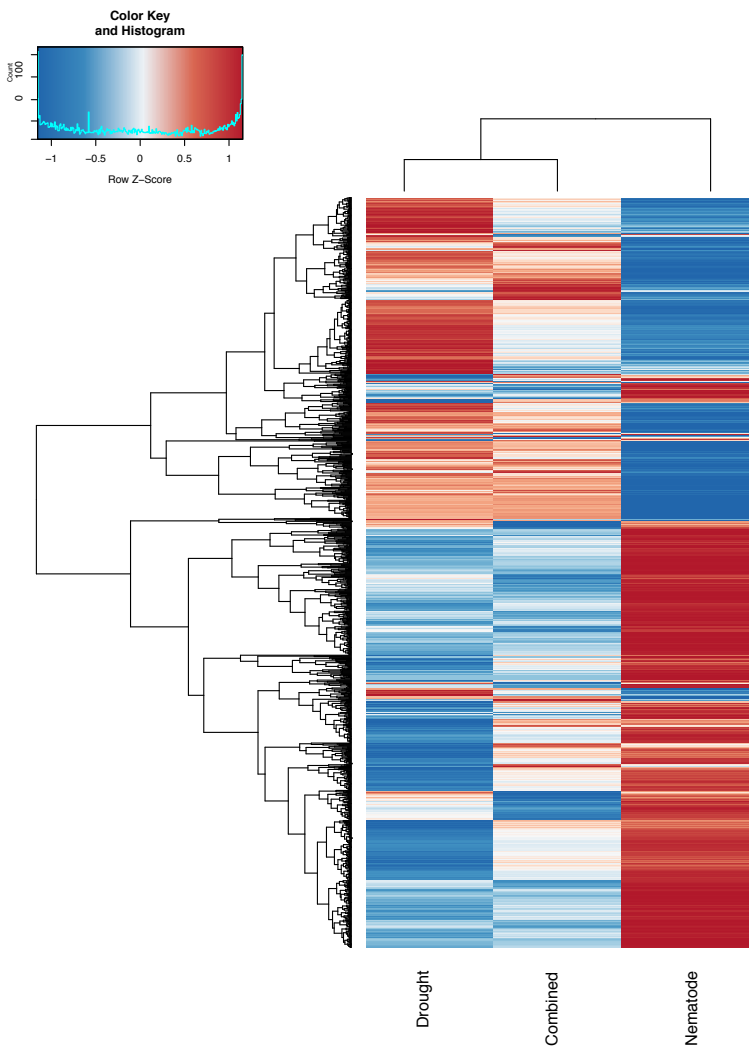



**Figure 2**: Expression profile of *A. stenosperma* DEGs (FDR < 0.05 and $\log_2$FC > 2 or < -2) under biotic and abiotic treatments. Each column represents drought, combined and nematode treatments. The colour key values are represented in Log2FC.

## Functional analyses of differentially expressed genes (DEGs)

*Gene Ontology analysis*. DEGs identified in drought, nematode and combined stresses were classified into GO categories. Whilst for the nematode and combined stresses treatments, most enriched terms were limited to the molecular function category, genes in the drought treatment enriched all three gene ontology categories (molecular function, biological process and cellular component). None of the enriched terms were common to all three treatments, suggesting

specialized responses. In the GO categories enriched by nematode infection we observed terms related to ROS, such as iron ion binding (GO:0005506) and oxidoreductase activity (GO:0016705), whilst for the combined stresses cysteine-type activity (GO:0008234), hydrolase activity (GO:0016788) and thiolester hydrolase activity (GO:0016790) categories were the most enriched. These enzymes activity are often associated with stress, regulating the physiological homeostasis of the cell (Álvarez et al., 2012).

Contrastingly, in the drought treatment, all the three main gene ontology categories were enriched, with molecular function showing the highest number of enriched terms. Mostly, terms involved with "oxireductase, antioxidant, and hydrolase activities" and "heme, cation, metal and magnesium binding". In the biological processes category, most terms were enriched for metabolic processes (inositol, alcohol and polyol), response to stress and defense response. In cellular component, high levels of transcripts were annotated for membrane parts (GO: 0044425) (Figure 3).

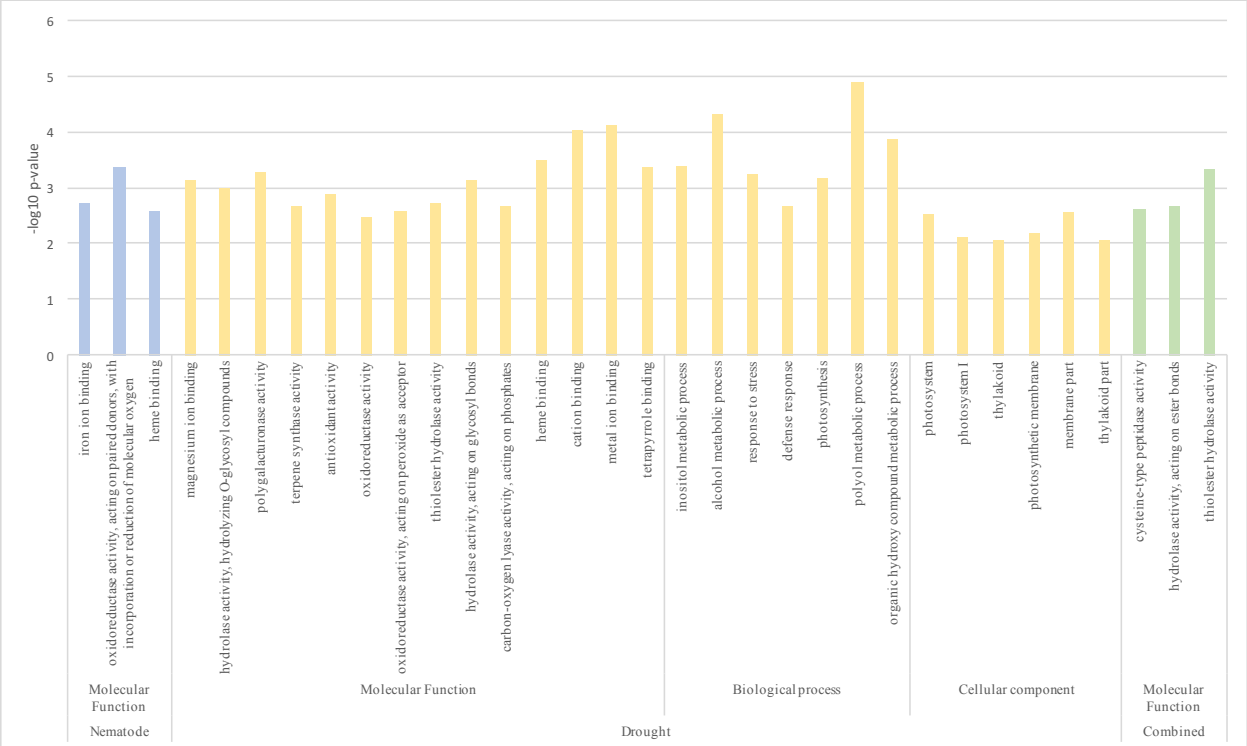



**Figure 3:** Gene Ontology enrichment analysis. GO terms enriched for drought, nematode and combined treatments. DEGS responsive to nematode treatment is represented by blue bars, to drought by yellow bars and to the combined treatment by green bars. The values are in –log10 of the p-value.

*Transcription factors families.* In total, 78 TFs belonging to 31 different families were DEGs in at least one of the three treatments applied. The drought treatment encompassed most of the exclusive TFs (40), while the combined stresses and the nematode infection contained only 13 and three exclusive TF genes. Most of the TFs identified in the drought challenge belonged to the MYB family (Myeloblastosis), with 18 genes annotated, followed by bHLH (16) and WRKY (7). The combined stresses treatment contained three exclusive TF families: GRAS, TALE and DBB, which are involved in plant defense response against abiotic and biotic stresses in plants (Bogdanove et al., 2010; Ma et al., 2010; Vinson et al., 2018; Xu et al., 2015; Zhao et al., 2019). Eleven exclusive TF families were found to be differentially expressed for drought stress alone: NF-YA, FAR1, Trihelix, NF-YB, Dof, HSF, Nin-like, ARF, AP2, CO-like, and ARR-B, with NF-YA previously reported as responsive to drought stress in *A. stenosperma* (Vinson et al., 2018). Although in the nematode treatment there was no exclusive TF families, it triggered the upregulation of two genes from WRKY and one from bHLH families that were exclusively differentially expressed. Contrastingly, the exclusive DEGs from the WRKY family showed a downregulation for the other two treatments (drought and combined).

All but three TF families (AP2, ERF and NAC) showed contrasting expression profile between nematode and drought treatments, suggesting their possible involvement in the combined stresses response (Figure 4).

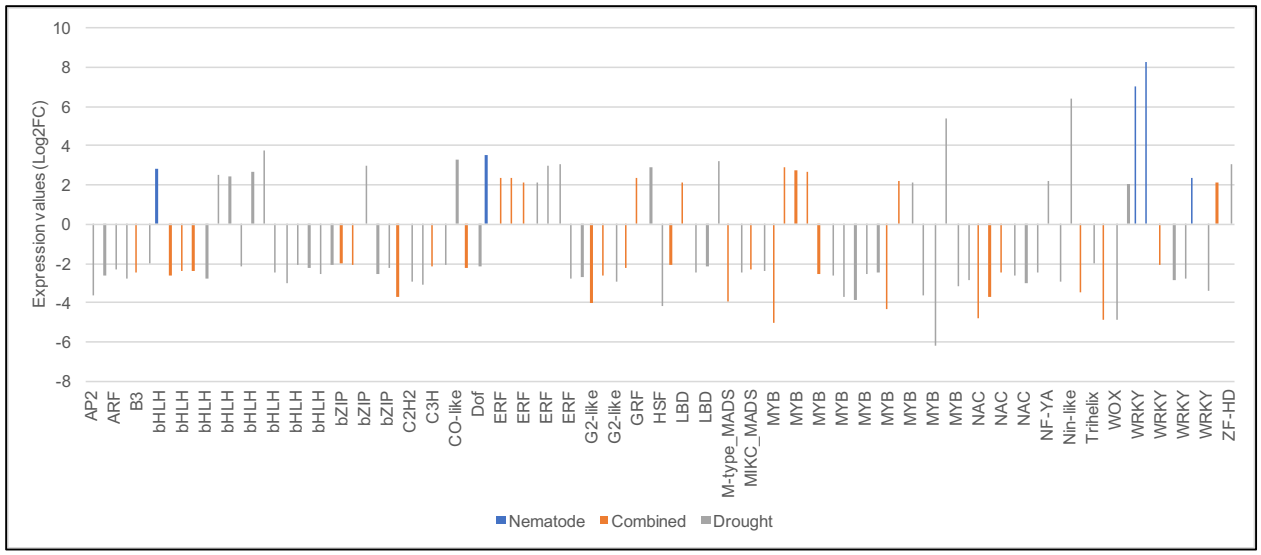


**Figure 4:** Expression values of *A. stenosperma* transcription factors under drought, nematode and combined stresses treatments. The values are in $Log_2FC$.

**Genes exclusively expressed in combined stresses**

Overall, the arsenal and expression trend of genes involved in plant response to drought and combined stresses were more akin than those in response to nematode infection alone. Remarkably, from the 570 DEGs responsive to both stresses simultaneously, almost 30% (209) were exclusive to the combined stresses treatment. This suggests the activation of a diverse molecular response, possibly leading to different metabolic and physiological feedbacks, when the plant is exposed to both stresses simultaneously. These combined stresses exclusive genes were mostly downregulated (136 genes), with an expression magnitude ranging from -10 to 6 ($Log_2FC$). Despite not being DEGs, they also showed a similar downregulation trend under drought stress, while an opposite expression trend under nematode infection (Figure 5).

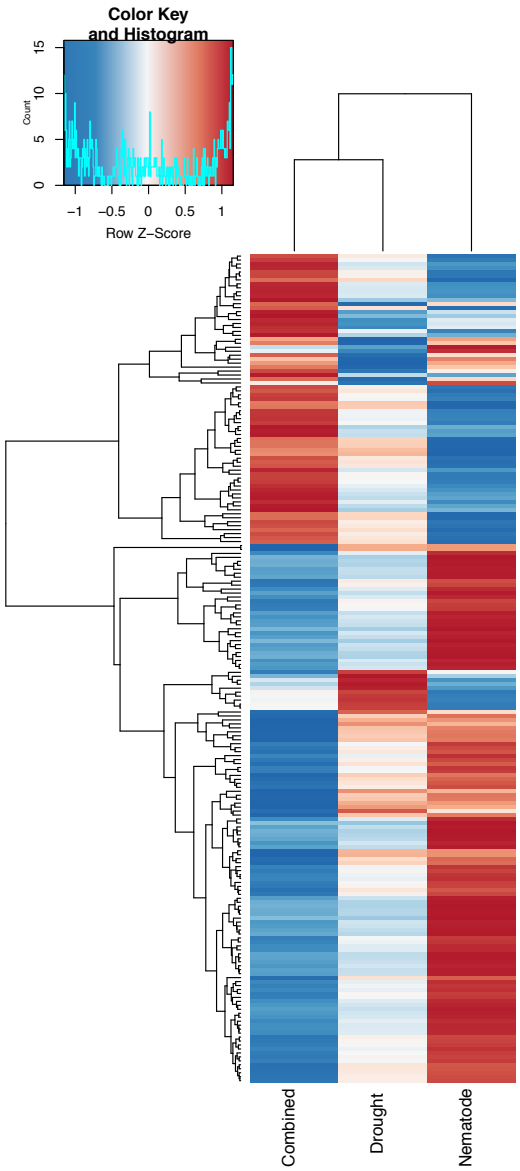



**Figure 5:** Transcript expression of DEGs exclusive to combined stresses treatment under combined and single stresses. The columns represent the combined stress, drought stress and nematode stress, respectively. The values are in $\log_2$FC.

The functional annotation of the DEGs exclusive to combined stresses using Mercator showed their distribution in 15 categories with most genes annotated to solute transport, RNA biosynthesis, external stimuli and phytohormones (Figure 6).

The depiction of exclusive DEGs in Mapman categories showed to be very similar for the combined and drought stresses, albeit different for DEGs under nematode stress, which shows an

overall dominance of genes regulated by the drought treatment. Most genes assigned to the RNA biosynthesis category belonged to the MYB family of transcription factors, which are widely known to be involved in drought responses through the abscisic acid signaling pathway (Seo et al., 2009). We could also classify genes involved in the production of the phytohormones cytokinin and ethylene. Cytokinin has previously been described as negatively regulated in drought stress, as decreased levels of cytokinin leads to an apical root dominance, that together with high levels of ABA favours the plant adaptation to drought stress (O'Brien and Benková, 2013). Contrastingly, during nematode infection genes involved in conjugation, degradation and synthesis of cytokinin were upregulated (regardless their FDR), as they are involved with giant cell formation induced by the nematode (Siddique et al., 2015). The identification of upregulated genes encoding ethylene responsive factors (ERF) and those involved in ethylene synthesis in all treatments, indicates that different phytohormone pathways are triggered by these stresses and their combinations.

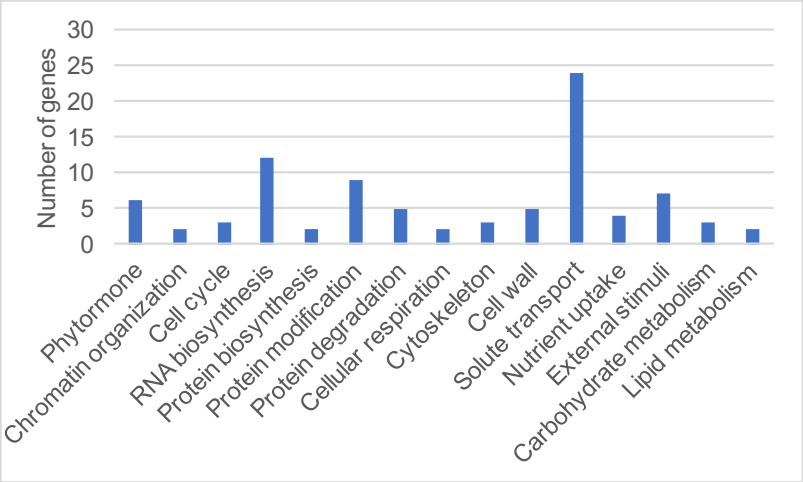



**Figure 6:** Functional classification by Mercator of DEGs exclusive to combined stresses treatment.

**Genes regulated by biotic and abiotic stresses**

In total, we found 14 genes regulated under drought, nematode and the combined stresses in common. These genes could be assigned to functional categories related to plant defense such as: phytohormones (Ethylene and salicylic acid), cell wall biosynthesis (expansin like-B), oxireductases (L-aspartate oxidase (FAD)), transmembrane enzymes, hydrolases and cytochrome

p450. Eight of these genes (64%) showed a similar expression profile under all three stresses, with the majority being upregulated (Figure 7).

Only two genes were downregulated in all treatments, transmembrane ATPase (Aradu.EAR0Q) and Adenosine deaminase (Aradu.7M8TY), and although poorly studied in plants have been linked to responses to different diseases in other organisms (Ndimba et al., 2005; Weretilnyk, 2002).

Five genes (Salycicate carboxymethyltransferase, phytepsin, leguminosin, glutamine amidotransferase and cytochrome p450) previously associated with plant stress responses showed a contrasting expression profile between nematode and the other two treatments (Figure 7).

Previous studies described the involvement of aspartic proteins in response to different abiotic stresses (Milisavljević et al., 2009). Here, an aspartic protein (Aradu.A3AX6) was strongly induced in the drought and combined stresses (9-fold), while downregulated at 7-fold in the nematode treatment. Likewise, another protein known for its role in abiotic stress responses, the glycine rich protein (GRP: Aradu.EZW4U), or leguminosin secreted protein, was downregulated for nematode infection and upregulated in the combined and drought treatments.

An Expansin like-b (Aradu.KH5IZ) and a chitinase (Aradu.4196P) DEG showed an upregulated profile in all three treatments applied in this study. Previously, another member of the expansin like-b subfamily in Arachis, *EXLB8,* was identified as responsive to both drought, UV and nematode infection in different *Arachis* species (Carmo et al., 2019; Guimaraes et al., 2017). Guimaraes et al. (2017) also showed that *EXLB8* caused a reduction on the number of *M. arenaria* galls when overexpressed in soybean (*Glycine max)* composite plants.

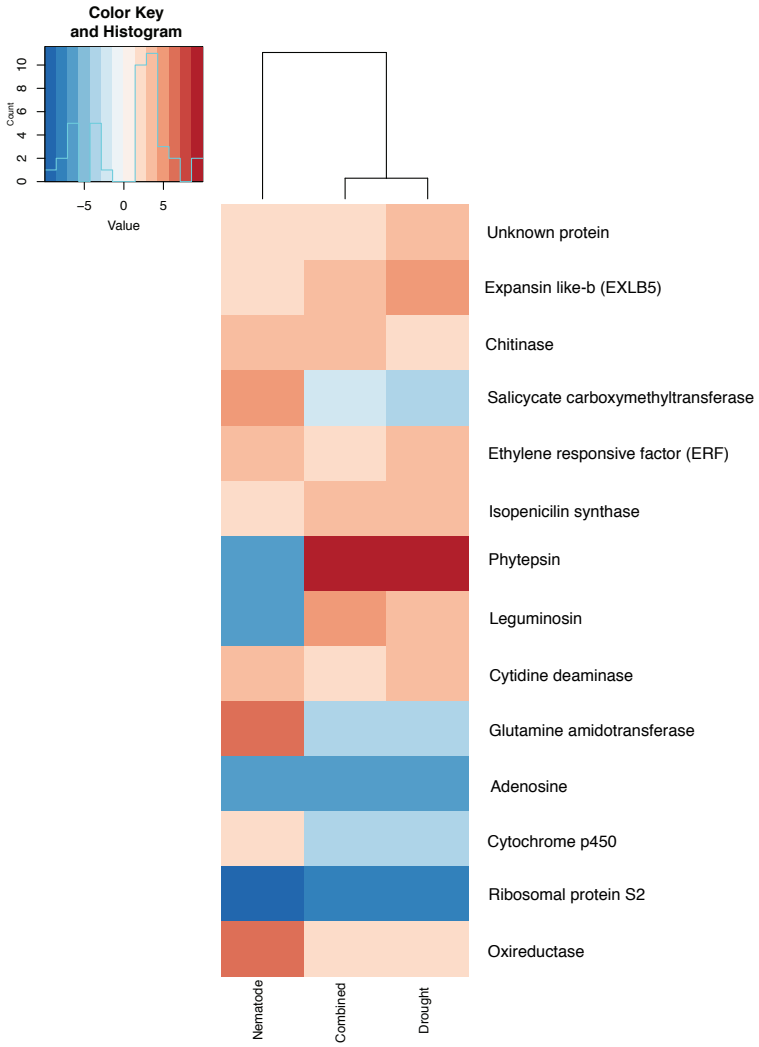



**Figure 7**: Relative expression values of 14 genes expressed in common under nematode, combined and drought stresses. The colour key values are in $\log_2$FC.

### *EXLB5* relative expression analysis

Amongst the 14 genes responsive to drought, nematodes and combined stresses, an expansin like-b gene (*EXLB5*), an ortholog of *AdEXLB8* (Guimaraes et al., 2017) was upregulated under all the above stresses. The validation of this expression profile by qRT-PCR in *A. stenosperma* roots submitted to all the three treatments (drought, nematode and combined stress) corroborated this upregulation pattern (Figure 8). The highest relative expression for *EXLB5* was observed in plants submitted exclusively to drought (23.19), while the combined and nematode treatments showed

proximate expressions (3.18 and 4.57, respectively) (Figure 8). This is in accordance to the overall gene expression defense response observed in *Arachis* to drought, that included not only a greater number of responsive genes, but also genes with higher expression magnitudes, when compared to nematode and combined stresses.

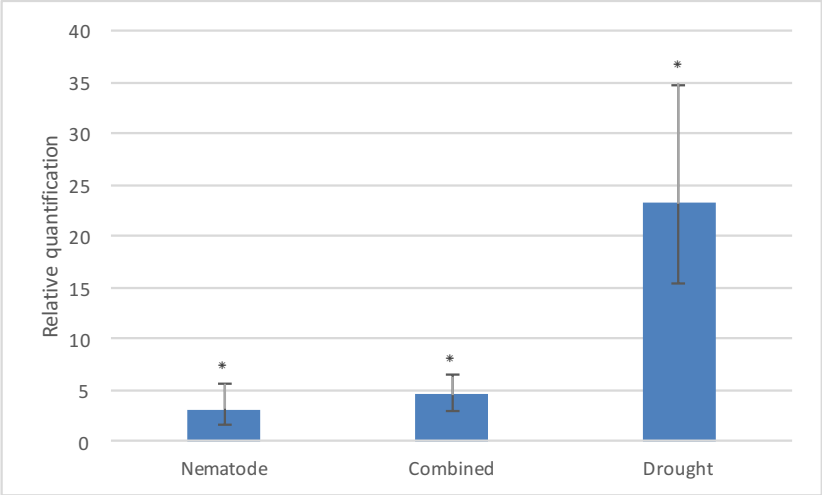



**Figure 8:** Relative expression of *EXLB5* under drought, combined stresses and nematode infection. Statistically significant regulated genes (*) and error bars represented.

## 5. DISCUSSION

In nature, plants are constantly being challenged by combined stresses, making the identification of the physiological, molecular and the *crosstalks* among the signaling pathways in response to various stresses crucial to produce more adapted cultivars. Nonetheless, plant transcriptional responses to multiple stresses, in particular to abiotic and biotic simultaneously, has been remarkably understudied, with most research covering drought and other abiotic hazards or various fungi (Suzuki et al., 2014). Also, few studies analysed the whole plant transcriptome under stress, which provides a more robust overview of the plant mechanisms of response to these challenges (Van Verk et al., 2013). In this study, we analysed *A. stenosperma* responses to individual and combined root-knot nematode and drought stresses, in order to outline genes and molecular mechanisms involved in the responses to individual and combined challenges.

### Single and combined stress responses

Overall, the differential gene expression analysis of *A. stenosperma* roots subjected to drought, nematode, and combined stresses showed distinct metabolic pathways, and hormonal signaling for each treatment applied. The higher number of exclusive genes in comparison to shared genes among treatments suggested a highly specialized plant response to each stimulus.

We used previous transcriptome studies on wild *Arachis* spp. subjected to individual stresses (drought and nematode) (Mota et al., 2018; Vinson et al., 2018), to evaluate if the sum of genes regulated in individual stresses was equivalent to the response observed in the combined stresses. Our data showed that in *A. stenosperma* roots, the application of combinatorial stresses led to a remarkable change on the identity and magnitude of transcripts expressed, in comparison to the stresses applied separately. This is in accordance to recent studies (Kissoudis et al., 2016; Olivas et al., 2016) that showed that plant responses to combinations of two or more stress conditions is not only unique and diverse from the response of plants to each stress applied individually, but are largely controlled by different, and sometimes opposing, signaling pathways that may interact and inhibit each other (Suzuki et al., 2014).

Findings in this study show that the biggest change in transcriptional responses, represented by the number of DEGs, was observed in *A. stenosperma* plants subjected to drought (1,034 DEGs). This was expected, as extensive molecular, physiological and adaptive changes occur in plants under drought conditions, such as ROS, calcium waves, electrical, and hydraulic (stomatal closure) signals that can occur within seconds or gradually (Kollist et al., 2019; Ramegowda and Senthil-Kumar, 2015).

These acclimation responses require energy and resources, and lead to an enormous transcriptional reprograming of the plant cell (Mittler and Blumwald, 2010).

Most drought responsive DEGs found in this analysis were included in oxireductase and antioxidant activities categories, with the upregulation of genes such as cytochrome p450 and thyoredoxin indicating that the ABA responsive pathway is the most prevalent responsive metabolic pathway utilized by *A. stenosperma* in response to water shortage.

The second major overall transcriptional change in *Arachis* roots occurred under the combined stresses treatment, with 570 DEGs identified. Although variations have been reported (Bowler and Fluhr, 2000; Mittler and Blumwald, 2010; Suzuki et al., 2014), it is generally assumed that plants

exposed to abiotic stress conditions were more prone to the successful infection of pathogens (Atkinson and Urwin, 2012). This antagonistic behavior was recently observed in pine seedlings, in which resistance to herbivore insects, generally governed by JA, was altered by drought (Suárez-Vidal et al., 2019). In addition to ABA, reports have demonstrated the involvement of JA in water stress responses (Brossa et al., 2011; de Ollas et al., 2015; De Ollas et al., 2018), with its accumulation leading to an enhancement on ABA concentration, with JA-dependent signaling remaining active even when its accumulation is transient (Wasternack, 2014). ABA and JA signaling pathways can interact at several points, and are regulated by MYC2 and MYB transcription factors, suggesting a role for JA in the response to water stress (Kazan and Manners, 2013).

Previous studies on the transcriptional profile of *A. stenosperma* roots infected with *M. arenaria* showed that most defense related genes appeared between the 6$^{th}$ and the 9$^{th}$ DAI (days after inoculation) (Guimaraes et al., 2015). Here, most of the DEGs identified at the 8$^{th}$ DAI were included in functional categories related to ROS and metal binding, and participate in the hormonal defense response led by the jasmonic pathway (JA). These transcripts included transcription factors related to JA (WRKY and bHLH) and an ERF transcription factor, related to the ET pathway. The transcription factor bHLH, from the MYC family of transcription factors, is known as an important regulator of the *crosstalk* and antagonism between SA and JA pathways, which regulates the plant response to different pathogens (Kazan and Manners, 2013). The occurrence of an exclusive bHLH gene (Aradu.VS58Y) and its strong upregulation under nematode infection alone, is also an indicative of the JA involvement on the resistance of *A. stenosperma* to *M. arenaria*. This is in accordance to Mota *et al.*, (2018), that showed that the main response to RKN in *Arachis* was regulated by JA and ET and was also reported in sweet potato infected with *M. incognita* (Lee et al., 2018). The induction of WRKY, that can bridge the response to both biotic and abiotic stresses, and trigger PTI and ETI defense responses (Bai et al., 2018; Li et al., 2015), suggests its participation in this *crosstalk*, as all WRKY genes that were activated during the nematode treatment showed an opposite expression profile under drought and combined stresses treatments (Figure 4).

In the present study, we observed an overall reduction on the number and expression magnitude of genes responsive to nematode infection in *A. stenosperma* roots subjected to combined stress (drought and nematodes) in comparison to the imposition of nematode infection alone. This included the reduction of defense related genes, such as pathogen-related and NBS-LRR genes, which have previously been correlated with resistance in this species (Guimaraes et al., 2015; Mota et al., 2018). We suggest that the antagonistic effects of drought stress on *A. stenosperma* nematode resistance, might be due to its interference in the plant's ability to produce chemical defenses mediated by JA-ET in response to damage, causing functional impairment and reducing the plant resistance against the nematode.

**Plant responses to stress combination includes more than one hormone pathway**

In the combined stresses treatment, we observed an overall downregulation of genes related to JA pathway, including bHLH transcription factors, cullins and cystatins (Dutt et al., 2011) and of genes regulating both the ABA-independent (NAC branch) and ABA–dependent (bZIP/MYC branch) hormone pathways. Concurrently, Ethylene and ABA-dependent (MYC branch) related genes (ERF and MYB transcription factors) were upregulated (Figure 9).

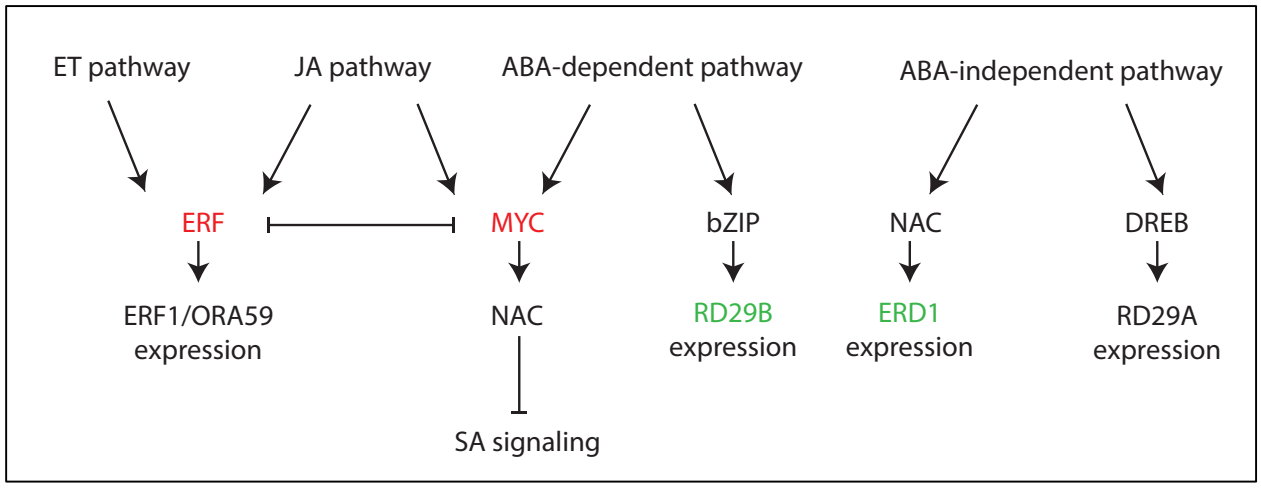



**Figure 9:** Schema of the crosstalk among hormone-regulated defense pathways. ERF and MYC branches of the Jasmonic acid (JA) pathway. The ERF branch interacts with the ethylene (ET) activating the ERF transcription factor and subsequently the expression of marker genes ERF1 and ORA59. The MYC branch interacts with the ABA-dependent pathway, activating the NAC transcription factors, which blocks the

signaling to the salicylic acid (SA) via. The second ABA-dependent branch activates the expression of RD29B through the transcription factor bZIP. The ABA-independent pathway can activate the expression of ERD1 and RD29A through NAC and DREB transcription factors, respectively. Adapted from Bostock et al., (2014).

In addition to ABA-dependent and -independent mediated signaling events, alternative pathways encoding ethylene responsive factors (ERF), and genes involved in ethylene synthesis were also identified, such as ERF TFs and ORA59 (Dubois et al., 2013, 2018; Pre et al., 2008), suggesting a broader crosstalk between regulatory pathways that govern biotic and abiotic responses.
However, ERF is a big TF family, with 122 ERF members only in *Arabidopsis*, and can be assigned to different hormones pathways, being also induced by JA, making the depiction of JA-ET-ABA communication and its interaction under water stress conditions very complex (Bostock et al., 2014).
Overall, we found that the defense response displayed by *A. stenosperma* roots subjected to nematode/drought combined stresses, tend to follow the JA-ABA-ET schematic model adapted from Bostock *et al.*, (2014), that puts ABA as the central regulator in the stress response (Figure 9). In this representation, we suggest that the ABA/MYC branch is activated by both, JA induced in response to the nematode attack and ABA molecules originated in response to drought stress. On its turn, JA and ABA induce the TF NAC which activate ERD1 and induce the production of VSP1 and LOX which are osmoprotectants that contribute to increase plant drought tolerance. Concomitantly, JA pathway can also activate the ERF branch, which will activate the ERF1/ORA59 transcription factors and induce downstream pathogen related (PR) proteins, such as PDF1.2, known as a defense protein against biotic stress.

Despite some advances on plant hormone crosstalk mechanisms, how the plant perception is altered due to the combination of stresses is yet unknown. Recent studies have suggested that calcium ($Ca^{+}$) signaling acts as a hub in the plant response to abiotic and biotic stresses (Ku et al., 2018). One of the most studied $Ca^{+}$ modulated gene is calmodulin (*CaM*), which can act as part of a $Ca^{+}$ signal transduction pathway by modifying its interactions with various target proteins such as kinases or phosphatases, and have different roles in the regulation of plant response to stresses (Bouché et al., 2005). The expression of *CaM* genes can be induced by different stresses and

hormones, with *Arabidopsis* lines overexpressing calmodulin *AtCaM* leading to an increased sensitivity to ABA and consequently enhanced tolerance to abiotic stresses (Magnan et al., 2008), whilst *AtCaM* decreased the resistance to *Pseudomonas* infection (Leba et al., 2012). In the present study, we observed an upregulation of *CaMBP* (Calmodulin binding proteins), which we suggest that might contribute to the antagonistic response between the nematode and the combined/drought treatments.

**Genes responsive to combined stresses for crop improvement**

Findings in the present study show that plant responses to individual and combined stresses (nematodes and drought) are highly specialized, with few genes in common. Understanding the complex regulation of biotic and abiotic responses and their interactions are crucial to avoid trade-off effects that limit plant growth and yield, and to allow a more efficient biotechnological crop improvement. The search for genes to be used for plant genetic engineering has mainly focused on the solution of individual stresses, with few examples aiming the identification of genes responsive to more than one stress. Here, we found 14 genes responsive to drought, *Meloidogyne* infection and the combined stresses. These are interesting candidate genes that once validated in planta can contribute to plant resistance against single or joint stresses (abiotic, biotic or combined). Amongst these, an expansin-like b gene (*AdEXLB5*) was selected for further analyses, as previous studies from our group have demonstrated that another representative of the EXLB sub-family, the *AraEXLB8* gene, was highly induced in response to water deficit in two different *Arachis* species, *A. duranensis* and *A. magna* (Brasileiro et al., 2015; Guimaraes et al., 2017) and modulated in response to RKN infection (Guimaraes et al., 2017). *AdEXLB5* showed an upregulation for both abiotic and biotic stresses and their combinations, constituting a promising candidate gene for *in planta* validation and plant genetic engineering against multiple stresses.

# Chapter III

Candidate genes: Gene family characterization of dehydrin and expansin genes of *Arachis* spp.

**Contrasting effects of wild Arachis dehydrin under abiotic and biotic stresses**

Published in Frontiers in Plant Science – 2019

Ana Paula Zotta Mota, Thais Nicolini de Oliveira, Christina Cleo Vinson, Thomas Christopher Rhys Williams, Marcos Mota do Carmo Costa, Ana Claudia Guerra Araujo, Etienne G.J. Danchin, Maria Fatima Grossi-de-Sá, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 April 2019
doi: 10.3389/fpls.2019.00497

# Contrasting Effects of Wild *Arachis* Dehydrin Under Abiotic and Biotic Stresses

*Ana Paula Zotta Mota[1,2], Thais Nicolini Oliveira[1,3], Christina Cleo Vinson[1,3], Thomas Christopher Rhys Williams[3], Marcos Mota do Carmo Costa[1], Ana Claudia Guerra Araujo[1], Etienne G. J. Danchin[4], Maria Fatima Grossi-de-Sá[1], Patricia Messenberg Guimaraes[1] and Ana Cristina Miranda Brasileiro[1]**

[1] EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, Brazil, [2] Departamento de Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [3] Departamento de Botânica, Universidade de Brasília, Brasília, Brazil, [4] INRA, Université Côte d'Azur, CNRS, ISA, Sophia-Antipolis, France

Plant dehydrins (DNHs) belong to the LEA (Late Embryogenesis Abundant) protein family and are involved in responses to multiple abiotic stresses. DHNs are classified into five subclasses according to the organization of three conserved motifs (K-; Y-; and S-segments). In the present study, the DHN protein family was characterized by molecular phylogeny, exon/intron organization, protein structure, and tissue-specificity expression in eight Fabaceae species. We identified 20 DHN genes, encompassing three ($Y_nSK_n$, $SK_n$, and $K_n$) subclasses sharing similar gene organization and protein structure. Two additional low conserved DHN Φ-segments specific to the legume $SK_n$-type of proteins were also found. The *in silico* expression patterns of DHN genes in four legume species (*Arachis duranensis*, *A. ipaënsis*, *Glycine max*, and *Medicago truncatula*) revealed that their tissue-specific regulation is associated with the presence or absence of the Y-segment. Indeed, DHN genes containing a Y-segment are mainly expressed in seeds, whereas those without the Y-segment are ubiquitously expressed. Further qRT-PCR analysis revealed that, amongst stress responsive dehydrins, a $SK_n$-type DHN gene from *A. duranensis* (*AdDHN1*) showed opposite response to biotic and abiotic stress with a positive regulation under water deficit and negative regulation upon nematode infection. Furthermore, transgenic *Arabidopsis* lines overexpressing (OE) *AdDHN1* displayed improved tolerance to multiple abiotic stresses (freezing and drought) but increased susceptibility to the biotrophic root-knot nematode (RKN) *Meloidogyne incognita*. This contradictory role of *AdDHN1* in responses to abiotic and biotic stresses was further investigated by qRT-PCR analysis of transgenic plants using a set of stress-responsive genes involved in the abscisic acid (ABA) and jasmonic acid (JA) signaling pathways and suggested an involvement of DHN overexpression in these stress-signaling pathways.

**Keywords: *Arabidopsis*, drought, freezing, genome-wide, *Meloidogyne*, root-knot nematode**

OPEN ACCESS

***Edited by:***
*Alejandra A. Covarrubias, National Autonomous University of Mexico (Morelos), Mexico*

***Reviewed by:***
*Steffen P. Graether, University of Guelph, Canada*
*Pedro Carrasco, University of Valencia, Spain*

***Correspondence:**
*Ana Cristina Miranda Brasileiro*
*ana.brasileiro@embrapa.br*

***Specialty section:***
*This article was submitted to Plant Abiotic Stress, a section of the journal Frontiers in Plant Science*

***Received:*** *17 December 2018*
***Accepted:*** *01 April 2019*
***Published:*** *18 April 2019*

***Citation:***
*Mota APZ, Oliveira TN, Vinson CC, Williams TCR, Costa MMdC, Araujo ACG, Danchin EGJ, Grossi-de-Sá MF, Guimaraes PM and Brasileiro ACM (2019) Contrasting Effects of Wild Arachis Dehydrin Under Abiotic and Biotic Stresses. Front. Plant Sci. 10:497. doi: 10.3389/fpls.2019.00497*

## INTRODUCTION

The dehydrin (DHN) family of proteins, or dehydration proteins, belongs to group II of the LEA (Late Embryogenesis Abundant) proteins and is considered the most important LEA group due to its involvement in tolerance to several abiotic stresses. DHNs are defined by the presence of three conserved motifs, named the K-, Y-, and S-segments, that classify DHNs into five subclasses

according to their number and organization: $Y_nSK_n$; $K_n$; $SK_n$; $K_nS$; and $Y_nK_n$ (Close, 1996). DHN proteins typically contain at least one copy of the signature K-segment, a 15-amino acid lysine-rich repeat (EKKGIMDKIKEKLPG) near the C-terminus (Malik et al., 2017). The two other conserved motifs, despite being common in DHN, are not essential to characterize the protein: the Y-segment ([V/T]D[E/Q]YGNP), near to the N-terminus, and the S-segment, a group of four to eight serine residues (Malik et al., 2017). The K-, Y-, and S-segments are separated by poorly conserved motifs, called Φ-segments, that are usually rich in glycine and polar amino acids (Graether and Boddington, 2014).

DHNs often accumulate in vegetative tissues of higher plants in response to different types of abiotic stresses, such as low temperature, drought, salinity, or wounding, and this response represents part of the molecular arsenal developed by plants to withstand the damaging effects of the intracellular water loss caused by these stresses (Graether and Boddington, 2014). However, despite considerable research, the mechanisms underlying DHN responses to dehydration remain unclear and each type of DHN might have a specific function. It is generally accepted that DHNs act as chaperones to prevent the aggregation, damage, and inactivation of proteins during stress imposition (Hara et al., 2005; Kovacs et al., 2008). The highly disordered nature of DHN structure also contributes to biochemical properties that permit these proteins to function in enzyme activity protection, membrane binding and stabilization, ROS scavenging, and ion sequestration (Banerjee and Roychoudhury, 2016).

Plant DHN gene expression is generally modulated in response to abiotic stress and varies according to the type of condition imposed and the subclass to which they belong (Abedini et al., 2017). In particular, the DHN $SK_n$-types are commonly associated with cold tolerance and found near the plasma membrane where they exert a protective role (Graether and Boddington, 2014). In addition to cold stress, there is evidence that DHNs belonging to the $SK_n$-type can also enhance the ability of plants to cope with multiple and simultaneous abiotic stresses. For example, the overexpression in tobacco plants of $SK_n$-type genes isolated from freezing-tolerant species (arctic chickweed and Chinese mei) improved tolerance to cold, drought, salt, and osmotic stress (Hill et al., 2016; Bao et al., 2017). Also, a $SK_3$ DHN from wild tomato and a $SK_2$ from rubber-tree enhanced tolerance to more than one abiotic stress in transgenic tomato and *Arabidopsis* plants, respectively (Liu H. et al., 2015; Cao et al., 2017). Likewise, the overexpression of a $SK_3$ DHN from a stress-tolerant *Musa* genotype led to superior performance in a commercial cultivar of transgenic banana under both drought and salinity (Shekhawat et al., 2011).

In contrast, some $SK_n$-type genes appear to be stress-specific. For instance, those isolated from cold-adapted species (wheat and wild potato), that have been shown to enhance tolerance to freezing in transgenic strawberry and cucumber, respectively (Houde et al., 2004; Yin et al., 2006); or that isolated from a drought-tolerant wild olive species, which increased tolerance to water deficit in transgenic *Arabidopsis* (Chiappetta et al., 2015). Similarly, a sorghum $SK_3$ increased protection against oxidative stress when overexpressed in tobacco (Halder et al., 2018), and a $SK_3$ from the halophyte *Ipomoea pes-caprae* increased the salt tolerance when overexpressed in *Arabidopsis* (Zhang et al., 2018).

Unlike in abiotic stress, the involvement of plant DHNs in biotic stress is poorly studied, with few studies showing modulation of DHN expression associated with responses to fungal infection either on its own or in association with abiotic stress (Turco et al., 2004; Yang et al., 2012). Although DHNs are known to respond to wounding and exogenous hormones that play a vital role in pathogen defense signaling and disease resistance, such as Abscisic acid (ABA), Jasmonic acid (JA), Salicylic acid (SA), and Ethylene (ET), their role in plant defense against pathogens remains to be elucidated (Shen et al., 2004; Hanin et al., 2011; Rosales et al., 2014). In addition, despite the importance of DHNs in abiotic stress tolerance mechanisms, little is known about their potential role in legume crop yields in stress-prone environments (Rémus-Borel et al., 2010; Araújo et al., 2015). Furthermore, the evolutionary history of the legume DHN family has never been phylogenetically investigated, with a previous review focusing only on overall sequence similarity analysis of LEA proteins in six legume species (Battaglia et al., 2013).

To date, there is only one report on the use of DHN genes isolated from legume species for the improvement of stress tolerance in crop legumes through transgenic approaches (Xie et al., 2012). However, rapid advances and the improvement in accuracy of bioinformatics approaches as well as re-annotated versions of legume genomes, including the recent genome sequencing of two *Arachis* species (Bertioli et al., 2016), can contribute to new insights into the molecular function and evolution of the DHN family in legumes.

In the present study, we characterized the DHN gene family in legumes regarding their phylogenetic classification, chromosomal localization, duplication events, molecular structure and spatial expression patterns. We then focused on the transcriptional behavior of a DHN gene from the wild peanut *Arachis duranensis* (*AdDHN1*) in response to abiotic and biotic stresses, shedding light on its role in stress responses. The overexpression of *AdDHN1* in *Arabidopsis* plants improved their tolerance to multiple abiotic stresses (freezing and drought) but increased their susceptibility to the biotrophic root-knot nematode (RKN) *Meloidogyne incognita*. This trade-off effect: improved resistances to abiotic stresses vs. increased susceptibility to biotic stress is an important finding that must be taken into consideration in research into the improvement of crop plants.

# MATERIALS AND METHODS

## Identification and Phylogenetic Analysis of Dehydrins in Fabaceae

The complete genomes of eight species (*Arachis duranensis*, *A. ipaënsis*, *Cicer arietinum*, *Cajanus cajan*, *Glycine max*, *Lotus japonicus*, *Medicago truncatula*, and *Phaseolus vulgaris*) belonging to the Fabaceae family were downloaded from their respective public databases (**Supplementary Table 1**). To identify

the putative DHN proteins in these eight proteomes, we used two independent approaches. First, the PFAM (Finn et al., 2016) dehydrin domain (PF00257) was used as a query for the program hmmsearch, from the HMMER3 suite (Mistry et al., 2013) against the predicted proteome of each species. In the second approach, the K-segment described by Malik et al. (2017) was used as input for the Find Individual Motif Occurrences (FIMO) software (Grant et al., 2011), applying a threshold $q$-value of $< 0.05$. The proteins found using both approaches were verified for their motifs, presence and organization. Only proteins with more than half of the total length of the PF00257 domain (168 aa) and with a canonical dehydrin motif organization were considered in the present study.

All the identified putative DHN protein sequences were aligned using the MAFFT software with automatic detection of the most appropriate alignment strategy (Katoh et al., 2002). The poorly aligned regions (more than 10% of gaps) were eliminated with the trimAl software (Capella-Gutiérrez et al., 2009). We performed the phylogenetic analysis of the selected protein sequences using RAxML software (Stamatakis, 2006), with automatic search of the fittest evolutionary model and parameters and a bootstrap search automatically stopped upon congruence.

## Synteny Analysis, Duplication Pattern, and Gene/Protein Structure of Dehydrins in Fabaceae

We retrieved the physical location of the putative DHN genes in the chromosomes of the eight Fabaceae species from the GFF-formatted file in their respective databases. The MCScanX software (Wang et al., 2012) was used to determine the syntenic relationship and duplication patterns between these species and Circa[1] to plot graphical representation of DHN genes location and their syntenic relationships.

The intron/exon organization of dehydrin genes was extracted from the GFF-formatted file of each genome and submitted to the GSDS software[2] for graphical representation. We predicted the consensus sequence of the DHN motifs of the proteins using the Multiple Expectation maximization for Motif Elicitation (MEME) (Bailey et al., 2009).

## Spatial Expression Pattern of Fabaceae Dehydrin Genes

The gene expression atlas of *A. duranensis*, *A. ipaënsis, G. max,* and *M. truncatula*[3] was used to analyze the spatial expression patterns of DHNs. Respective expression values, represented as FPKM, were retrieved in table format and the mean FPKM values for each of the following tissues were determined: dry seeds, roots, leaves and stems in each species. In addition, the RT-PCR analysis of transcript abundance of *G. max* DHN genes by Yamasaki et al. (2013) was also considered in our analysis.

[1]http://omgenomics.com

[2]http://gsds.cbi.pku.edu.cn/

[3]http://bar.utoronto.ca

## Analysis of Dehydrin Expression in Wild *Arachis* in Response to Stress

Our previously published transcriptome RNA-Seq data were exploited to determine the *in silico* expression profile of the two *A. duranensis* DHN genes in response to both abiotic (water deficit) and biotic (nematode inoculation) stresses (Mota et al., 2018; Vinson et al., 2018). Quantitative RT-PCR (qRT-PCR) analysis was also conducted, essentially as described by Morgante et al. (2013), to determine the relative expression of the *AdDHN1* and *AdDHN2* genes (**Supplementary Table 2**) in *A. duranensis* plants under dehydration and upon nematode infection, using the RNA isolated in Vinson et al. (2018) and Mota et al. (2018), respectively. ACT1 and UBI2 were used as reference genes for *Arachis* root samples subjected to dehydration; and 60S and GAPDH for samples inoculated with nematodes, in accordance with Morgante et al. (2011).

## *AdDHN1* Cloning

To identify the complete coding sequence of *AdDHN1*, the Aradu.IF4XP gene model[4] was aligned with the four best BLASTn hits of *A. duranensis* databases available on NCBI[5]. The consensus sequence (675 bp) was synthesized and cloned (Epoch Life Science Inc., TX, United States), under the control of the *Arabidopsis* actin 2 promoter (ACT-2) and the nopaline synthase (NOS) terminator, at the XhoI restriction site of pPZP-201BK-EGFP (Chu et al., 2014). This binary vector, hereafter called pPZP-AdDHN1, also contains two additional cassettes for the constitutive expression of the enhanced green fluorescent protein (eGFP) reporter gene and the hygromycin phosphotransferase (hpt) selection marker gene. The pPZP-AdDHN1 binary vector was then introduced into the disarmed *Agrobacterium tumefaciens* strain "GV3101" by standard electroporation protocol. Transformed colonies were selected by PCR using primer pairs flanking the eGFP or *AdDHN1* sequences (**Supplementary Table 2**).

## *Arabidopsis* Transformation

Wild type (WT) *Arabidopsis thaliana* ecotype Columbia (Col-0) was transformed with *A. tumefaciens* containing the pPZP-AdDHN1 vector using the floral dip method (Clough and Bent, 1998) and maintained in a growth chamber at 21°C, 60% relative humidity, 12 h photoperiod. Seeds originating from the dipped-plants were germinated on plates containing solid (0.8% w/v agar) half-strength Murashige and Skoog (MS) basal medium (Sigma-Aldrich, St. Louis City, United States) with sucrose (30 g/L) and hygromycin (15 mg/L). Hygromycin-resistant T0 plants were transferred to pots containing substrate (Carolina Soil®, CSC, Brazil) and maintained in the controlled growth chamber to produce the T1 generation. Transformants were then screened repeatedly for hygromycin resistance and grown on to obtain homozygous T3 generation *AdDHN1* overexpressing (OE) lines for further analysis.

[4]https://peanutbase.org/

[5]http://www.ncbi.nlm.nih.gov

## Stress Assays in Transgenic *Arabidopsis* Overexpressing (OE) *AdDHN1*

Seeds harvested from 13 OE lines at T3 generation were used to determine the effect of freeze-shock treatment on seedling growth. Two weeks old seedlings germinated on MS medium plates (12 individuals per OE line and WT per plate, three plates per line) were placed in a temperature-regulated freezer at −18°C for 1 h in the dark. After complete freezing of the MS medium, the seedlings on plates were returned to normal controlled growth conditions. The survival rate after the freezing treatment was assessed by recording the number of seedlings that regained turgidity and displayed a normal phenotype compared to that of non-treated seedlings, 3 days after the treatment. Seedlings that showed healthy growth in treated and non-treated plates were further used for quantification of total soluble sugars (Buysse and Merckx, 1993) and qRT-PCR analysis (see below).

The *Arabidopsis* OE lines with enhanced freezing tolerance were then selected for subsequent stress assays. For that, OE and WT plants, grown in the controlled growth chamber, were submitted to a gradual water deficit (dry-down) treatment and to infection with *M. incognita*. We analyzed the data using single factor ANOVA ($P < 0.05$).

For the dry-down assay, irrigation of 3 weeks old plants grown on substrate was interrupted for a group of 10 individuals per OE line and the WT (stressed group) during 8 days whilst the control group of individuals was kept under irrigated conditions, i.e., around 70% of field capacity (FC). During the assay, two SPAD chlorophyll meter readings (SCMR; SPAD-502, Konica Minolta Sensing, Japan) was recorded every 2 days from the same leaf of each individual. At the end of the water stress (8th day), we collected three leaf discs (0.4 $cm^2$) per individual to assess Relative Water Content (RWC) according to de Brito et al. (2011). The aerial and root plant fresh biomass were also determined at the end of the assay and the leaf area was estimated using the Rosette Tracker software (De Vylder et al., 2012).

For the *M. incognita* bioassay, roots from 4 weeks old plants grown on a 2:1 sand:substrate mixture (v:v) were inoculated with approximately 500 J2 infective larvae of *M. incognita*, essentially as described by Morgante et al. (2013). At 60 days after inoculation (DAI), roots were stained with acid fuchsin and the number of nematode females on roots (10 individuals per OE line and WT) assessed under a stereomicroscope.

## Analysis of Transcription Levels in *Arabidopsis* by qRT-PCR

We conducted qRT-PCR analysis in the T3 seedlings to study the expression levels of the *AdDHN1* transgene and other stress-responsive genes in *Arabidopsis* OE lines and WT plants. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), treated with DNAse and reverse transcribed as previously described (Morgante et al., 2013). qRT-PCR reactions were performed in three biological replicates on the StepOne Plus Real-Time PCR System (Applied Biosystems, Foster City, United States), as previously described (Vinson et al., 2018), using specific primers (**Supplementary Table 2**). The online real-time PCR Miner tool (Zhao and Fernald, 2005) was used to estimate the average cycle threshold (Cq) values. The relative quantification (RQ) of mRNA levels was normalized with AtACT2 and AtEF-1α reference genes (**Supplementary Table 2**) and determined for the *AdDHN1* transgene using the qGENE software[6] and for the stress-responsive *Arabidopsis* genes using the REST 2009 v. 2.0.13 software (Pfaffl et al., 2002).

# RESULTS

## Identification and Characterization of DHN Genes in Fabaceae

We searched for DHN proteins in the proteomes predicted from the whole genome sequences of eight Fabaceae species belonging to the Papilionoideae sub-family based on the presence of the conserved dehydrin PFAM domain PF00257. Initially, we identified 21 putative DHN proteins in the eight species (**Supplementary Table 1**). However, two protein sequences from *C. arietinum*, and their corresponding genes, were 100% identical and hence eliminated, yielding a total of 20 non-redundant putative DHN proteins. Based on their motif numbers and organization, we could classify all 20 putative DHNs identified into only three out of the five known subclasses: $Y_nSK_n$, $Y_nK_n$, and $SK_n$. Subsequently, we named each DHN according to the corresponding chromosomal position of the gene in each species (**Supplementary Table 1**).

## Phylogenetic Analyses, Gene Structure, and Protein Motifs of DHNs in Fabaceae Species

We performed a Maximium-likehood (ML) phylogenetic analysis using the 20 deduced DHN protein sequences identified in the eight Fabaceae species. The phylogenetic analysis found the JTT model as the most appropriate and the analysis converged after 599 bootstrap replicates. This analysis could separate, with high confidence values, the putative DHN proteins into two distinct groups, the proteins with a Y-segment (belonging to the $Y_nSK_n$ and $Y_nK_n$ subclasses) and the proteins without a Y-segment (belonging to the $SK_n$ subclass) (**Supplementary Figure 1**). The number of DHN proteins in the two groups was identical (ten in each group). With the exception of *C. cajan*, all the Fabaceae species possessed at least one $SK_n$ and one $Y_nSK_n$ DHN protein. *L. japonicus* was the only species with a predicted additional $Y_nK_n$-type DHN. This *L. japonicus* protein holds an outgroup position in the Y-segment DHN clade (**Supplementary Figure 1**).

We further analyzed the structural diversity of the 20 DHN genes through their exon/intron organization. Regardless of their phylogenetically-determined protein group, the legume DHN genes showed a very conserved 2-exon/1-intron organization, except for the *LjDHN3* gene from *L. japonicus* which was devoid of introns and was the only legume DHN gene that did not contain an S-segment (**Figure 1A**). Notably, the presence of a conserved intron in DHN genes seems to be associated with the presence of an S-segment in *Arabidopsis*, potato and rapeseed

[6]http://www.gene-quantification.de/download.html

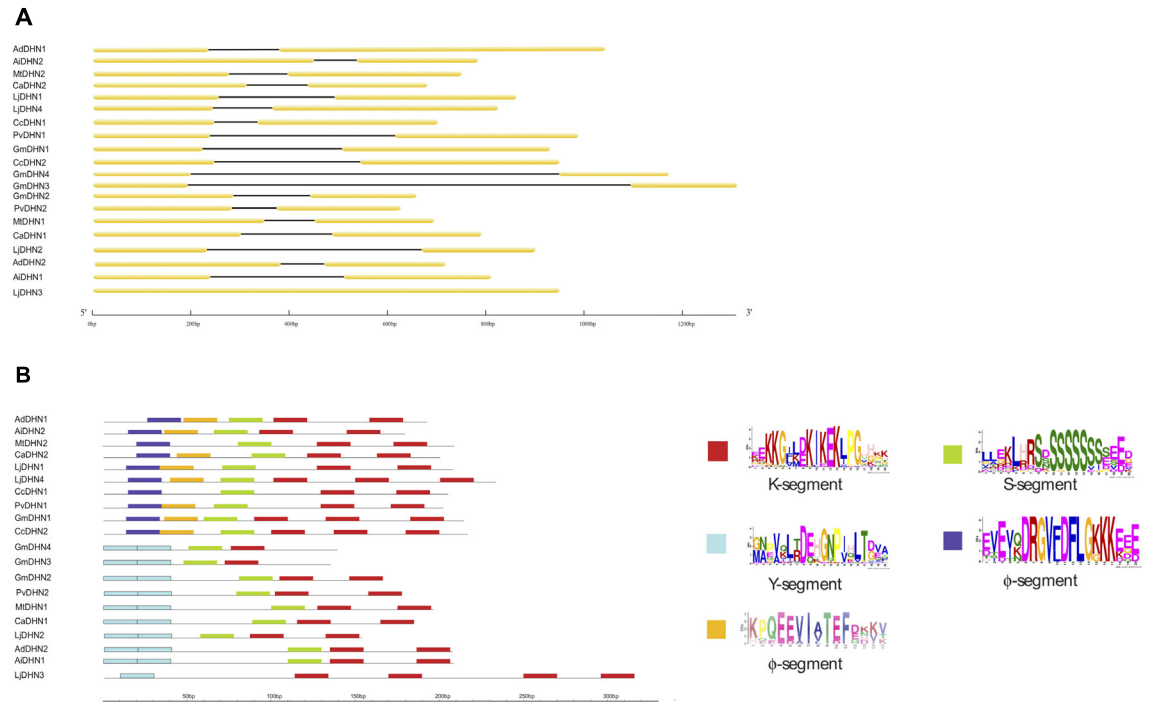



**FIGURE 1 |** Gene and protein organization of the 20 DHNs in eight Fabaceae species. **(A)** Exon/intron organization of the DHN genes; **(B)** Motif organization and conserved segments of the DHN proteins represented by different colors (left) and their corresponding motif logos (right). The scale length is represented in base pairs (bp) and amino acids.

(Jiménez-Bremont et al., 2013; Charfeddine et al., 2015; Liang et al., 2016). The protein structures of all 20 legume DHNs showed that, when present, the Y-segment generally occurred in two consecutive copies and the K-segment in one to four copies (**Figure 1B** and **Supplementary Figure 2**). In accordance with the phylogenetic analysis (**Supplementary Figure 1**), half of these proteins presented one or two Y-segments and are therefore classified as $Y_nSK_n$- and $Y_nK_n$-types. All DHNs identified in the eight legumes species showed a single S-segment, except LjDHN3 which lacked this segment (**Figure 1B** and **Supplementary Figure 2**). In this study, in addition to the location of the three conserved DHN motifs, we could identify two additional consensus sequences preceding the S-segment, represented by (DRGV[FL]DFLG) and (EE[VA]I[AV]TEF), near the N-terminal region of proteins belonging to the $SK_n$-type (**Figure 1B** and **Supplementary Figure 2**). These two DHN motifs have not yet been described, and hereafter are considered as Φ-segments, previously noticed as poorly conserved segments located between the K-, S-, and Y-segments (Graether and Boddington, 2014).

We studied the predicted subcellular localization of the 20 legume DHNs using the WoLF PSORT Prediction Software[7]. Interestingly, 19 out of 20 proteins studied were predicted to have a nuclear localization with only the $Y_nSK_n$-type CaDHN1 predicted to be cytoplasmic (**Supplementary Table 1**). This preferential nuclear localization of legume DHNs is not in accordance with previous studies showing that plant DHNs subcellular localization does not tend to be more nuclear than the cytoplasmic and are present in different parts of the cell (Abedini et al., 2017; Yu et al., 2018). Since the presence of phosphorylation sites is commonly associated with the subcellular localization of DHN proteins in the plant cell (Abedini et al., 2017), a further analysis revealed the presence of one phosphorylation motif and one nuclear localization site (NLS) in legume DHNs. The phosphorylation site (LXRXXS) was identified in 14 out of the 20 DHNs (**Supplementary Figure 2** and **Supplementary Table 1**), regardless of their subclasses, whilst the remaining six DHN proteins that did not present this motif belonged to the $Y_nSK_n$-type (AdDHN2, AiDHN1, GmDHN3, and GmDHN4), $SK_n$-type (LjDHN4) and the only example in legumes of the $Y_nK_n$-type (LjDHN3). Despite the absence of a phosphorylation motif, these proteins are still predicted to be in the nucleus. Interestingly, the presence of a NLS was observed only in DHN proteins belonging to the $Y_nSK_n$ subclass, with this specificity also observed in grapevine and barley (Yang et al., 2012; Abedini et al., 2017). The presence of these phosphorylation and NLS sites allow the phosphorylation of the S-segment and has previously been associated with the translocation of DHNs from the cytoplasm to the nucleus (Yang et al., 2012; Abedini et al., 2017; Yu et al., 2018). However, in legumes DHN traffic in the cell could also be associated with other motifs, as LjDHN3, which lacks both phosphorylation motifs, is predicted to be located in the nucleus, whereas CaDHN1, a $Y_nSK_n$-type which contains both phosphorylation motifs, is predicted to have a cytoplasmic location.

[7]https://wolfpsort.hgc.jp

## Chromosomal Location and Syntenic Relationships of DHN Genes in Fabaceae Species

The chromosomal location of the DHN genes varied according to the legume species and the number of representatives per species. Interestingly, with the exception of *C. cajan*, which does not possess a $Y_nSK_n$-type DHN, all the species possessed one $SK_n$-type and one $Y_nSK_n$-type in physical proximity on the same chromosome (**Supplementary Figure 3**). The two *C. cajan* DHN genes belonging to the $SK_n$ type were distributed on distinct chromosomes (Cc06 and Cc07). *G. max* and *L. japonicus* possess more DHN genes (four representatives each) and, in addition to the co-localization of the conserved $SK_n$-type and $Y_nSK_n$-type on a same chromosome (Gm04 for *G. max* and Lj01 for *L. japonicus*) (**Supplementary Figure 3**), the two extra representatives were on different chromosomes. For *G. max*, the two extra $Y_nSK_n$-type copies were present on chromosomes Gm12 and Gm13. For *L. japonicus*, the extra $SK_n$-type DHN was on chromosome Lj05 while the $Y_nK_n$-type, so far specific to this species, was on chromosome Lj02.

From a McScanX analysis at the whole genome level for the eight species, we further focused on the duplication and syntenic relationships of the 20 legume DHN genes (**Supplementary Figure 3**). The association of one $Y_nSK_n$- and one $SK_n$-type representative located on the same chromosome, observed in seven out of eight species (the exception is *C. cajan*), was part of a larger conserved synteny block for six species (*A. duranensis*, *A. ipaënsis*, *G. max, L. japonicus*, *M. truncatula,* and *P. vulgaris*). Surprisingly, the two *C. arietinum* representative genes, although also on the same chromosome, did not show conserved synteny with the conserved blocks from the other six species. Likewise, the extra representatives for *G. max* and *L. japonicus*, distributed on different chromosomes, did not exhibit synteny among the legume species (**Supplementary Figure 3**).

Previous reports described that the DHN gene family had undergone duplication events in plants, resulting mostly from whole-genome (WGD) and tandem duplications (Yang et al., 2012; Liang et al., 2016). However, this does not seem to be the case for the DHN genes from Fabaceae species, where only one pair of DHN genes from *G. max* (*GmDHN3*/*GmDHN4*) was considered as duplicated through a WGD/tandem event. These results indicate that except for *G. max*, duplication events of the DHN gene families in Fabaceae probably occurred before speciation, which is also consistent with our phylogenetic analysis (**Supplementary Figure 1**).

## DHN Gene Spatial Expression in Fabaceae

We analyzed the possible relationship between the presence/absence and organization of the DHN conserved segments and their spatial expression pattern in plant tissues using the expression atlas[8] publicly available for four (*A. duranensis*, *A. ipaënsis, G. max,* and *M. truncatula*) of the Fabaceae species and a transcript abundance analysis of *G. max* (Yamasaki et al., 2013). Within a DHN type, the expression across the different species tended to be similar (**Supplementary Table 3**). Three of the $SK_n$-type genes (*AdDHN1*, *AiDHN2,* and *GmDHN1*) were predicted to be ubiquitous with high FPKM expression values (from 4.26 to 449.51) in dry seeds, roots, leaves, and stems. This expression behavior was confirmed for the *G. max* gene (*GmDHN1*) by Yamasaki et al. (2013) through RT-PCR analysis of seeds, leaves, and stems. The expression of the *MtDHN2* gene from *M. truncatula* was not detected in any of the experiments of Benedito et al. (2008) included in the expression atlas (**Supplementary Table 3**). Contrastingly, half of the legume DHN genes that belong to the $Y_nSK_n$ type (*AdDHN2*, *AiDHN1,* and *MtDHN1*) exhibited a more tissue-specific expression, restricted mainly to seeds (FPKM > 16,000 for *M. truncatula*), and a very low gene expression in other tissues. Moreover, expression of the *G. max* $Y_nSK_n$ genes (*GmDHN2* and *GmDHN2*) was only detected in seeds by Yamasaki et al. (2013) (**Supplementary Table 3**).

## *A. duranensis* DHN Genes Regulation Under Abiotic and Biotic Stresses

We studied in more detail the regulation of two *A. duranensis* genes (*AdDHN1* and *AdDHN2*) in response to abiotic and biotic stresses. This wild *Arachis* species has been used by our group in functional genomics studies due to its contrasting responses to abiotic and biotic stresses: tolerant to water deficit but somewhat susceptible to *Meloidogyne arenaria* infection (Proite et al., 2008; Leal-Bertioli et al., 2012; Guimaraes et al., 2017). The *in silico* expression profile of *AdDHN1* and *AdDHN2* was determined using our previous RNA-seq data including *A. duranensis* roots submitted to a dehydration treatment (Vinson et al., 2018) and roots challenged with the RKN *M. arenaria* (Mota et al., 2018). We then validated the *in silico* expression analysis based on RNA-seq using qRT-PCR.

The RNA-seq expression profiling of *AdDHN1* transcripts in *A. duranensis* roots showed opposite behaviors for expression of this gene under the two stresses analyzed, with upregulation in response to dehydration and downregulation in response to nematode infection (**Supplementary Figure 4**). This opposite expression behavior was further validated by qRT-PCR using specific primers (**Supplementary Table 2**), corroborating the predicted *in silico* analysis (**Supplementary Figure 4**). Contrastingly, no expression of *AdDHN2* could be detected either *in silico* (no mapped reads) or by qRT-PCR (no amplified samples) under the two stresses evaluated. These results are in accordance with the expression atlas (**Supplementary Table 3**), which did not show any basal expression of *AdDHN2* in *A. duranensis* roots. Therefore, based on the expression profiling of both *Arachis* DHN genes, only the *AdDHN1* gene was selected for a deeper characterization at functional levels as an interesting candidate involved in opposite responses to abiotic and biotic stresses.

## Analysis of *Arabidopsis* Plants Overexpressing *AdDHN1*

To further investigate the involvement of *AdDHN1* in the response to abiotic and biotic stresses through its overexpression

[8]http://bar.utoronto.ca

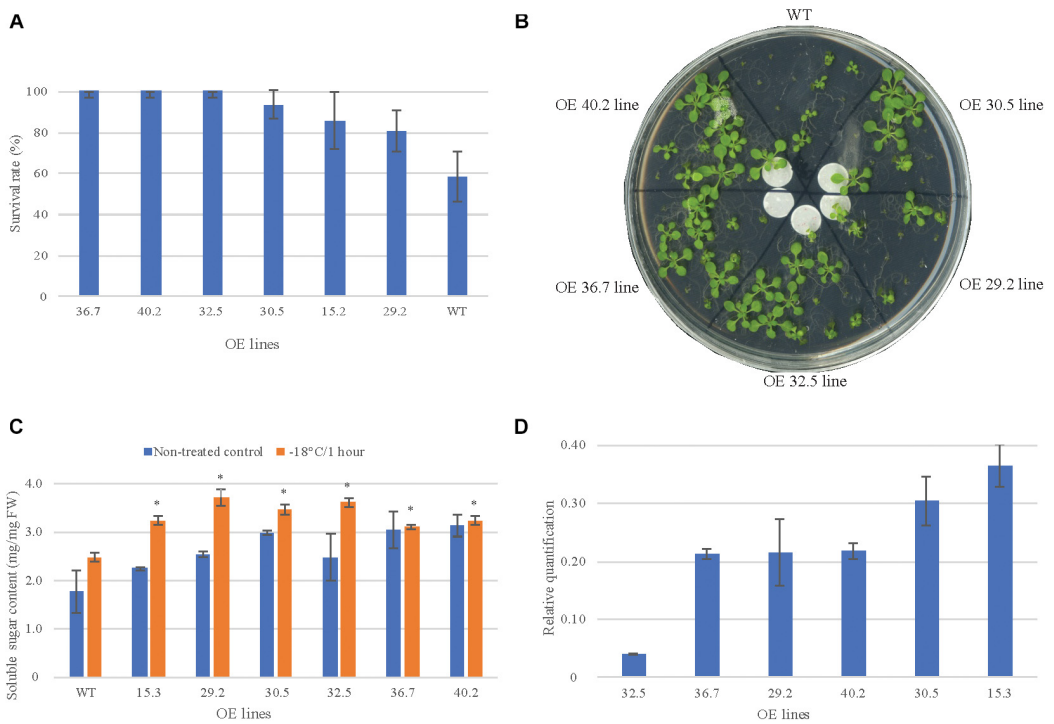



**FIGURE 2 |** Morphology and freezing tolerance in 2 weeks old *Arabidopsis* wild-type (WT) plants and in six *AdDHN1* overexpressing (OE) lines. **(A)** *Arabidopsis* seedlings and **(B)** their survival rates 3 days after the freezing-shock treatment (−18°C for 1 h); **(C)** Total sugar content in leaves from treated and non-treated control seedlings. Values are means ± SD of three independent replicates and expressed as μg/mg of fresh weight (FW). Significant differences ($P < 0.05$) between WT and OE lines are marked with an asterisk. **(D)** The relative expression of *AdDHN1* transgene quantified by qRT-PCR, using the ACT2 gene from *Arabidopsis* as the reference gene. Error bars are the standard errors of the means from three samples of 10 plants.

in transgenic plants, a consensus coding sequence was determined. The alignment of five *A. duranensis* sequences showed high nucleotide conservation, except for the gene model Aradu.IF4XP[9] which present a gap of 128 bp (**Supplementary Figure 5**). The consensus of *AdDHN1* coding sequence (675 bp) was then cloned and used to produce transgenic *Arabidopsis* plants.

## Freezing Treatment

The overexpressing (OE) lines at T3 generation were selected based on their response to freezing-shock treatment (survival rate and sugar content) and on the abundance of *AdDHN1* transcripts. We observed an enhanced freezing tolerance in six of the 13 OE lines submitted to freezing treatment (−18°C for 1 h) which showed a greater survival rate than the WT (**Figures 2A,B**). The plantlets from these six lines (15.2; 29.2; 30.5; 32.5; 36.7; and 40.2) displayed a normal regrowth after the treatment and appeared to be more tolerant to the freezing injury than WT plants, which were not able to resume growth. Moreover, all of these six OE lines accumulated more soluble sugars in leaves than WT plants under normal growth conditions and exhibited a significant increase in sugar content in response to the freezing treatment (**Figure 2C**). Accumulation of sugars under adverse environmental conditions, including freezing, could contribute to maintain cell turgor and to protect membranes and proteins against stress damage (Krasensky and Jonak, 2012). Accordingly, the overexpression of *AdDHN1* in *Arabidopsis* seemed to improve freezing tolerance concomitant with a higher concentration of solutes, under both normal and stress conditions, as observed in previous studies (Mattana et al., 2005). The qRT-PCR analysis subsequently confirmed the overexpression of the transgene *AdDHN1* in these six OE lines at different expression levels (**Figure 2D**). Overall, the *AdDHN1* transcript abundance did not correlate with the enhanced freezing tolerance or the accumulation of soluble sugars, in particular for the OE 32.5 line, indicating a possible post-transcriptional regulatory mechanism of the transgene. These findings are in agreement with the general presumption that there is no clear relationship between the transcript levels of overexpressed transgenes and the phenotypic effects observed in transgenic lines, which can vary from a strong positive to no significant correlation (Kerr et al., 2018; Ko et al., 2018). Based on these analyses, the six freezing-tolerant OE lines were selected for further analysis.

## Dry-Down Assay

The water deprivation assays indicated that the six best freezing tolerant OE lines also exhibited enhanced drought tolerance, with less morphological damage compared to WT (**Figure 3A**). From the 5th day of the dry-down assay, WT plants showed

[9]https://peanutbase.org/

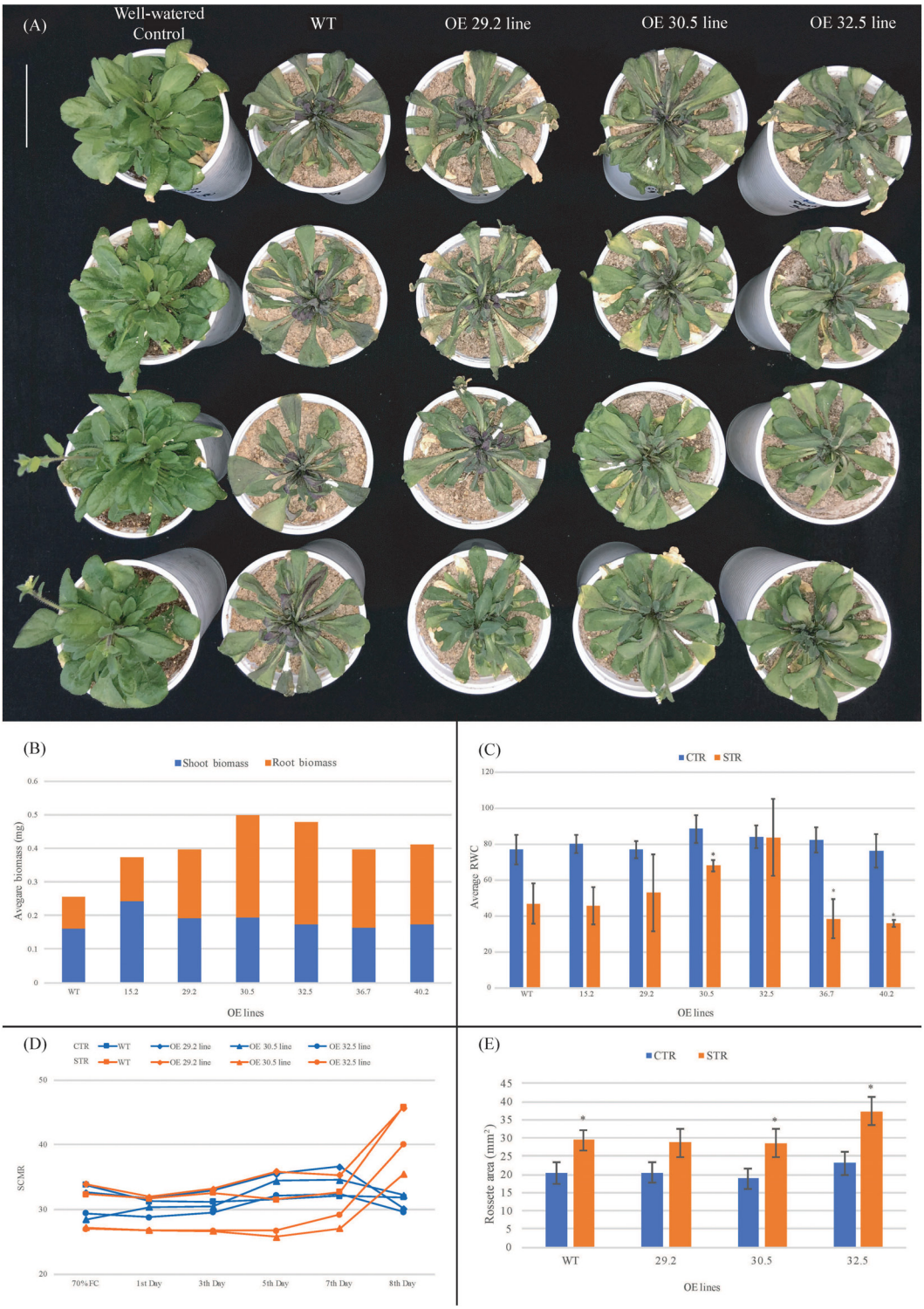



**FIGURE 3 |** Performance of 3 weeks old *Arabidopsis* plants from WT and *AdDHN1* OE lines submitted to a dry-down treatment for 8 days (stressed plants; STR) and the corresponding irrigated control (control plants; CTR). **(A)** Phenotype of the aerial part of CTR plants (first column) and STR plants (2nd–4th columns) from WT and the three OE lines (29.2, 30.5, and 32.5); **(B)** shoot and root biomass (miligrams of fresh weight) analysis of WT and OE lines; **(C)** Percentage of relative water content (RWC) in leaves from WT and OE lines. Values are means ± SD of 10 individuals and significant ($P \leq 0.05$) differences between CTR and STR plants are marked with an asterisk; **(D)** SCMR (SPAD chlorophyll meter reads) of WT and OE lines. The dry-down assay initiated with 70% FC, decreasing over time in STR plants from 65% FC (1st Day), 50% FC (3nd Day), 40% FC (5rd Day), 30% FC (7th Day) to 20% FC (8th Day); **(E)** Average of rosette area ($mm^2$) of WT and OE plants at 8th day of dry-down treatment. Values are means ± SD of four individuals and significant ($P \leq 0.05$) differences between CTR and STR plants are marked with an asterisk.

progressive symptoms of water deficiency (leaf wilt), followed by growth retardation that resulted in 38.9% lower biomass at the 8th day, compared to the average of OE lines (**Figures 3A,B**). This lower biomass was particularly noticeable in roots, with around 55.4% less biomass in the WT compared to the average of OEs. The RWC was also analyzed at the end of the dry-down assay, with no significant difference observed between leaf RWC values of OE and WT control plants, with values falling within the expected RWC range of 76–88% (**Figure 3C**). However, OE and WT stressed plants exhibited differences in the reduction in RWC compared to the corresponding control plants that occurred over the course of the experiment. The reduction of RWC values observed in leaves of three OE lines (29.2; 30.5; and 32.5) in response to water deficit (average of 18%) was significantly smaller than in the other lines (15.2; 36.7; and 40.2) which were similar to the WT plants (around 39%). The OE 32.5 line showed a distinct behavior with a RWC value of 84% under drought conditions, which was almost the same value found for plants under irrigated conditions and a good indicator of drought tolerance due to the capacity of these plants to maintain high leaf water status. These results suggest that leaves of three OE lines (29.2; 30.5; and 32.5) can maintain their levels of internal water higher than WT for eight consecutive days without irrigation. For a more detailed investigation, a new dry-down assay (biological repetition) was carried out with only these three OE lines, under the same experimental conditions. SCMRs recordings confirmed the responsiveness of OE and WT plants to drought imposition that initiated with 70% FC and decreased to 20% FC at the 8th day of treatment (**Figure 3D**). While the control (irrigated) plants maintained a mean SCMR of 32 over the treatment, stressed (non-irrigated) plants presented a gradual increase from the 3th day, reaching values of approximately 41 five days after. In addition, two OE lines (30.5 and 32.5) displayed reduced SCMR values up to the 6th day of treatment, compared to WT plants, indicating a possible adaptive response to preserve photosynthetic efficiency. Measurements of rosette area showed a significant increase in the total rosette area of WT and OE lines over the course of the stress treatment (**Figure 3E**). The larger rosette surface area observed in the 32.5 line suggests that this line was better able to maintain growth under stress compared to the WT and in 29.2 and 30.5 OE lines.

### Nematode Challenge

To examine the potential effect of *AdDHN1* overexpression on the nematode infection process in *Arabidopsis*, 4 weeks old OE and WT plants were challenged with 1,000 *M. incognita* juveniles (J2) and the level of infection assessed by the number of nematode females in each root system at 60 DAI. Overall, the three selected OE lines were severely affected by *M. incognita*, with a significant higher number of females found in the roots of all three OE lines, when compared to the WT plants, indicating an increase in the susceptibility due to *AdDHN1* overexpression (**Figure 4**). The increase in the *M. incognita* infection rates in the OE lines ranged from 60% in 29.2 line to around 41% in 30.5 and 32.5 lines. No visible difference was observed in the morphology and biomass of roots from OE lines and WT, which ranged from 298.5 to 188.8 mg of roots per plant (**Figure 4**).

### qRT-PCR Analysis of Stress-Responsive Genes

To determine whether the overexpression of *AdDHN1* could interfere with *Arabidopsis* hormone signaling involved in stress responses, we analyzed the expression of a subset of nine marker genes associated with ABA and JA pathways (Seo et al., 2009; Yang et al., 2011; Sasaki-Sekimoto et al., 2013; Naznin et al., 2014; Guo et al., 2015; Singh and Laxmi, 2015; Zhao et al., 2018; **Supplementary Table 2**). The relative expression of these marker genes was compared between the three OE lines (29.2, 30.5, and 32.5) and the WT plants. The *AdDHN1* overexpression negatively regulated the expression of two marker genes (*ERD1* and *RD29A*) from the ABA-independent pathway which was both downregulated in the three OE lines (**Figure 5A**). Conversely, transcripts of the *RD29B* and *RD22* genes from the ABA-dependent pathway seemed to be activated by *AdDHN1* overexpression, as *RD29B* could be detected in two OE lines but not in the WT plants (**Supplementary Figure 6**) whereas *RD22* was positively regulated in the three OE lines (**Figure 5B**). Regardless of the apparent interference with ABA-independent signaling, the *AdDHN1* overexpression induced the activation of downstream ABA-responsive genes that promote improvement of drought tolerance, including the orthologous of *AdDHN1* gene in *Arabidopsis* (*AtDHN*) (**Figures 5**, **6**).

Concurrently, *AdDHN1* overexpression promoted the downregulation of three genes involved in the JA pathway (*JAZ1*, *MYC2,* and *ORA59*; **Figure 5A**) and the upregulation of *PDF1.2*, a marker gene for JA/ET-mediated responses, which expression was exclusive in OE plants (**Supplementary Figure 6**). These alterations in the JA pathway could be owed to the blockage of *MYC2* due to *AdDHN1* overexpression, as previously suggested by Hanin et al. (2011) in *Arabidopsis*.

## DISCUSSION

### Fabaceae DHNs Are Mainly of Two Types and Were Already Present in Their Common Ancestor

Dehydrins (DHNs) are LEA proteins known to act in multiple developmental processes and in response to various stresses, and despite their role in abiotic stress tolerance, they remain insufficiently studied in legumes (Fabaceae). The structure in motifs and presence/absence pattern for the whole LEA gene family has previously been described for six legume species (Battaglia et al., 2013). However, this study neither analyzed the evolutionary relationships between the DHN sequences nor their distribution throughout whole genomes in relation to conserved synteny. Here, by using phylogenomics methods and taking advantage of the re-annotated versions of these six genomes, as well as the recent genome sequencing of two *Arachis* species, we provide new insights into molecular evolution of the DHN family in legumes. Based on the presence and conservation of the PFAM domain of DHNs, containing the signature K-segment, we identified 20 likely functional DHN genes in eight species of Fabaceae.

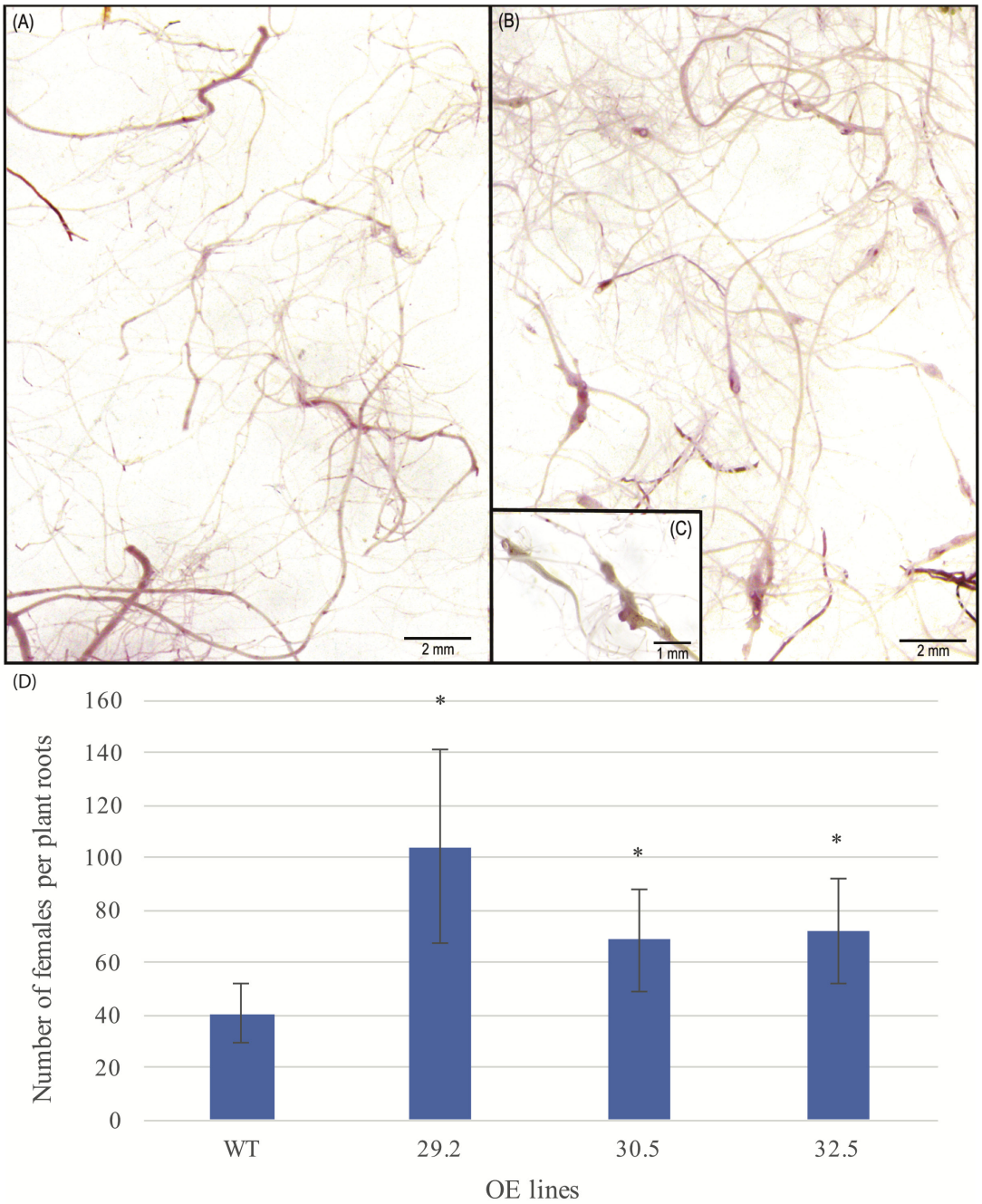



**FIGURE 4 |** Roots of *Arabidopsis* plants stained by acid fuchsin from **(A)** WT and **(B)** 29.2 OE line, at 60 days after inoculation (DAI) with *Meloidogyne incognita*. **(C)** Zoom in galls observed in 29.2 OE line. **(D)** Average number of females per plant roots of WT and OE lines inoculated with *M. incognita*. Values are means ± SD of 10 individuals and significant ($P \leq 0.05$) differences between WT and OE lines are marked with an asterisk.

Most DHN (19 out of 20) genes studied here were distributed between two sub-types and all the species, except *C. cajan*, possessed at least one representative of each sub-type, suggesting that the common ancestor of all these species already possessed at least one DHN of each type.

The presence of one type of each DHN in the common ancestor was confirmed by current phylogenetic analysis, which showed a clear separation of DHNs into two distinct groups, those containing the Y-segment (nine $Y_nSK_n$- and one $Y_nK_n$-type), and those without the Y-segment (10 $SK_n$-type). However, the McScanX duplication and synteny analysis suggests that the $Y_nSK_n$- and $SK_n$-type DHNs in these legumes do not originate from the WGD event described in papilionoid (PWGD) (Cannon et al., 2015). Indeed, the two DHN types tend to

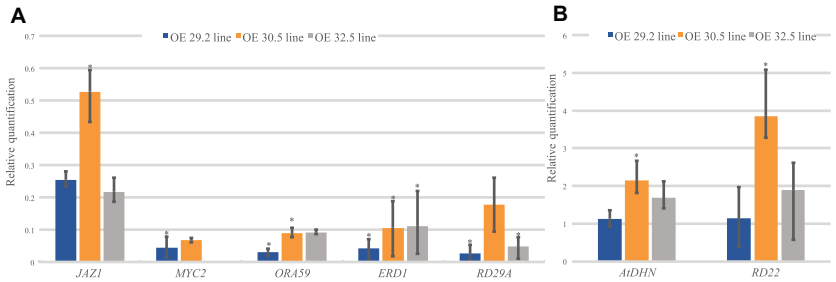



**FIGURE 5 |** Relative quantification of mRNA levels of seven stress-responsive *Arabidopsis* genes (*JAZ1; MYC2; ORA59; ERD1; RD29A; AtDHN;* and *RD22)* in the three OE lines (29.2, 30.5, and 32.5) relative to the WT plants. **(A)** Downregulated and **(B)** upregulated genes. Values are means ± SD of three biological replicates and significant ($P \leq 0.05$) differences between WT and OE lines are marked with an asterisk.

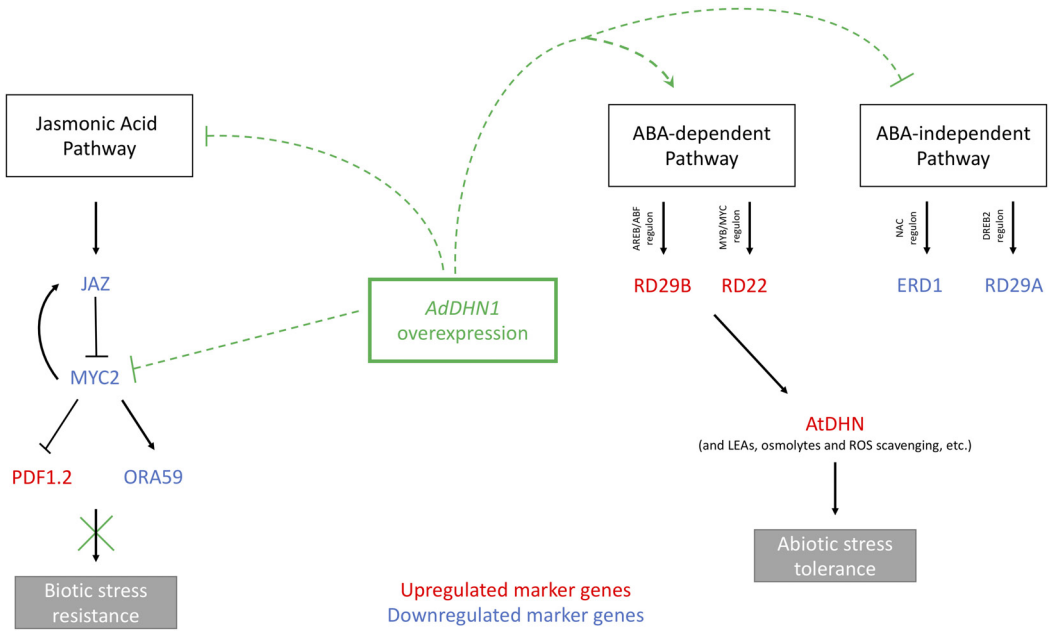



**FIGURE 6 |** Schematic representation of the ABA and JA pathways in *Arabidopsis*. Black lines represent the expected role of nine marker genes on ABA and JA pathways in WT plants. Green dotted lines represent the hypothetical interference of the *AdDHN1* overexpression in *Arabidopsis* OE lines in these pathways, in accordance with Hanin et al. (2011). The red color indicates marker genes that are upregulated in OE lines and blue color those downregulated.

be co-localized in close proximity on the same chromosome and were never connected by a WGD/tandem duplication relationship in our analysis. This co-localization on the same chromosome of one $Y_nSK_n$ and one $SK_n$-type was probably derived from a common ancestral proximal duplication.

The extra copy of a $Y_nSK_n$-type in *G. max*, in contrast, was clearly due to a WGD event in this species. It is known that *G. max* underwent a WGD that occurred after the differentiation of the other *Glycine* species (Schmutz et al., 2010).

Overall, and despite the PWGD event, the number of DHNs in each Fabaceae species is relatively low when compared to other vascular plant species studied (e.g., six DHN genes per species on average in poplar, *Arabidopsis*, barley and rice (Wang et al., 2007; Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008; Tommasini et al., 2008). This reduced number is likely due to multiple losses of DHN genes over the course of Papilionoideae evolution, through events called Legume Lost Genes (LLGs), recently proposed by Gu et al. (2016).

In Fabaceae, we found full length DHNs for only three of the five known types: nine $Y_nSK_n$ as well as one $Y_nK_n$, both described as more frequent in monocot species (Abedini et al., 2017), and 10 $SK_n$-type. Previous studies have described the presence of additional segments, other than the three commonly associated to the DHN genes (K-, S-, and Y-segments), called Φ-segments, preceding the S-segment at the N-terminal region and with no sequence conservation between plant species (Graether and Boddington, 2014; Abedini et al., 2017). These Φ-segments are characterized by an enrichment of glycine and threonine,

and, more rarely, tryptophan, cysteine and phenylalanine (Abedini et al., 2017; Malik et al., 2017), however, their function is still not clear. Here, we identified two novel Φ-segments (DRGV[FL]DFLG) and (EE[VA]I[AV]TEF), in the legume DHN genes from the $SK_n$ subclass, preceding the S-segment at the N-terminal end of the proteins. These motifs have not yet been described in plant DHNs.

## DHN Segment Organization Does Not Correlate With Their Subcellular Localization but With Their Spatial Expression Pattern

The analyses associating the presence and organization of the conserved segments observed here (K-, S-, and Y-) demonstrated that there is no correlation with the subcellular localization of the DHN genes. However, previously described phosphorylation motifs, may be involved in their location or translocation to the nucleus. The motif LXRXXS can be phosphorylated by a kinase and trigger translocation of the protein from the cytosol to the nucleus. This conserved motif was observed in most of the DHN protein sequences of Fabaceae, explaining the preferential nuclear location of these proteins (Abedini et al., 2017).

While the conserved segments have no influence on DHN subcellular location, it seems to be associated with their expression in different plant tissues. Expression of genes encoding $Y_nSK_n$ proteins was observed preferentially in seeds of the Fabaceae species, agreeing with previous studies in grapevine and *Arabidopsis* (Hundertmark and Hincha, 2008; Yang et al., 2012). Conversely, $SK_n$-type DHN genes display a ubiquitous pattern, as they are expressed in both the vegetative tissues (roots, leaves, and stems) and the dry seeds. This difference in the spatial distribution of expression is correlated with the phylogenetic separation of the genes, and could be associated with distinct roles of the subclasses of DHN in the plant, as suggested by Graether and Boddington (2014). The response of the *AdDHN1* gene to drought imposition and its high expression levels in all the analyzed tissues, led to our choosing to clone this gene, and analyze its function further using the model plant *Arabidopsis*.

## Overexpression of *AdDHN1* in Transgenic *Arabidopsis* Plants Led to Better Resistance to Abiotic Stresses but Increased Susceptibility to Nematodes

Several studies have shown the positive effect of DHN overexpression on plant tolerance to different abiotic stress conditions, mainly low temperatures, drought and salinity (Graether and Boddington, 2014). Most of these studies have concentrated on a few DHN genes isolated from model plants or major crops, with few focusing on native species that are well adapted to adverse environmental conditions.

Here, we isolated and studied the DHN gene *AdDHN1* from *A. duranensis*, a wild species native to low rainfall regions in South America that has evolved molecular and morphological adaptation mechanisms that aid survival in adverse and water-limited environments (Leal-Bertioli et al., 2012). We cloned the *AdDHN1* coding sequence under the control of a constitutive promoter to generate transgenic *Arabidopsis* lines expressing this candidate gene, which is potentially involved in both abiotic and biotic stresses. The heterologous overexpression of *AdDHN1* conferred different degrees of tolerance to a freezing-shock treatment in seedlings and to a gradual water deficit assay in substrate-grown plants. These results were supported by plant phenotypes and physiological indices, such as sugar and RWCs, chlorophyll meter readings, biomass and rosette area. This suggests that *AdDHN1* confers a similar protective role in both abiotic stresses and corroborates previous studies showing that the overexpression of $SK_n$-type DHNs enhanced the tolerance to low temperatures and drought in transgenic plants (Liu Y. et al., 2015; Bao et al., 2017). Similarly, the overexpression of a $YK_n$-type DHN isolated from *Saussurea involucrate*, a wild species that also grows in adverse environmental conditions, also conferred enhanced tolerance to both cold and drought in transgenic tobacco plants (Guo et al., 2017).

Although the role of DHNs in the response to a large range of abiotic stresses is well reported, little is known about their involvement in biotic stress responses. Some studies have reported the induction of DHNs in response to filamentous pathogen attack, such as *Erysiphe necator* in grapevine (Yang et al., 2012) or in combination with water deficit as with *Phytophthora cinnamomi* in oak (Turco et al., 2004), suggesting a putative role of DHNs in modulating pathogen defense responses (Hanin et al., 2011; Rosales et al., 2014). However, to date, the potential effect of DHNs in other interactions with different pathogens, including Metazoa such as nematodes, has not been reported.

Here, in contrast to abiotic stress, the overexpression of *AdDHN1* enhanced the susceptibility of transgenic *Arabidopsis* lines to the RKN *M. incognita*, a biotrophic plant pathogen with a wide host range and a very sophisticated strategy of host colonization. The present study demonstrates for the first time the increased susceptibility to a pathogen attack due to the overexpression of a DHN gene in transgenic plants. However, the mechanisms that are involved in these defense responses, as well as the tradeoffs affecting the plant normal metabolism are still unclear.

## Overexpression of *AdDHN1* in *Arabidopsis* Has Effects on ABA and JA Plant Defense Pathways

It is well known that ABA- and JA-signaling pathways are involved in response to both biotic and abiotic stresses, and, although JA-dependent signaling is generally effective against necrotrophic pathogens (Pieterse et al., 2012), it can also be activated by wounding and some biotrophic pests (Robert-Seilaniantz et al., 2011). However, these hormone signaling pathways often interact in diverse manners in response to a stimulus, a phenomenon referred to as "crosstalk" (Pieterse et al., 2012; Van der Does et al., 2013; Singh and Laxmi, 2015), leading to different, postponed or subdued defense responses.

Many studies support the idea that DHNs are ABA-regulated proteins, as implied by the presence of motifs linked to ABA-dependent and ABA-independent pathways in the promoter of several DHN genes, in particular those containing the S-segment (Zolotarov and Strömvik, 2015; Tiwari et al., 2018; Yu et al., 2018). Here, we showed that the overexpression of *AdDHN1* in transgenic *Arabidopsis* interfered with the expression of a subset of ABA-marker genes in comparison compared to WT plants. For instance, we observed in OE lines a contrasting regulation in the two major ABA signaling pathways, illustrated by the downregulation of *ERD1* and *RD29A*, which are part of the NAC and DREB2 regulons, respectively, in the ABA-independent signal transduction pathway, and the upregulation of *RD22* and *RD29B*, which are part of the MYB/MYC and AREB/ABF regulons, respectively, in the ABA-dependent pathway (**Figure 6**). These results suggest that the overexpression of *AdDHN1* acts in the ABA-dependent pathway to promote freezing and drought tolerance in transgenic *Arabidopsis* plants, despite the unnoted function of the ABA-independent pathway. This is supported by the induction in OE lines of the endogenous *Arabidopsis* DHN gene (*AtDHN*), which is the orthologous of *AdDHN1*. $SK_n$-type DHNs, such as *AtDHN* and *AdDHN1*, are part of the overall macromolecular protection mechanisms activated by the complex transcriptional cascade downstream of the DREB2A stress-regulatory system to avoid water loss during drought stress (Yu et al., 2018). Moreover, Tiwari et al. (2018) recently proposed that, apart from their role as a protective protein, DHNs interact with other proteins to act as a positive regulator involved in ABA-mediated drought stress signaling.

In addition to ABA, other hormone signaling pathways normally associated with biotic and wounding stresses, such as the JA pathway, can also be activated by dehydration stress due to cellular damage (Hanin et al., 2011). In our work, we hypothesized that the overexpression of *AdDHN1* suppresses the expression of *MYC2*, a central regulator in JA synthesis, causing in turn *ORA59* downregulation and enabling the expression of *PDF1.2* (**Figure 6**). We also found that *JAZ1*, a negative regulator of JA signaling, is also downregulated in the OE lines, possibly through a negative regulatory loop (Chini et al., 2007), due to *MYC2* repression (**Figure 6**). The overexpression of DHNs causing *MYC2* suppression in transgenic plants and affecting their responses to pathogen attacks has previously been suggested by Hanin et al. (2011). These findings indicate that the interference of *AdDHN1* overexpression in the JA pathway, compromising the modulation of JA- and wound-responsive genes (Hanin et al., 2011), contribute to an increase in nematode susceptibility in the OE lines compared to WT.

Indeed, our recent transcriptome studies of wild *Arachis* showed that both JA and ABA pathways are, similarly, activated in response to dehydration in the drought-tolerant *A. duranensis* (Vinson et al., 2018) whereas the JA pathway is the preferential route triggered in the resistance response to the RKN *M. arenaria* in *A. stenosperma* (Guimaraes et al., 2015; Mota et al., 2018). However, the detailed mechanisms by which the overexpression of *AdDHN1* gene from *A. duranensis* interferes in the responses to dehydration and nematode attack in transgenic *Arabidopsis* and the crosstalk between both stresses remains unclear, although its role in the modulation of ABA and JA signaling pathways are suggested.

Overall, our findings reveal an important tradeoff between biotic and abiotic defense responses following DHN gene overexpression. Although enhanced resistance to abiotic stresses tends to confirm and reinforce the protective role of DHN proteins in abiotic stress tolerance, this comes with a cost for the plant. Indeed, susceptibility to the infection by RKNs is increased, suggesting a negative regulation in disease-resistance responses. Tradeoffs between different processes in plants, including stress responses, may occur due to resource restriction and have an important effect on plant productivity and fitness, with hormone crosstalk having a major role in regulating this balance and prioritizing plant responses. A better understanding of the molecular aspects involved in these defenses signaling interactions is therefore vital to optimize plant engineering programs aiming at improving resistance to abiotic and biotic stresses and their tradeoffs in plants.

## AUTHOR CONTRIBUTIONS

AM contributed to the design of the work, data analysis and interpretation, and drafting the manuscript. TO and MC performed the data collection and analysis. CV and TW contributed to the data collection and analysis and critical revision of the manuscript. AA and MG-d-S contributed to the data interpretation and critical revision of the manuscript. ED contributed to the design of the work, data analysis and interpretation, and critical revision of the manuscript. PG contributed to the conception and design of the work, data analysis, and critical revision of the manuscript. AB contributed to the conception and design of the work, data analysis, interpretation, drafting, and critical revision of the manuscript. All the authors have approved the final version of the manuscript to be published.

## FUNDING

This work was financially supported by grants from EMBRAPA; CNPq-Brazilian National Council for Scientific and Technological Development; Special Visitor Researcher program (PVE grant); INCT PlantStress (project number 465480/2014-4); CAPES- Coordination for the Improvement of Higher Education Personnel and FAP-DF- Distrito Federal Research Foundation. Each of the funding bodies granted the funds based on a research proposal. They had no influence over the experimental design, data analysis or interpretation, or writing the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00497/full#supplementary-material

## REFERENCES


Abedini, R., Ghanegolmohammadi, F., and Pishkamrad, R. (2017). Plant dehydrins: shedding light on structure and expression patterns of dehydrin gene family in barley. *J. Plant Res.* 130, 747–763. doi: 10.1007/s10265-017-0941-5

Araújo, S. S., Beebe, S., Crespi, M., Delbreil, B., González, E. M., Gruber, V., et al. (2015). Abiotic stress responses in legumes: Strategies used to cope with environmental challenges. *Crit. Rev. Plant Sci.* 34, 237–280. doi: 10.1080/07352689.2014.898450

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, 202–208. doi: 10.1093/nar/gkp335

Banerjee, A., and Roychoudhury, A. (2016). Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress. *Plant Growth Regul.* 79, 1–17. doi: 10.1007/s10725-015-0113-3

Bao, F., Du, D., An, Y., Yang, W., Wang, J., Cheng, T., et al. (2017). Overexpression of *Prunus mume* dehydrin genes in tobacco enhances tolerance to cold and drought. *Front. Plant Sci.* 8:151. doi: 10.3389/fpls.2017.00151

Battaglia, M., Covarrubias, A. A., Marina, B., and Alejandra, A. C. (2013). Late embryogenesis abundant (LEA) proteins in legumes. *Front. Plant Sci.* 4:190. doi: 10.3389/fpls.2013.00190

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* 55, 504–513. doi: 10.1111/j.1365-313X.2008.03519.x

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaënsis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517

Bies-Ethève, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., et al. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* 67, 107–124. doi: 10.1007/s11103-008-9304-x

Buysse, J., and Merckx, R. (1993). An improved colorimetric method to quantify sugar content of plant tissue. *J. Exp. Bot.* 44, 1627–1629. doi: 10.1093/jxb/44.10.1627

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., et al. (2015). Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210. doi: 10.1093/molbev/msu296

Cao, Y., Xiang, X., Geng, M., You, Q., and Huang, X. (2017). Effect of HbDHN1 and HbDHN2 genes on abiotic stress responses in *Arabidopsis*. *Front. Plant Sci.* 8:470. doi: 10.3389/fpls.2017.00470

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Charfeddine, S., Saïdi, M. N., Charfeddine, M., and Gargouri-Bouzid, R. (2015). Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. *Mol. Biol. Rep.* 42, 1163–1174. doi: 10.1007/s11033-015-3853-2

Chiappetta, A., Muto, A., Bruno, L., Woloszynska, M., Lijsebettens, M., and Van Bitonti, M. B. (2015). A dehydrin gene isolated from feral olive enhances drought tolerance in *Arabidopsis* transgenic plants. *Front. Plant Sci.* 6:392. doi: 10.3389/fpls.2015.00392

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., et al. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 448, 666–671. doi: 10.1038/nature06006

Chu, Y., Guimaraes, L. A., Wu, C. L., Timper, P., Holbrook, C. C., and Ozias-Akins, P. (2014). A technique to study Meloidogyne arenaria resistance in *Agrobacterium rhizogenes*-transformed peanut. *Plant Dis.* 98, 1292–1299. doi: 10.1094/PDIS-12-13-1241-RE

Close, T. J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiol. Plant.* 97, 795–803. doi: 10.1034/j.1399-3054.1996.970422.x

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

de Brito, G. G., Sofiatti, V., de Andrade Lima, M. M., de Carvalho, L. P., and da Silva Filho, J. L. (2011). Physiological traits for drought phenotyping in cotton. *Acta Sci. Agron.* 33, 117–125. doi: 10.4025/actasciagron.v33i1.9839

De Vylder, J., Vandenbussche, F., Hu, Y., Philips, W., and Van Der Straeten, D. (2012). Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects. *Plant Physiol.* 160, 1149–1159. doi: 10.1104/pp.112.202762

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Graether, S. P., and Boddington, K. F. (2014). Disorder and function: a review of the dehydrin protein family. *Front. Plant Sci.* 5:576. doi: 10.3389/fpls.2014.00576

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Gu, Y., Xing, S., and He, C. (2016). Genome-Wide analysis indicates lineage-specific gene loss during *Papilionoideae* evolution. *Genome Biol. Evol.* 8, 635–648. doi: 10.1093/gbe/evw021

Guimaraes, L. A., Mota, A. P. Z., Araujo, A. C. G., de Alencar Figueiredo, L. F., Pereira, B. M., de Passos Saraiva, M. A., et al. (2017). Genome-wide analysis of expansin superfamily in wild *Arachis* discloses a stress-responsive expansin-like B gene. *Plant Mol. Biol.* 94, 1–18. doi: 10.1007/s11103-017-0594-8

Guimaraes, P. M., Guimaraes, L. A., Morgante, C. V., Silva, O. B., Araujo, A. C. G., Martins, A. C. Q., et al. (2015). Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. *PLoS One* 10:e0140937. doi: 10.1371/journal.pone.0140937

Guo, R., Zhao, J., Wang, X., and Wang, X. (2015). Constitutive expression of a grape aspartic protease gene in transgenic *Arabidopsis* confers osmotic stress tolerance. *PCTOC* 121, 275–287. doi: 10.1007/s11240-014-0699-6

Guo, X., Zhang, L., Zhu, J., Liu, H., and Wang, A. (2017). Cloning and characterization of SiDHN, a novel dehydrin gene from *Saussurea* involucrata Kar. et Kir. that enhances cold and drought tolerance in tobacco. *Plant Sci.* 256, 160–169. doi: 10.1016/j.plantsci.2016.12.007

Halder, T., Upadhyaya, G., Basak, C., Das, A., Chakraborty, C., and Ray, S. (2018). Dehydrins impart protection against oxidative stress in transgenic tobacco plants. *Front. Plant Sci.* 9:136. doi: 10.3389/fpls.2018.00136

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal. Behav.* 6, 1503–1509. doi: 10.4161/psb.6.10.17088

Hara, M., Fujinaga, M., and Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *J. Exp. Bot.* 56, 2695–2703. doi: 10.1093/jxb/eri262

Hill, W., Jin, X.-L., and Zhang, X.-H. (2016). Expression of an arctic chickweed dehydrin, CarDHN, enhances tolerance to abiotic stress in tobacco plants. *Plant Growth Regul.* 80, 323–334. doi: 10.1007/s10725-016-0169-8

Houde, M., Dallaire, S., N'Dong, D., and Sarhan, F. (2004). Overexpression of the acidic dehydrin WCOR410 improves freezing tolerance in transgenic strawberry leaves. *Plant Biotechnol. J.* 2, 381–387. doi: 10.1016/j.nicl.2017.06.031

Hundertmark, M., and Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9:118. doi: 10.1186/1471-2164-9-118

Jiménez-Bremont, J. F., Maruri-López, I., Ochoa-Alfaro, A. E., Delgado-Sánchez, P., Bravo, J., and Rodríguez-Kessler, M. (2013). lea gene introns: is the intron of dehydrin genes a characteristic of the serine-segment? *Plant Mol. Biol. Rep.* 31, 128–140. doi: 10.1007/s11105-012-0483-x

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kerr, T. C. C., Abdel-mageed, H., Aleman, L., Lee, J., Payton, P., Cryer, D., et al. (2018). Ectopic expression of two AREB / ABF orthologs increases drought tolerance in cotton (*Gossypium hirsutum*). *Plant Cell Environ.* 41, 898–907. doi: 10.1111/pce.12906

Ko, D. K., Nadakuduti, S. S., Douches, D. S., and Buell, C. R. (2018). Transcriptome profiling of transgenic potato plants provides insights into variability caused by plant transformation. *PLoS One* 13:e0206055. doi: 10.1371/journal.pone.0206055

Kovacs, D., Kalmar, E., Torok, Z., and Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol.* 147, 381–390. doi: 10.1104/pp.108.118208

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63, 1593–1608. doi: 10.1093/jxb/err460

Leal-Bertioli, S. C. M., Bertioli, D. J., Guimaraes, P. M., Pereira, T. D., Galhardo, I., Silva, J. P., et al. (2012). The effect of tetraploidization of wild *Arachis* on leaf morphology and other drought-related traits. *Environ. Exp. Bot.* 84, 17–24. doi: 10.1016/j.envexpbot.2012.04.005

Liang, Y., Xiong, Z., Zheng, J., Xu, D., Zhu, Z., Xiang, J., et al. (2016). Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci. Rep.* 6:24265. doi: 10.1038/srep24265

Liu, H., Yu, C., Li, H., Ouyang, B., Wang, T., Zhang, J., et al. (2015). Overexpression of ShDHN, a dehydrin gene from *Solanum habrochaites* enhances tolerance to multiple abiotic stresses in tomato. *Plant Sci.* 231, 198–211. doi: 10.1016/j.plantsci.2014.12.006

Liu, Y., Zhang, J., Li, W., Guo, C., and Shu, Y. (2015). In silico identification, phylogeny and expression analysis of expansin superfamily in *Medicago truncatula*. *Biotechnol. Biotechnol. Equip.* 30, 197–203. doi: 10.1080/13102818.2015.1093919

Malik, A. A., Veltri, M., Boddington, K. F., Singh, K. K., and Graether, S. P. (2017). Genome analysis of conserved dehydrin motifs in vascular plants. *Front. Plant Sci.* 8:709. doi: 10.3389/fpls.2017.00709

Mattana, M., Biazzi, E., Consonni, R., Locatelli, F., Vannini, C., Provera, S., et al. (2005). Overexpression of Osmyb4 enhances compatible solute accumulation and increases stress tolerance of *Arabidopsis thaliana*. *Physiol. Plant* 125, 212–223. doi: 10.1111/j.1399-3054.2005.00551.x

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263

Morgante, C. V., Brasileiro, A. C. M., Roberts, P. A., Guimaraes, L. A., Araujo, A. C. G., Fonseca, L. N., et al. (2013). A survey of genes involved in arachis stenosperma resistance to meloidogyne arenaria race 1. *Funct. Plant Biol.* 40, 1298–1309. doi: 10.1071/FP13096

Morgante, C. V., Guimaraes, P. M., Martins, A., Araujo, A. C. G., Leal-Bertioli, S. C. M., Bertioli, D. J., et al. (2011). Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes* 4:339. doi: 10.1186/1756-0500-4-339

Mota, A. P. Z., Vidigal, B., Danchin, E. G. J., Togawa, R. C., Leal-Bertioli, S. C. M., Bertioli, D. J., et al. (2018). Comparative root transcriptome of wild Arachis reveals NBS-LRR genes related to nematode resistance. *BMC Plant Biol.* 18:159. doi: 10.1186/s12870-018-1373-7

Naznin, H. A., Kiyohara, D., Kimura, M., Miyazawa, M., Shimizu, M., and Hyakumachi, M. (2014). Systemic resistance induced by volatile organic compounds emitted by plant growth-promoting fungi in *Arabidopsis thaliana*. *PLoS One* 9:e86882. doi: 10.1371/journal.pone.0086882

Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30:36. doi: 10.1093/nar/30.9.e36

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2012). Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol.* 28, 489–521. doi: 10.1146/annurev-cellbio-092910-154055

Proite, K., Carneiro, R., Falcão, R., Gomes, A., Leal-Bertioli, S., Guimarães, P., et al. (2008). Post-infection development and histopathology of *Meloidogyne arenaria* race 1 on *Arachis* spp. *Plant Pathol.* 57, 974–980. doi: 10.1111/j.1365-3059.2008.01861.x

Rémus-Borel, W., Castonguay, Y., Cloutier, J., Michaud, R., Bertrand, A., Desgagnés, R., et al. (2010). Dehydrin variants associated with superior freezing tolerance in alfalfa (*Medicago sativa* L.). *Theor. Appl. Genet.* 120, 1163–1174. doi: 10.1007/s00122-009-1243-7

Robert-Seilaniantz, A., Grant, M., and Jones, J. D. G. (2011). Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu. Rev. Phytopathol.* 49, 317–343. doi: 10.1146/annurev-phyto-073009-114447

Rosales, R., Romero, I., Escribano, M. I., Merodio, C., and Sanchez-Ballesta, M. T. (2014). The crucial role of Φ- And K-segments in the in vitro functionality of *Vitis vinifera* dehydrin DHN1a. *Phytochemistry* 108, 17–25. doi: 10.1016/j.phytochem.2014.10.006

Sasaki-Sekimoto, Y., Jikumaru, Y., Obayashi, T., Saito, H., Masuda, S., Kamiya, Y., et al. (2013). Basic helix-loop-helix transcription factors JASMONATE-ASSOCIATED MYC2-LIKE1 (JAM1), JAM2, and JAM3 are negative regulators of jasmonate responses in *Arabidopsis*. *Plant Physiol.* 163, 291–304. doi: 10.1104/pp.113.220129

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Seo, P. J., Xiang, F., Qiao, M., Park, J., Lee, Y. N., Kim, S., et al. (2009). The MYB96 transcription factor mediates abscisic acid signaling during drought stress response. *Plant Physiol.* 151, 275–289. doi: 10.1104/pp.109.144220

Shekhawat, U. K. S., Srinivas, L., and Ganapathi, T. R. (2011). MusaDHN-1, a novel multiple stress-inducible SK3-type dehydrin gene, contributes affirmatively to drought- and salt-stress tolerance in banana. *Planta* 234, 915–932. doi: 10.1007/s00425-011-1455-3

Shen, Y., Tang, M. J., Hu, Y. L., and Lin, Z. P. (2004). Isolation and characterization of a dehydrin-like gene from drought-tolerant *Boea crassifolia*. *Plant Sci.* 166, 1167–1175. doi: 10.1016/j.plantsci.2003.12.025

Singh, D., and Laxmi, A. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front. Plant Sci.* 6:895. doi: 10.3389/fpls.2015.00895

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Tiwari, P., Indoliya, Y., Singh, P. K., Singh, P. C., Chauhan, P. S., Pande, V., et al. (2018). Role of dehydrin-FK506-binding proteins complex in enhancing drought tolerance through ABA-mediated signaling pathway. *Environ. Exp. Bot.* 158, 136–149. doi: 10.1016/j.envexpbot.2018.10.031

Tommasini, L., Svensson, J. T., Rodriguez, E. M., Wahid, A., Malatrasi, M., Kato, K., et al. (2008). Dehydrin gene expression provides an indicator of low temperature and drought stress: transcriptome-based analysis of Barley (*Hordeum vulgare* L.). *Funct. Integr. Genomics* 8, 387–405. doi: 10.1007/s10142-008-0081-z

Turco, E., Close, T. J., Fenton, R. D., and Ragazzi, A. (2004). Synthesis of dehydrin-like proteins in *Quercus ilex* L. and *Quercus cerris* L. seedlings subjected to water stress and infection with *Phytophthora cinnamomi*. *Physiol. Mol. Plant Pathol.* 65, 137–144. doi: 10.1016/j.pmpp.2004.11.010

Van der Does, D., Leon-Reyes, A., Koornneef, A., Van Verk, M. C., Rodenburg, N., Pauwels, L., et al. (2013). Salicylic acid suppresses jasmonic acid signaling downstream of SCFCOI1-JAZ by targeting GCC promoter motifs via transcription factor ORA59. *Plant Cell* 25, 744–761. doi: 10.1105/tpc.112.108548

Vinson, C. C., Mota, A. P. Z., Oliveira, T. N., Guimaraes, L. A., Leal-Bertioli, S. C. M., Williams, T. C. R., et al. (2018). Early responses to dehydration in contrasting wild Arachis. *PLoS One* 13:e0198191. doi: 10.1371/journal.pone.0198191

Wang, H., Zhang, H., and Li, Z. (2007). Analysis of gene expression profile induced by water stress in upland rice (*Oryza sativa* L. var. IRAT109) seedlings using subtractive expressed sequence tags library. *J. Integr. Plant Biol.* 49, 1455–1463. doi: 10.1111/j.1672-9072.2007.00553.x

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Xie, C., Zhang, R., Qu, Y., Miao, Z., Zhang, Y., Shen, X., et al. (2012). Overexpression of MtCAS31 enhances drought tolerance in transgenic *Arabidopsis* by reducing stomatal density. *New Phytol.* 195, 124–135. doi: 10.1111/j.1469-8137.2012.04136.x

Yamasaki, Y., Koehler, G., Blacklock, B. J., and Randall, S. K. (2013). Dehydrin expression in soybean. *Plant Physiol. Biochem.* 70, 213–220. doi: 10.1016/j.plaphy.2013.05.013

Yang, S., Seo, P. J., Yoon, H., and Park, C. (2011). The *Arabidopsis* NAC transcription factor VNI2 integrates abscisic acid signals into leaf senescence via the COR / RD genes. *Plant Cell* 23, 2155–2168. doi: 10.1105/tpc.111.084913

Yang, Y., He, M., Zhu, Z., Li, S., Xu, Y., Zhang, C., et al. (2012). Identification of the dehydrin gene family from grapevine species and analysis of their

responsiveness to various forms of abiotic and biotic stress. *BMC Plant Biol.* 12:140. doi: 10.1186/1471-2229-12-140

Yin, Z., Rorat, T., Szabala, B. M., Ziółkowska, A., and Malepszy, S. (2006). Expression of a *Solanum sogarandinum* SK3-type dehydrin enhances cold tolerance in transgenic cucumber seedlings. *Plant Sci.* 170, 1164–1172. doi: 10.1016/j.plantsci.2006.02.002

Yu, Z., Wang, X., and Zhang, L. (2018). Structural and functional dynamics of dehydrins: A plant protector protein under abiotic stress. *Int. J. Mol. Sci.* 19:3420. doi: 10.3390/ijms19113420

Zhang, H., Zheng, J., Su, H., Xia, K., Jian, S., and Zhang, M. (2018). Molecular cloning and functional characterization of the dehydrin (IpDHN) gene from Ipomoea pes-caprae. *Front. Plant Sci.* 9:1454. doi: 10.3389/fpls.2018.01454

Zhao, J., Zhang, X., Guo, R., Wang, Y., Guo, C., Li, Z., et al. (2018). Over-expression of a grape WRKY transcription factor gene, VlWRKY48, in *Arabidopsis thaliana* increases disease resistance and drought stress tolerance1. *Plant Cell Tissue Organ Cult.* 132, 359–370. doi: 10.1007/s11240-017-1335-z

Zhao, S., and Fernald, R. D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.* 12, 1047–1064. doi: 10.1089/cmb.2005.12.1047.Comprehensive

Zolotarov, Y., and Strömvik, M. (2015). De novo regulatory motif discovery identifies significant motifs in promoters of five classes of plant dehydrin genes. *PLoS One* 10:e0129016. doi: 10.1371/journal.pone.0129016

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Genome-wide analysis of expansin superfamily in wild Arachis discloses a stress-responsive expansin-like B gene**

Published in Plant Molecular Biology – 2017

Larissa Arrais Guimaraes, Ana Paula Zotta Mota, Ana Claudia Guerra Araujo, Lucio Flavio de Alencar Figueiredo, Bruna Medeiros Pereira, Mario Alfredo de Passos Saraiva, Raquel Bispo Silva, Etienne G. J. Danchin, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

Plant Mol Biol (2017) 94:79–96
DOI 10.1007/s11103-017-0594-8

# Genome-wide analysis of expansin superfamily in wild *Arachis* discloses a stress-responsive expansin-like B gene

**Larissa Arrais Guimaraes[1] · Ana Paula Zotta Mota[1,2] · Ana Claudia Guerra Araujo[1] · Lucio Flavio de Alencar Figueiredo[3] · Bruna Medeiros Pereira[1,3] · Mario Alfredo de Passos Saraiva[1] · Raquel Bispo Silva[1,3] · Etienne G. J. Danchin[4] · Patricia Messenberg Guimaraes[1] · Ana Cristina Miranda Brasileiro[1]**

Received: 7 October 2016 / Accepted: 13 February 2017 / Published online: 27 February 2017


**Abstract** Expansins are plant cell wall-loosening proteins involved in adaptive responses to environmental stimuli and various developmental processes. The first genome-wide analysis of the expansin superfamily in the *Arachis* genus identified 40 members in *A. duranensis* and 44 in *A. ipaënsis*, the wild progenitors of cultivated peanut (*A. hypogaea*). These expansins were further characterized regarding their subfamily classification, distribution along the genomes, duplication events, molecular structure, and phylogeny. A RNA-seq expression analysis in different *Arachis* species showed that the majority of these expansins are modulated in response to diverse stresses such as water deficit, root-knot nematode (RKN) infection, and UV exposure, with an expansin-like B gene (*AraEXLB8*) displaying a highly distinct stress-responsive expression profile. Further analysis of the *AraEXLB8* coding sequences showed high conservation across the *Arachis* genotypes, with eight haplotypes identified. The modulation of *AraEXLB8* expression in response to the aforementioned stresses was confirmed by qRT-PCR analysis in distinct *Arachis* genotypes, whilst in situ hybridization revealed transcripts in different root tissues according to the stress imposed. The overexpression of *AraEXLB8* in soybean (*Glycine max*) composite plants remarkably decreased the number of galls in transformed hairy roots inoculated with RKN. This study improves the current understanding of the molecular evolution, divergence, and gene expression of expansins in *Arachis*, and provides molecular and functional insights into the role of expansin-like B, the less-studied plant expansin subfamily.

**Keywords** Drought · *Meloidogyne* · Ultraviolet (UV) · Composite plant · *Glycine max*

## Introduction

Expansins are cell wall loosening proteins involved in extension and relaxation of cells in growing tissues through a non-enzymatic activity (McQueen-Mason et al. 1992). Plant expansins are also implicated in responses to many abiotic and biotic stresses, such as drought; salinity; cold; heat; oxidative stress; herbivore attack and phytopathogen infection (Marowa et al. 2016), indicating that these proteins constitute a common component in the response of plants to stress. The role of expansins in stress responses has been reinforced by studies showing that the ectopic overexpression in transgenic plants of some expansin-coding genes leads, in general, to enhanced overall plant growth and tolerance to abiotic stress (Cosgrove 2015; Marowa et al. 2016; Sasidharan et al. 2011). In contrast, the reduction of expansin transcripts, which inhibits cell expansion, leads to an increase in resistance to plant diseases

Accession Numbers The *AraEXLB8* coding sequences from 13 *Arachis* genotypes have been deposited at NCBI GenBank under the accession KX588113 to KX588125.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-017-0594-8) contains supplementary material, which is available to authorized users.

✉ Ana Cristina Miranda Brasileiro
ana.brasileiro@embrapa.br

[1] Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Final W5 Norte, Brasília, DF CP 02372, Brazil

[2] Universidade do Rio Grande do Sul, Porto Alegre, RS, Brazil

[3] Universidade de Brasília, Campus Darcy Ribeiro, Brasília, DF, Brazil

[4] Institut Sophia Agrobiotech, INRA, University of Nice Sophia Antipolis, CNRS, 06900 Sophia Antipolis, France

possibly due to a more efficient cell wall physical protection (Ding et al. 2008; Gal et al. 2006). Expansins are also involved in responses to oxidative stress (Han et al. 2015) derived from enhanced production and accumulation of Reactive Oxygen Species (ROS) (Baxter et al. 2014). As a secondary stress, oxidative damage is generally ubiquitous in almost all types of stresses, including ultraviolet (UV) exposure, thus supporting the idea that expansins are common and crucial players in the response to multiple and simultaneous stresses. Nevertheless, the non-enzymatic mechanisms by which expansins influence the ability of plants to withstand associated stresses remain uncertain.

The expansin superfamily is ubiquitous in the plant kingdom and classified into four subfamilies: α-expansin (EXPA), β-expansin (EXPB), expansin-like A (EXLA) and expansin-like B (EXLB) (Kende et al. 2004). It has been assumed that expansin genes belonging to the same phylogenetic group share almost similar functions and effects regarding plant development and growth (Marowa et al. 2016). To date, research efforts have been concentrated on elucidating the mode of action and the biological function of expansins belonging to the two largest subfamilies, EXPA and EXPB, and, to a lesser extent EXLA, whilst little attention was given to the EXLB members (Cosgrove 2016; Marowa et al. 2016).

Recently, genome-wide investigations of expansins in species belonging to Fabaceae family, such as *Glycine max* (soybean), *Phaseolus vulgaris* (common bean) and *Medicago truncatula* (Cosgrove 2015; Liu et al. 2015; Zhu et al. 2014) have brought new insights into the molecular and evolutionary history of expansins in legumes, as well as their functions and regulatory mechanisms. In *G. max*, the expansin superfamily has expanded predominantly by segmental duplication and presents a broad functional divergence among subfamilies (Zhu et al. 2014). Likewise, studies in *M. truncatula* revealed that both segmental and tandem duplications contributed to the evolution and diversification of expansins and that genes involved in same processes are closely located on chromosomes (Liu et al. 2015). In this regard, the recently available genome sequences of *Arachis duranensis* and *A. ipaënsis*, the wild progenitors of cultivated peanut (*A. hypogaea*) (Bertioli et al. 2016; Chen et al. 2016), represents a great opportunity to advance the knowledge on plant expansins in *Arachis* and other legumes.

Wild *Arachis* species have been exploited in the last years as sources of alleles to enhance environmental adaptability and to disclose candidate genes, genetic markers, and genomic sequences for peanut breeding improvement (Brasileiro et al. 2014; Janila et al. 2016). In this context, previous transcriptome analysis revealed several differentially expressed genes responsive to drought, including expansin genes, in *A. duranensis* and *A. magna* (Brasileiro et al. 2015; Guimaraes et al. 2012). Moreover, *A. stenosperma* transcriptome survey identified, amongst other candidates, expansin genes putatively involved in resistance to the root-knot nematode (RKN) *Meloidogyne arenaria* (Guimaraes et al. 2015). Either through harboring improved performance under water-limited conditions (*A. duranensis* and *A. magna)* or higher resistance to several pathogens (*A. stenosperma*), these wild species proved to be valuable for stress-related gene discovery.

The potential of expansin-coding genes to enhance stress tolerance, associated with the limited knowledge of their roles in plant responses to environmental cues, highlights the importance of characterizing the expansin superfamily and their functional mechanisms, in order to enable their use for plant improvement. The present study reports the first genome-wide identification and analysis of the expansin superfamily in the genus *Arachis* and the molecular and functional characterization of a novel stress-responsive expansin-like B gene (*AraEXLB8*), a member of the less-studied expansin subfamily.

## Results

### Genome-wide identification and analysis of *Arachis* expansins

Overall, 40 and 44 putative expansins were identified in the genomes of *A. duranensis* and *A. ipaënsis*, respectively, with most of them ranging between 250 and 275 amino acids (AAs) in size and harboring signal peptides for secretion (Supplementary Tables 1 and 2), as expected from canonical plant expansins (Sampedro and Cosgrove 2005). Using a phylogenetic analysis, most of the *Arachis* expansin genes could be assigned to the four subfamilies proposed by Kende et al. (2004) with high confidence values (Fig. 1a; Supplementary Figs. 1, 2, 3). The phylogeny showed that EXPA, EXPB and EXL (A and B) subfamilies formed highly supported monophyletic groups. However, within the EXL group, EXLB proteins did not form a highly supported monophyletic group but were interspersed by the highly supported EXLA monophyletic group, although this separation of the EXLBs in two groups received only a moderate bootstrap support value of 47 (Fig. 1a). A further phylogenetic analysis, including only EXLAs and EXLBs, confirmed either the separation of EXLB in two subgroups (Supplementary Figs. 4a, 5) or low support for the monophyly of EXLBs (Supplementary Fig. 4b). Among *A. duranensis* expansins, the EXPA subfamily constituted the largest clade with 25 members, followed by eight EXLB, six EXPB and one EXLA member, whilst for *A. ipaënsis*, expansins were classified as

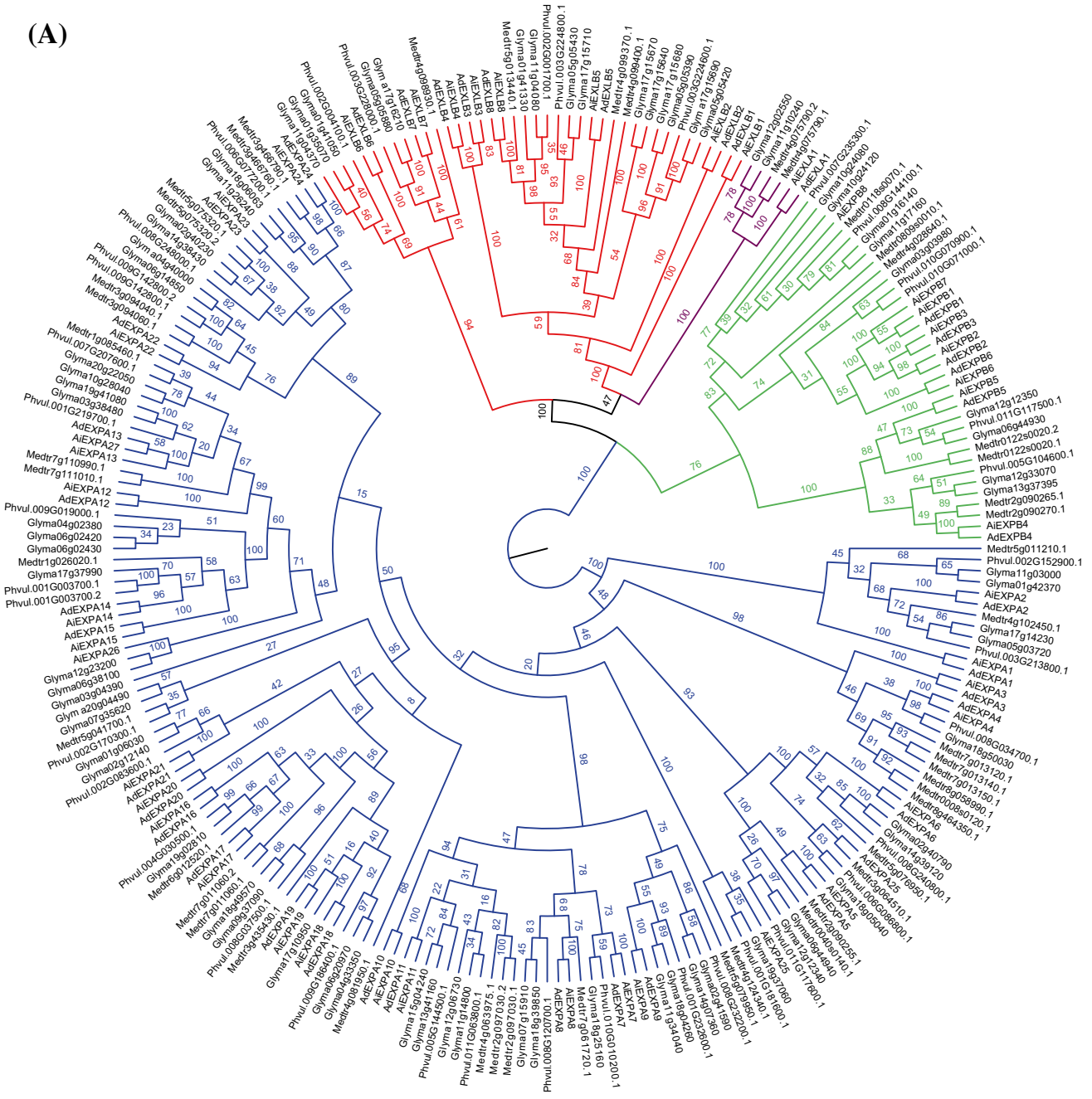



**Fig. 1** Phylogenetic analysis and expansin distribution in *Arachis duranensis* and *Arachis ipaënsis*. **a** Phylogenetic analysis of the expansin superfamily in *A. duranensis* (Ad); *A. ipaënsis* (Ai); *Glycine max* (Glyma); *Phaseolus. vulgaris* (Phvul) and *Medicago truncatula* (Medtr). **b** Distribution of expansin genes in the ten chromosomes of each *A. duranensis* (*light gray*; A01–A10) and *A. ipaënsis* (*dark gray*; B01–B10). Synteny between the two genomes is represented by *lines*. EXPA (*blue*), EXPB (*green*), EXLA (*purple*) and EXLB (*red*)

27 EXPA, eight EXLB, eight EXPB, and a single EXLA (Fig. 1a; Supplementary Tables 1, 2, 3).

Expansins were unevenly distributed on all the chromosomes of both *Arachis* species (Fig. 1b). Whilst the EXPA members (blue) were distributed on each chromosome of both species, EXLB (red) and EXPB (green) tend to be grouped on chromosomes 1 and 2, respectively, and appear to be mutually exclusive in their locations in each species. Similar uneven distribution of the EXLB members, clustered in one or two chromosomes, was previously observed in *Malus* x *Domestica* (Zhang et al. 2014b) and *G. max* (Zhu et al. 2014).

Whole genome level analysis showed that the majority of the predicted protein-coding genes for *A. duranensis*

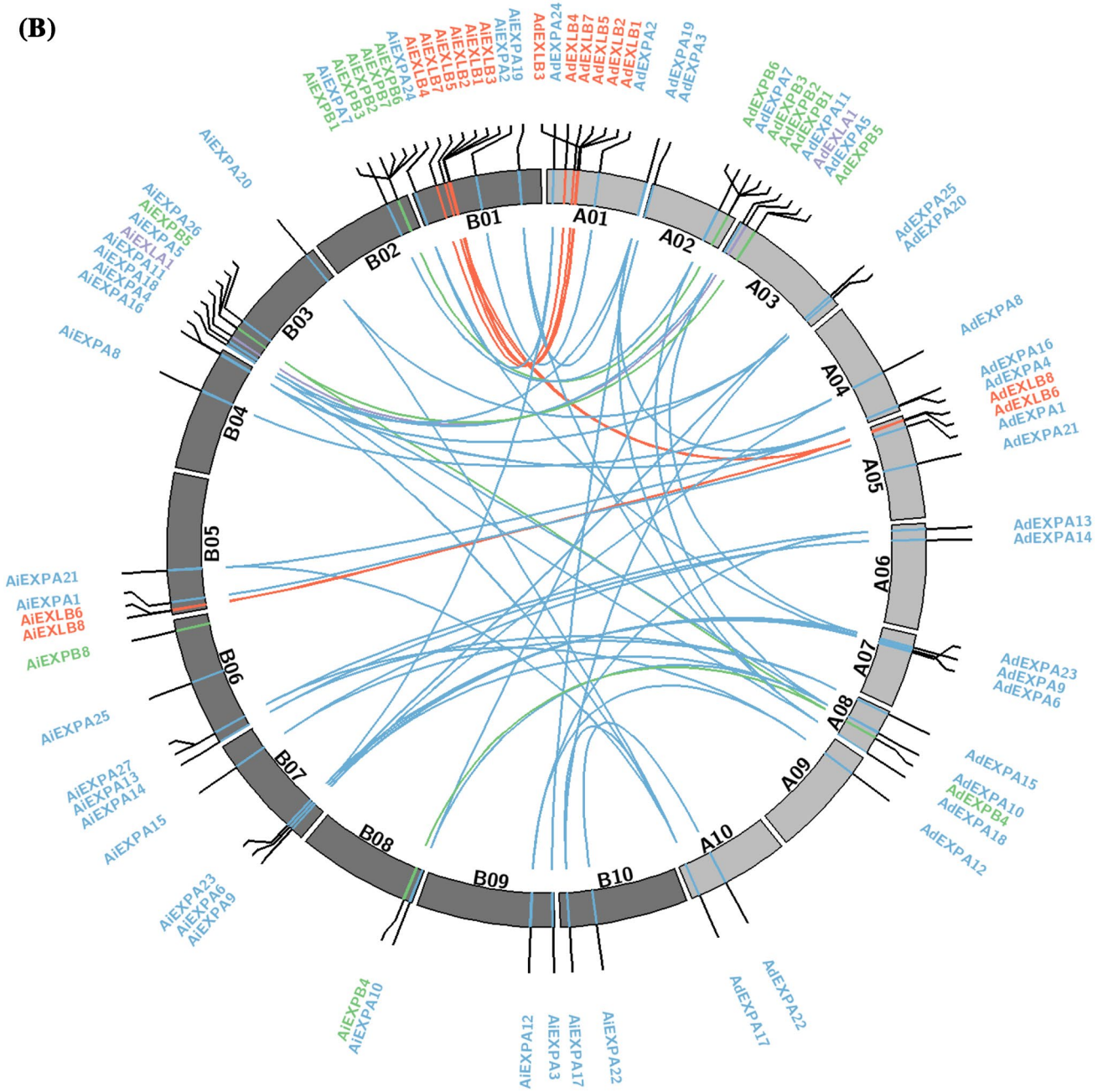



**Fig. 1** (continued)

(87.2%) and *A. ipaënsis* (88.6%) were duplicated. In both genomes, these duplicated copies resulted mostly from dispersed gene duplication (>68%) all around the genomes, whilst less than 17% resulted from whole genome duplication (WGD) (Supplementary Table 4). In contrast, expansion of the expansin superfamily genes mainly resulted (>50%) from WGD/segmental duplication events in both species rather than from dispersed duplications (Supplementary Table 7). Within the expansin subfamilies (Supplementary Table 5), the most common type of duplication for EXPA and EXLB was WGD/segmental duplication, whilst for EXPB, tandem duplications represented the majority. Further analysis also showed that the expansion of the expansin subfamilies in *Arachis* was comparable to that observed in *Arabidopsis thaliana* and *G. max* (Supplementary Table 5).

To predict the selective pressure between *Arachis* expansins duplicated genes, the ratio of *Ka* (rate of non-synonymous mutations)/*Ks* (rate of synonymous mutations) was calculated and revealed that only the

*AiEXLB5*/*AiEXLB8* genes pair from *A. ipaënsis* appeared to have undergone positive selection (*Ka*/*Ks* ratio > 1).

Phylogenetic analysis indicated that most duplication events of the expansin genes preceded *A. duranensis* and *A. ipaënsis* divergence (Fig. 1a) and might have occurred in a common ancestor, most probably since the divergence of the Dalbergioid clade (Bertioli et al. 2011). Although large inversions of both arms have been previously observed in chromosome 1 of *Arachis* species (Bertioli et al. 2016), most of the conserved syntenic blocks of duplicated expansins are located on chromosomes 1, 2 and 3 (Fig. 1b). Moreover, the high density of expansin genes in *A. duranensis* chromosome 8 is in accordance with the gene-rich characteristic of this abnormally small 'A chromosome' that contains low repetitive content and many genes (Bertioli et al. 2016).

### *Arachis* expansin gene and protein structures

*Arachis duranensis* and *A. ipaënsis* shared conserved expansin gene structures (Fig. 2) and showed diverse organizations comprising of two to six exons. Members of the same phylogenetically-determined subfamily were characterized by a similar intron/exon organization; in particular, those of the EXPB subfamily, in which all members consisted of four exons, except *AiEXPB8*. The majority of the EXPA members (82%) had three exons, while EXLB genes presented the largest number of exons, ranging from four to six, similar to *Malus* x *Domestica* and *G. max* (Zhang et al. 2014b; Zhu et al. 2014). The EXLA members had the same gene structure (4-exon/3-intron) in both *Arachis* species which differs from the usual 5-exon/4-intron EXLA structure of other plants (Ding et al. 2016; Krishnamurthy et al. 2014; Sampedro et al. 2005).

Predicted expansin protein structures (Fig. 2) showed a highly conserved pattern of DPBB and CBM63 Interpro domains (Cosgrove 2015) (Supplementary Figs. 2, 3), preceded by a signal peptide with 21 to 36 AA in most N-terminus. Also, the highly conserved intron/exon structure of *Arachis* expansins within each subfamily supports the classification proposed here, as well as the evidence of the close evolutionary relationship among them. Two EXPA members (*AdEXPA3* and *AdEXPA13*) contained unusual additional copies of DPBB or CBM63 domains in their predicted AA sequences. Since these atypical structures could be the result of gene prediction errors, they were not considered for further characterization.

### RNA-seq expression profiling of expansin genes under biotic and abiotic stresses

In order to gain broader understanding of the *Arachis* expansin superfamily gene regulation in response to stress, our previous transcriptome data of *A. duranensis* and *A. stenosperma* plants submitted to abiotic (water deficit and UV exposure) and biotic (RKN *M. arenaria* inoculation) stresses were here exploited. The expression profile (Fig. 3) indicated that all the *Arachis* expansin genes were modulated in response to at least one of the imposed stresses, with distinct expression behaviors depending on the species and stress. The hierarchical clustering analysis classified the expansin genes into five clusters (A–E) according to similarities in their transcript abundance patterns (Fig. 3). Clusters A and C were the most distinct, as each harbored only one representative, *AdEXLB8* and *AdEXPB6*, respectively, as a result of their unique expression profiles. Clusters B and D contained a similar amount of genes (14 and 13, respectively), encompassing representatives of three subfamilies (EXPA; EXPB and EXLB), whereas cluster E, with nine genes, included a single member of EXLA subfamily (*AdEXLA1*) but no representatives of the EXPB subfamily. Genes forming cluster D showed slight variations on their expression levels in response to the variable stresses applied. Differently, higher levels of gene expression and contrasting expression behaviors of *A. duranensis* expansin genes and *A. stenosperma* orthologs were observed between clusters B and E (Fig. 3). In cluster E, the majority of expansins were upregulated in *A. stenosperma* orthologs upon *M. arenaria* inoculation in roots (SN3, SN6 and SN9) or UV exposure in leaves (SUV), whereas a general downregulation was observed in *A. duranensis* (DN3, DN6 and DN9; Fig. 3). Interestingly, in cluster B, the opposite expression behavior occurred in response to *M. arenaria* infection (Fig. 3), in particular for *AdEXLB1, AdEXLB3* and *AdEXPB3* genes, which showed a strong upregulation in *A. duranensis*. Regarding the drought imposition treatment, however, the expression profile displayed by expansins belonging to clusters B, C and E (Fig. 3) were quite similar between *A. duranensis* (DDro) and *A. stenosperma* orthologs (SDro). These results suggest that *Arachis* expansin genes may exhibit both functional differentiation in the two species and overlapping responses between abiotic and biotic stresses.

An expansin-like B gene (*EXLB8*) has drawn attention as it represented an outlier among the *Arachis* expansin genes, with a highly specific expression in response to each stress (Fig. 3). Moreover, our previous studies suggested the involvement of *EXLB8* in RKN-resistance in *A. stenosperma* (Guimaraes et al. 2015), as well as in drought tolerance in *A. duranensis* and *A. magna* (Brasileiro et al. 2015). *EXLB8* gene from wild *Arachis* species (hereafter named *AraEXLB8*) was therefore selected for a broader characterization at molecular and functional levels, as the most promising candidate, among the *Arachis* expansin genes, to be further exploited for plant genetic engineering.

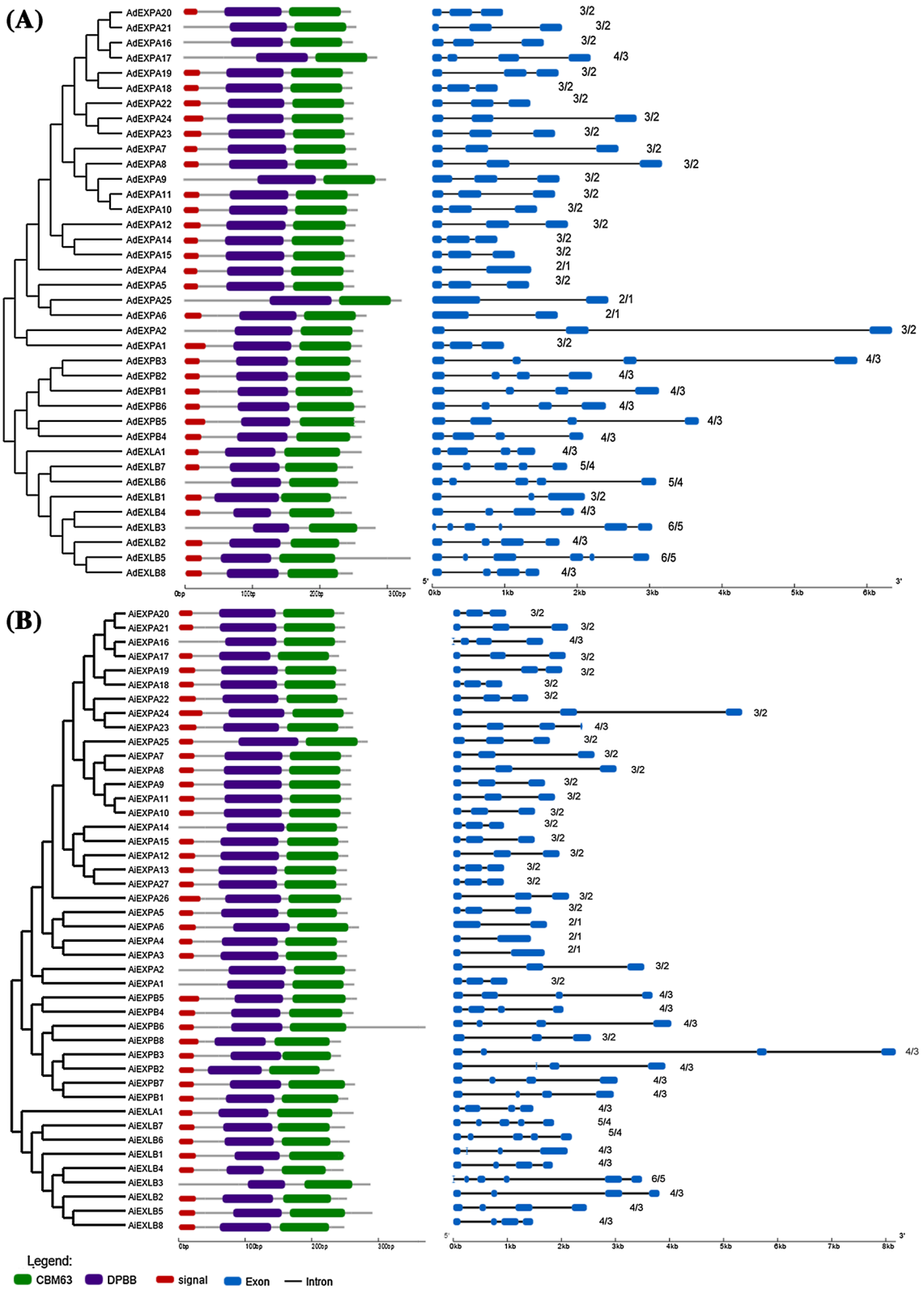



**Fig. 2** Protein and exon/intron gene structure of expansins in **a** *Arachis duranensis* and **b** *Arachis ipaënsis*. Protein structure diagram (*left*): DPBB and CBM63 domains are represented by *purple* and *green boxes*, respectively, and peptide signals by *red boxes*. Gene structure diagram (*right*): exons are represented by *blue boxes* and introns by linking *single lines*. The number of introns and exons, respectively, is after each diagram

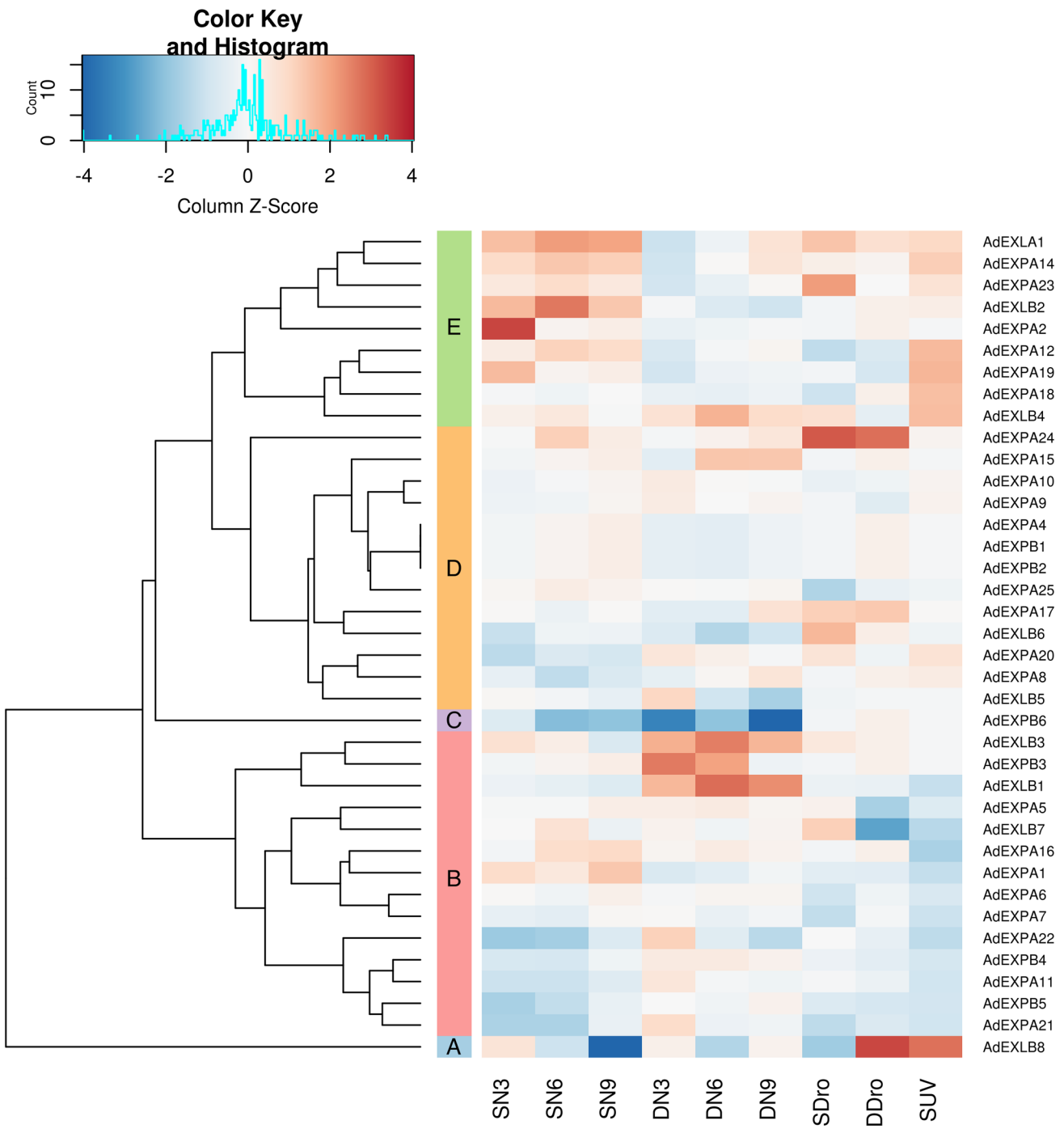



**Fig. 3** Heatmap of expansin transcripts of *Arachis duranensis* genes and *Arachis stenosperma* orthologs. Expression patterns (log2-based values) in roots of *A. duranensis* (D) expansin genes and *A. stenosperma* orthologs (S) at 3 (N3), 6 (N6) and 9 (N9) DAI with *M. arenaria*, during drought stress (Dro) and after UV exposure (UV), based on RNA-seq data

## Characterization and phylogenetic analysis of *AraEXLB8* gene

The 13 *AraEXLB8* coding sequences here identified (NCBI accession numbers KX588113 to KX588125), together with three additional available sequences, were used for the phylogenetic analysis. These 16 *AraEXLB8* genes from different *Arachis* genotypes showed 97.7% of identity, with aligned segment sizes ranging from 740 to 755 bp. Overall, 19 nucleotide substitutions were detected, with the majority (n = 12) occurring in exon 3 (Table 1; Supplementary Tables 6, 7). Based on Blosum62 matrix, most

**Table 1** Haplotypes of *AraEXLB8* gene in the 16 *Arachis* genotypes based on coding sequences segment (755 bp)

| SNP | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | Genome type | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H | Fs | Fr | E1 | | | | E2 | E3 | | | | | | | | | | | | E4 | | A | B | K | AB |
| H1 | 5 | 31 | G | C | G | C | C | A | G | A | C | A | A | C | A | A | T | C | C | C | C | – | – | – | 5 |
| H2 | 1 | 6 | * | * | * | * | * | * | * | * | * | * | * | * | * | * | C | * | * | * | * | 1 | – | – | – |
| H3 | 1 | 6 | * | * | T | T | * | * | * | * | * | * | * | * | * | * | C | * | * | * | * | 1 | – | – | – |
| H4 | 1 | 6 | * | T | * | * | * | * | * | * | * | G | * | * | * | G | * | * | * | * | * | – | – | – | 1 |
| H5 | 1 | 6 | * | * | T | * | * | T | * | * | * | * | C | * | * | * | C | G | * | G | T | – | – | 1 | – |
| H6 | 3 | 19 | * | * | T | * | * | * | T | * | G | * | * | * | T | * | C | * | T | G | T | 2 | 1 | – | – |
| H7 | 1 | 6 | * | * | T | * | G | C | * | T | * | * | C | T | * | * | C | * | * | G | T | – | 1 | – | – |
| H8 | 3 | 19 | A | * | T | * | * | * | * | * | * | * | C | T | * | * | C | * | * | G | T | 1 | 2 | – | – |

*H* Haplotype; *SNP* single nucleotide polymorphism; *Fs* simple frequency; *Fr* relative frequency (%); *E* exon; * the same base of the first haplotype (H1)

of substitutions can be considered relatively conservative. Although four SNPs (1; 2; 11 and 13) were identified as non-synonymous/non-conservative (Supplementary Tables 6, 7), these nucleotide substitutions did not seem to change protein conformation, as suggested by the *AraEXLB8* 3D protein modeling predicted for four species (*A. duranensis; A. stenosperma; A. ipaënsis* and *A. batizocoi*) which showed no differences among their tertiary structure (data not shown). Interestingly, even without any putative effect on protein folding/structure, SNPs 11 and 13 occurred in the vicinity of a catalytic site presented in the crystal structure of a native *Zea mays* EXPB1 (Yennawar et al. 2006).

Eight haplotypes were further identified without any recombination event occurring (Table 1). Haplotype 1 (H1) stands out as it comprised all tetraploid genotypes (AB) analyzed here, with the highest SNP frequency being 31%. Other haplotypes included only wild diploids, such as the H6 with two A (*A. villosa* and *A. stenosperma*) and one B (*A. gregoryi*) genome species, while the H8 comprised one A (*A. cardenasii*) and two B (both *A. ipaënsis*) species.

Phylogenetic analysis conducted with the above 16 *AraEXLB8* coding sequences showed two distinct groups, with a posterior probability support value from 0.52 to 1, based on Bayesian Inference analysis (Supplementary Fig. 6). The first clade comprised all six tetraploids [AB: wild (n = 1), synthetic (n = 1), and cultivated (n = 4)] derived from two distinct genotypes of *A. duranensis* (A genome) whilst the second clade was composed exclusively of diploids species with different genomes (A, B, and K).

### *AraEXLB8* expression in response to stresses

The expression levels of *AraEXLB8* analyzed by qRT-PCR in roots collected at the endpoint of the dry-down assay showed a significant and general upregulation in response to water deficit in all of the 13 genotypes studied (Fig. 4a). Notably, the three wild species with higher expression (*A. gregoryi, A. villosa* and *A. stenosperma* with RQ = 41.5-; 16.3- and 11.5-fold, respectively) are those that were grouped as H6 (Table 1).

Regarding the response to RKN, the *AraEXLB8* expression in roots during early stages of *M. arenaria* infection (3 DAI) displayed a clear contrasting behavior in the three *Arachis* species tested (Fig. 4b): upregulation (3.2-fold) in the RKN-resistant *A. stenosperma* and downregulation in the susceptible *A. hypogaea* 'Runner' (0.3-fold) and in the moderately resistant *A. duranensis* (0.6-fold). Despite the current lack of knowledge on the mechanisms through which *AraEXLB8* might operate to contribute to RKN resistance, our results showed a clear trend between *AraEXLB8* expression and plant resistance since the species displaying the lowest expression was the least resistant (*A. hypogaea* 'Runner') and the species with the highest expression was the most resistant (*A. stenosperma*). Distinct expression behavior was also observed in response to UV treatment in leaves, where *AraEXLB8* was highly induced in *A. stenosperma* (24.3-fold) but to a lesser extent, in *A. duranensis* and *A. hypogaea* 'Runner' (5.3- and 1.8-fold, respectively) (Fig. 4c).

To depict the spatial distribution of *AraEXLB8* transcripts in roots, eight *Arachis* genotypes were analyzed by *in situ* hybridization (ISH) using samples collected at the endpoint of the dry-down or few days after *M. arenaria* inoculation. RNA was preserved in all samples as shown by the orange acridine treatment (Fig. 5a). Regardless of the morphological disorganization of the root cells due to drought stress, ISH signals were detected in the outer cortical and epidermal cells, with evident signals in *A. gregoryi* (Fig. 5b) and *A. villosa* (Fig. 5c), and lack of signals in well-watered control roots (Fig. 5d). In RKN-inoculated roots, *AraEXLB8* transcripts were mostly present in the phloem and surrounding endodermal cells, with signals more evident in *A. stenosperma* (Fig. 5e) than in *A.*

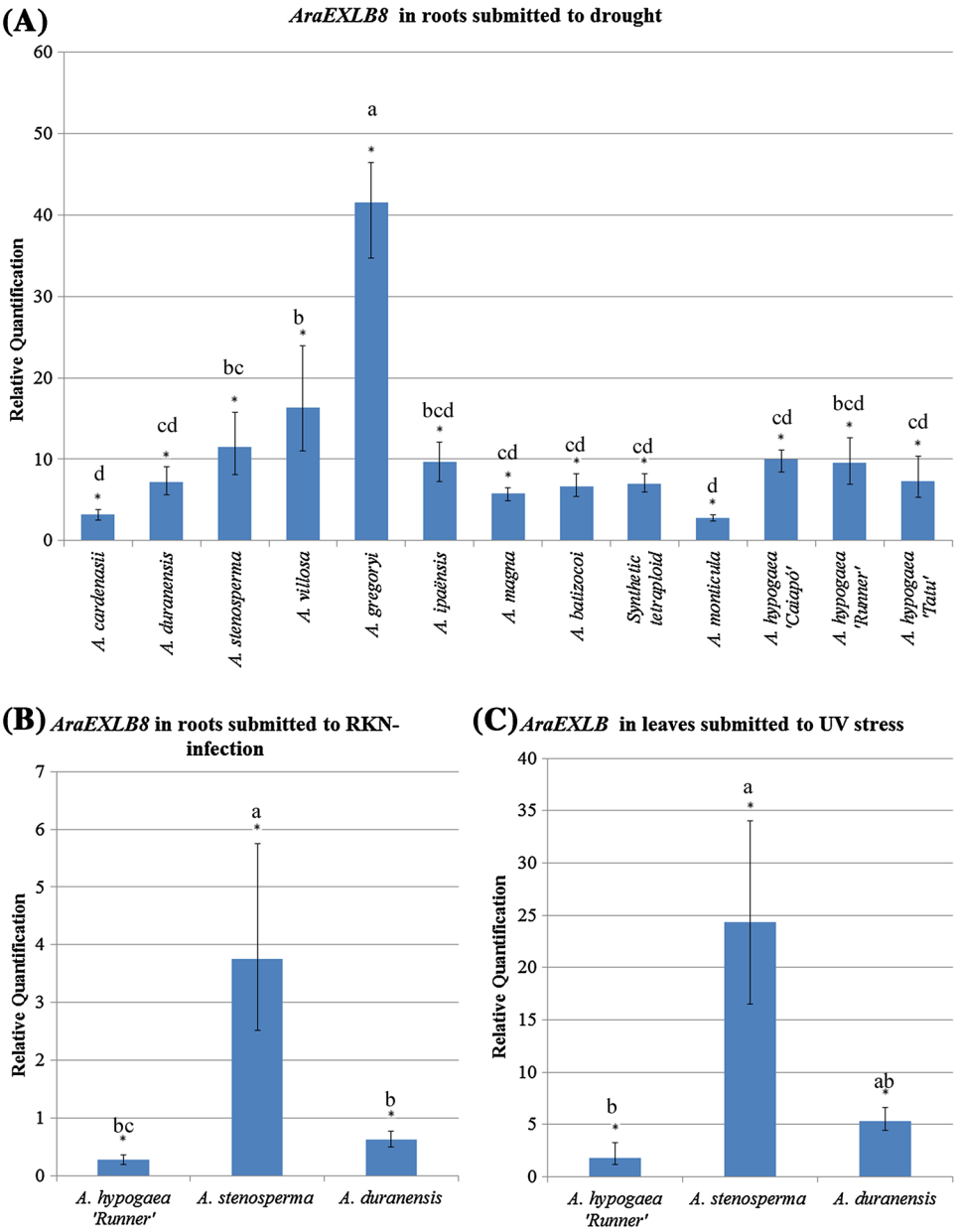



**Fig. 4** Relative quantification (RQ) of *AraEXLB8* transcripts. Expression profiles of *AraEXLB8* determined by qRT-PCR in **a** roots of 13 *Arachis* genotypes at the NTR around 0.3; in *A. hypogaea* 'Runner', *A. duranensis* and A. stenosperma **b** roots inoculated with *M. arenaria* at 3 DAI and **c** leaves 24 h after UV treatment. *Asterisks* significantly regulated genes; *letters* statistical differences between genotype samples

*hypogaea* 'Runner' (data not shown). No signals were observed in stressed samples hybridized with a sense (T7) probe (Fig. 5f). Hybridization signals showed different intensity depending on the plant genotype and tissue accordingly to the stress imposed.

## Overexpression of *AdEXLB8* in *G. max* composite plants

Three weeks after *Agrobacterium rhizogenes* transformation, *G. max* composite plants developed transgenic hairy roots at the wounding sites (with an efficiency of 78 and 60% with empty and pPZP-AdEXLB8 vectors, respectively), which were inoculated with 1,000 *M. javanica* juveniles. After 60 days, the amount of eGFP-positive roots per composite plant varied between empty and pPZP-AdEXLB8 vectors (36 and 19% of the total hairy roots, respectively) (Supplementary Table 8).

*Meloidogyne javanica* juveniles were able to complete at least one full life cycle in *G. max* transgenic hairy roots, as all of the eGFP-positive roots transformed by the empty vector showed gall and egg mass formation, with an average of 19 galls per gram of root (Fig. 6a, b; Supplementary Table 8). Conversely, galls were poorly observed in the eGFP-positive roots transformed with pPZP-AdEXLB8 (average of 3 galls per gram of root), which constitutes a significant reduction of 82% in the number of galls compared to the empty vector control, with results varying for

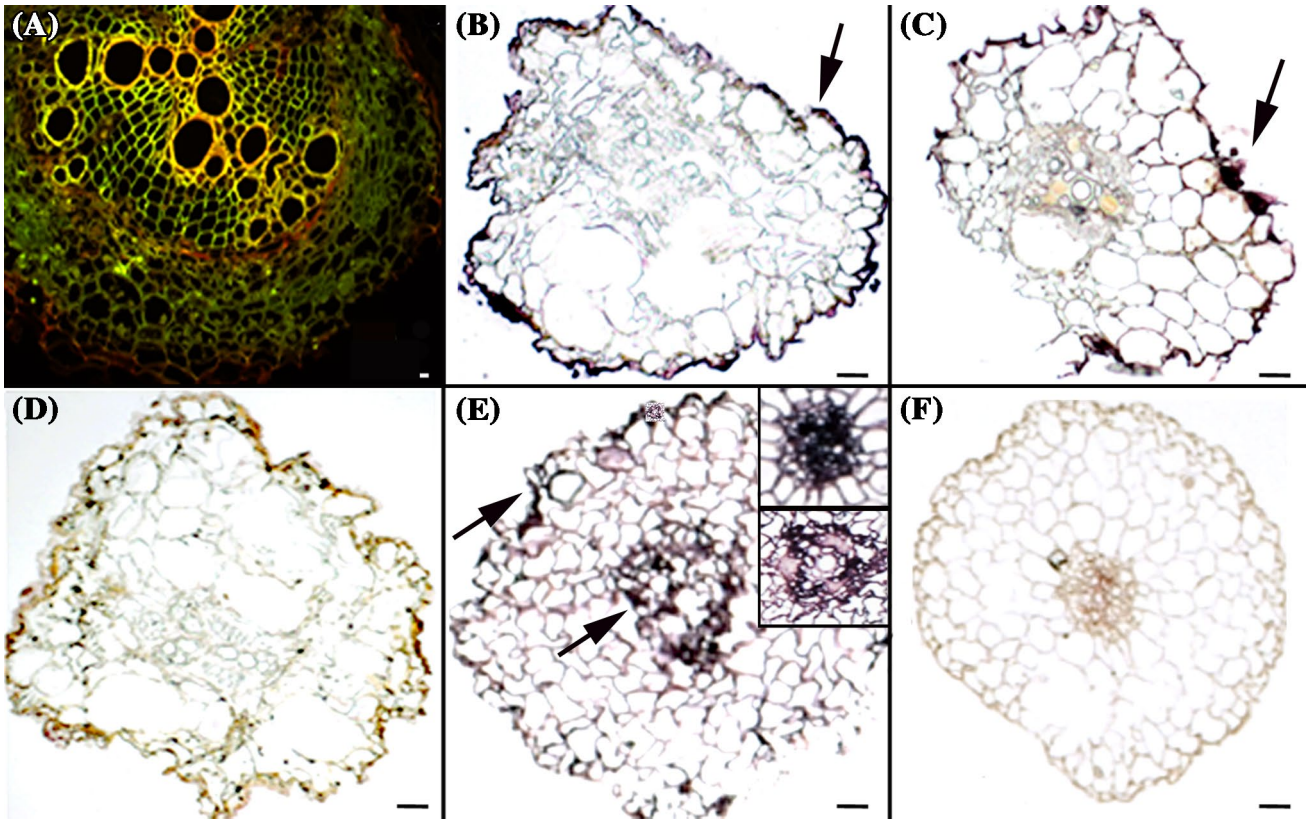



**Fig. 5** *AraEXLB8 in situ* hybridization. **a** Preserved RNA (*red*) after acridine orange treatment in root cells of a non hybridized RKN-inoculated *A. stenosperma* root (3 DAI). Drought-stressed roots of **b** *A. gregoryi* and **c** *A. villosa* with hybridization signals detected in cortical and epidermal root cells. **d** Well-watered (control) roots of *A. gregoryi*, with no hybridization signals. **e** RKN-inoculated roots showing evident signals in the vascular cylinder and surrounding endodermal cells, as well as some signals in epidermal cells, of *A. stenosperma* (3 DAI). **f** *A. stenosperma* inoculated roots (3 DAI) with no signals after T7 hybridization with the sense (T7) probe (technical control). ISH signals indicated by *arrows*. *Bars* = 20 µm

different transformation events (Fig. 6c–e; Supplementary Table 8).

## Discussion

Expansin induction as a preliminary step towards plant adaptation to environmental constraints and pathogen attack has been suggested as one of the mechanisms underlying plant resilience to overcome such stresses (Marowa et al. 2016). Despite their potential role in enhancing plant performance under adverse conditions, molecular and functional characterization of the expansin genes remains incomplete for most species. In this study, expansin superfamily organization, expression dynamics, and evolution was investigated in wild *Arachis* species known for harboring resistance to a large range of biotic and abiotic stresses (Bertioli et al. 2011). Their involvement in responses to drought, nematode infection, and UV treatment is further supported by current results.

### WGD contributed to the expansion of *Arachis* expansins

The number of expansin genes herein identified in *A. duranensis* (40) and *A. ipaënsis* (44) was consistent with that previously described for other legumes which share a genome duplication with *Arachis* at the estimated time of Papilionoid origin (Shirasawa et al. 2013), as *M. truncatula* (42), *P. vulgaris* (36) and *G. max* (75). The higher number of expansin genes in *G. max* may reflect its later lineage-specific duplication. *Arachis* expansins could be further classified into the four subfamilies as proposed by (Kende et al. 2004). However, while EXPA, EXPB and EXLA formed highly supported monophyletic groups, the monophyly of EXLB was not highly supported and EXLBs were separated into two groups interspersed by EXLAs in the main phylogenetic analysis (Fig. 1a). Interestingly, a similar nonsupport of EXLB monophyly and a separation in two subgroups was also observed in a recent Bayesian phylogeny including *Amborella, Arabidopsis* and *Oryza* (Seader et al. 2016). However, this separation of EXLBs in two groups is only supported by low bootstrap (Fig. 1a) or

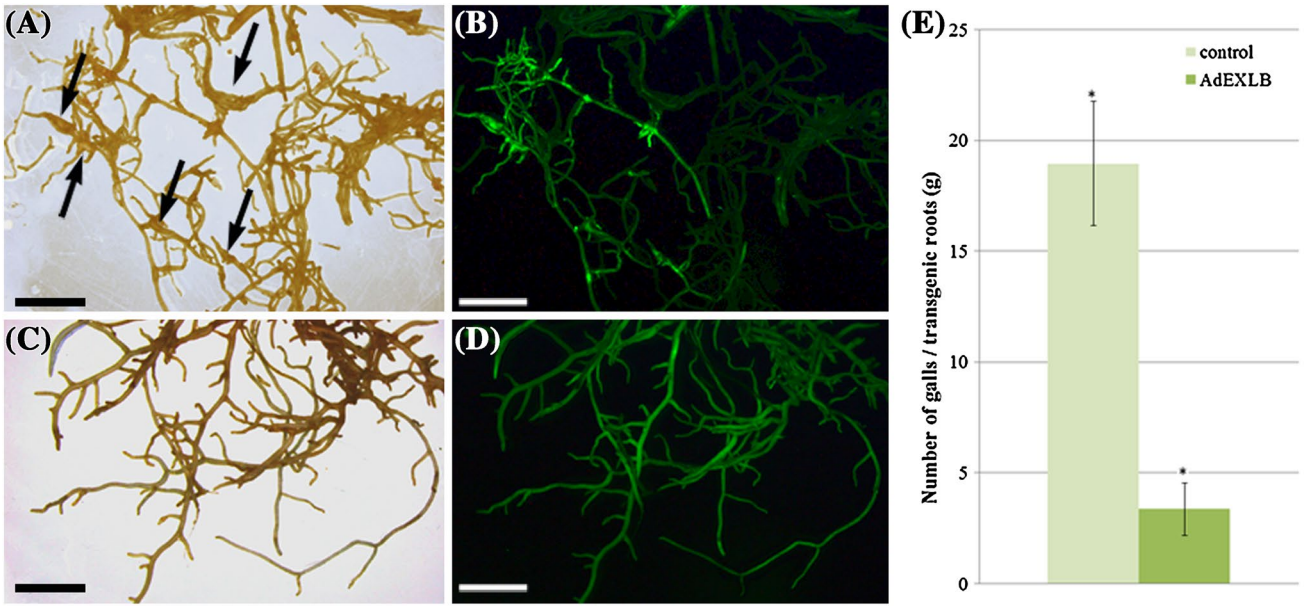



**Fig. 6** *Glycine max* hairy roots infected with *Meloidogyne javanica* at 60 DAI. **a, b** Hairy roots transformed with empty and **c, d** pPZP-AdEXLB8 binary vectors under **a** and **c** white light and **b** and **d** 480 ± 30 nm excitation GFP filter. **e** Number of galls per gram of hairy roots. Galls are indicated by *arrows*. *Bars* = 5 mm

posterior probability (Seader et al. 2016) values. Other phylogenies, including different species, found a monophyly of EXLBs and a separation between EXLAs and EXLBs (Dal Santo et al. 2013; Sampedro et al. 2005; Zhang et al. 2014b; Zhu et al. 2014). However, these topologies either received only moderate support for the monophyly of EXLB (e.g. Supplementary Figs. 1, 4b) or received higher support but either included less representatives of the EXLB family (Sampedro et al. 2005) or made use of less sophisticated phylogenetic methods, known to be more prone to biases such as long branch attraction (Dal Santo et al. 2013; Sampedro et al. 2006; Zhu et al. 2014). Overall, these different phylogenies cast doubt on the support for a monophyly (or polyphyly) of EXLBs, which depends on the taxon sampling and on the phylogenetic method used. However, all the phylogenies converge in supporting with high confidence values, the grouping of all EXLs together and the monophyly of all the other families.

The largely duplicated nature of plant genomes was no exception in *A. duranensis* and *A. ipaënsis*, with current whole genome analysis showing that the majority of protein-coding genes are duplicated, consistent with (Bertioli et al. 2016). The expansion of the *Arachis* expansin superfamily seemed to be mainly achieved by the retention of gene copies after whole genome duplication, as observed for *G. max* and *A. thaliana* (Zhu et al. 2014), or triplication, as for *Brassica rapa* (Krishnamurthy et al. 2014). Differently, in the monocot *Oryza sativa*, tandem duplication was the predominant contributor with a major influence in expansin superfamily expansion (Sampedro et al. 2005; Zhu et al. 2014). Hence, present results support previous evidence that the expansion of the plant expansin superfamily is lineage-specific, and is mostly due to whole genome duplications in eudicots and tandem duplication in monocots (Sampedro et al. 2006). Retention and diversification of gene copies after duplications is one of the main forces in the evolution of higher plants (Freeling 2009; Rodgers-Melnick et al. 2012) and could be associated with subfunctionalization or neofunctionalization of expansin gene duplicates in angiosperms, resulting, for example, in differences in cell wall composition of grasses compared with eudicots (Cosgrove 2015; Sampedro et al. 2015).

In the *Arachis* expansin subfamily, it seems that the abundance of EXPA genes, distributed in every chromosome of both *Arachis* species, might be a result of multiple duplication events, with WGD being the major contributor for their abundance. Conversely, the majority of EXPB gene copies derived from tandem duplications. These duplications were probably responsible for the large EXPB cluster observed in chromosomes 2 of both species. In general, tandem duplications are expected to have a lower impact on the genome than WGD that initially doubles the number of all genes. The EXLB subfamily appears to have expanded specifically in legume species, ranging from 4 to 15 members per genome, whilst in other eudicots, they range from 1 to 4 (Supplementary Table 3). Moreover, in monocots, EXLB duplication showed a different evolution pattern, lacking in *Zea mays* (Zhang et al.

2014c) and having only one representative in *O. sativa* (Sampedro et al. 2005). Herein, the results showed that 50% of the EXLB genes were assigned to the WGD category of duplication. Actually, the EXLB that are mainly concentrated on chromosome 1 are part of a long block of duplicated genes conserved with chromosome 5. The remnant of this WGD are the couples of duplicates *EXLB5* and *EXLB7* on chromosome 1 that correspond to *EXLB8* and *EXLB6* on chromosome 5. This observation holds true for both *Arachis* species. The following series of events for the expansion of EXLBs in *Arachis* can thus be hypothesized: first an ancestral series of proximal or tandem duplications gave rise to an ancestral cluster of EXLBs. Then following a WGD event all these EXLBs were initially duplicated and distributed to the ancestors of chromosome 1 and 5. Then, unbalanced gene loss led to preferential retention on chromosome 1. This phenomenon of genome dominance after WGD leading to more genes retained on one of the duplicated regions has been frequently observed in plants (Murat et al. 2014). Overall, the higher abundance of EXLB genes in *Arachis* suggests a selective advantage to retain these multiple copies after a series of duplication. This implies that the EXLB genes could be part of wild *Arachis* mechanisms that ensure successful adaptability in their native adverse environments (Rodgers-Melnick et al. 2012). Consistent with the possible recruitment of EXLB for new functions, a strong positive selection was identified between *AiEXLB5*/*AiEXLB8* gene copies, suggesting a functional divergence of this duplicated pair over the evolutionary history of (*A*) *ipaënsis* genome. The influence of positive selection in the expansin gene superfamily evolution has previously been reported for *G. max*, but not for other angiosperms such as (*B*) *rapa* or *Z. mays* (Krishnamurthy et al. 2014; Zhang et al. 2014c). This reinforces the present findings that *AraEXLB5* and *AraEXLB8* are stress-responsive genes (Brasileiro et al. 2015; Guimaraes et al. 2015), as positive selection is often observed in genes involved in defense/immunity and response to environmental constraints (Chen et al. 2010).

### *Arachis* expansin genes are widely modulated by stresses

Genome-wide expression analysis demonstrated that expansin genes, regardless of the subfamily, were broadly expressed in roots of *A. duranensis* and *A. stenosperma*, and in leaves of *A. stenosperma*, and were widely responsive to external stimuli, thus confirming their involvement in stress responses. This expansin ubiquitous expression in roots and leaves was previously observed in many plant species, including legumes (Liu et al. 2015; Zhu et al. 2014), which is consistent with their role in root and root hair development, water and nutrient uptake (Che et al. 2016; Cho and Cosgrove 2004) and leaf growth and stomatal plasticity (Lü et al. 2013; Zhang et al. 2011). Moreover, the recently published *A. hypogaea* expression atlas based on 22 different tissue types and ontogenies (Clevenger et al. 2016) showed that 36 homologs out of the 38 *A. duranensis* expansin genes are expressed (absolute expression above 1) in at least one of the tissues that represent the full development (vegetative, reproductive and seed) of peanut. Some of these expansin genes also displayed preferential tissue expression (Clevenger et al. 2016): two *AdEXPB* genes were exclusively expressed in roots or in reproductive shoot tips, whilst five *AdEXLB* transcripts were predominantly located in nodules. This indicates that, although expansins are implicated in a variety of plant processes, some *Arachis* genes might have evolved towards tissue-dependent functional specialization, as suggested for other species (Clevenger et al. 2016; Dal Santo et al. 2013; Lu et al. 2015; Zhang et al. 2014c). As the expression pattern of a gene is usually closely associated with its function, it is assumed that there are species-specific and/or stress-specific features in the biological functions of the *Arachis* expansin genes.

The involvement of expansins in adaptive responses to abiotic stresses has been widely reported and usually associated with the control of cell water loss and osmotic adjustment, probably by promoting cell wall extensibility, or with the general ROS-scavenging defense response to avoid oxidative damage (Marowa et al. 2016; Sasidharan et al. 2011). Here, the majority of the expansin genes were modulated according to the *Arachis* species and stress (drought and UV). Recently, Lu et al. (2015) also demonstrated that the responsiveness of *Solanum lycopersicum* expansins differs depending on the duration and type of stress imposed (drought, cold or heat). The combined evidence implies that different expansins could have distinct functions in response to abiotic stress via an ABA-independent pathway, as previously suggested (Cho and Cosgrove 2004).

Likewise, the overall expression levels of expansin transcripts upon a *M. arenaria* RKN infection revealed a general opposite profile in the RKN-resistant *A. stenosperma* and in the moderately RKN-resistant *A. duranensis*. This might reflect different mechanisms of resistance displayed by these species, which triggers the HR exclusively in *A. stenosperma* and an unknown defense response conferring partial resistance in *A. duranensis* (Proite et al. 2008). Further knowledge about the involvement of expansins in this parasitic association must, therefore, be pursued.

### *AraEXLB8* is a promising molecular marker

Among the *A. duranensis* expansin genes, *AdEXLB8* exhibited a distinct and highly specific expression profile in response to both stresses, corroborating its previously

association with *Arachis* drought tolerance (Brasileiro et al. 2015; Ding et al. 2014) and RKN resistance (Guimaraes et al. 2015; Tirumalaraju et al. 2011). *AraEXLB8* nucleotide characterization evidenced the synonymous SNP15 as a promising molecular marker since it occurred exclusively on tetraploid genotypes (natural and synthetic). As previously suggested by Barkley et al. (2007), this type of polymorphism can be further exploited to design new molecular markers to distinguish cultivated tetraploids and wild diploids species in botanical and evolutionary studies. The haplotypes established here based on the *AraEXLB8* nucleotide diversity indicated that they do not share ploidy level (diploidy and tetraploidy), plant life cycle (annual and perennial) or genome type (A, B, D and K), and do not necessarily reinforce phylogenetic proximities, as previously observed for *A. magna* and *A. ipaënsis* (Bechara et al. 2010; Krapovickas and Gregory 1994).

The *AraEXLB8* gene conservation based on cDNA sequences was higher (97.5%) than that observed in previous studies conducted with conserved regions or larger sampling of *Arachis* sequences (Bechara et al. 2010; Cunha et al. 2008; Moretzsohn et al. 2013), reinforcing the close phylogeny for the genotypes here observed. The high levels of conservation on *AraEXLB8* cDNA sequences might reflect *Arachis* reproductive habits, such as autogamy and geocarpy that limit its crossability and seed dispersion (Bertioli et al. 2011; Krapovickas and Gregory 1994).

### *AraEXLB8* is a promising candidate for plant genetic engineering

The cDNA substitution characterizing the haplotype 6 (H6) can be further associated with the *AraEXLB8* expression levels since the three H6 genotypes (*A. stenosperma, A. villosa*, and *A. gregoryi*) displayed the highest induction in response to drought. This overall induction of *AraEXLB8* is coincident with the progressive decline in transpiration previously observed in wild *Arachis* species during a dry-down assay (Leal-Bertioli et al. 2012). Given that expansins are involved in root water uptake and growth (Lü et al. 2013; Zhang et al. 2011), it is possible that this protein, via root-to-leaf signaling, is implicated in both drought response mechanisms triggered upon the perception in roots of soil water availability restriction. Moreover, *in situ* detection of the *AraEXLB8* transcripts indicates cell- and tissue-specific distribution, mostly in cortical and epidermal root cells, the primary site of water decrease perception. These cells might undergo morphology changes to avoid primary tissue damage by increasing expression of drought-responsive genes, such as *AraEXLB8*, involved in perception and signaling of this stress.

Several studies associate the *EXPA, EXPB* and *EXLA* genes with drought perception and adaptation (Marowa et al. 2016), but few reports have identified *EXLB* as drought-responsive genes. Previously, our group reported that *AraEXLB8* is highly upregulated in response to water deficit in roots and leaves of *A. duranensis* and *A. magna* (Brasileiro et al. 2015) with its induction considered to be a preliminary step toward drought acclimation through adjusting cell wall extensibility (Harb et al. 2010). The responsiveness of *AraEXLB8* orthologs to water deficit was also observed in roots of a drought-tolerant *A. hypogaea* genotype (Ding et al. 2014) and other legumes species, *M. truncatula* (Zhang et al. 2014a). Here, the increased expression levels of *AraEXLB8* in the 13 *Arachis* genotypes studied, particularly in those belonging to the H6 group, associated with its specific distribution in surface cells of the roots, suggest that *AraEXLB8* is involved in the network of molecular responses triggered by plants to circumvent dehydration damaging effects, making this gene a promising candidate for drought tolerance engineering.

Additional expression analysis in response to nematode infection in roots and UV treatment in leaves reinforce the prompt *AraEXLB8* responsiveness to other stresses, with different levels of expression displayed by the *Arachis* genotypes that harbor different levels of pathogens resistance. The upregulation of *AraEXLB8* orthologs in response to nematode infection was also reported in the *Arachis* RKN-resistant genotypes (Guimaraes et al. 2015; Tirumalaraju et al. 2011) as well as the downregulation of a *P. vulgaris EXLB* gene at the later stages of its compatible interaction with *M. incognita* (Santini et al. 2016). The shift in the spatial distribution of the *AraEXLB8* transcripts in RKN-inoculated *Arachis* roots, comparing to those detected in drought stressed, indicates the precise and stress-dependent regulation of this gene. The presence of *AraEXLB8* in some cells of vascular bundles reinforces its involvement in RKN infection response, as nematodes in the early parasitic stages (J2) are mainly located within the differentiating vascular bundle (Ozalvo et al. 2014). These results suggest that, regardless of the mechanism of resistance displayed, the *AraEXLB8* gene might have a more downstream role in the pathogen triggering response (PTI), and reinforces the need for a further understanding of molecular, biochemical and cellular events that culminate in pathogen resistance. The notable induction of *AraEXLB8* in *A. stenosperma* leaves in response to UV can be also associated with the general ROS-scavenging and signal transduction pathways that triggered changes in the expression of some genes, as expansins, and in the production of secondary metabolites, as resveratrol, to control numerous biological process, including pathogen defense (Baxter et al. 2014; Han et al. 2015; Lopes et al. 2013), and therefore, contributing to the HR resistance mechanism of *A. stenosperma*, as previously suggested (Guimaraes et al. 2010; Morgante et al. 2013). UV irradiation and other physical stresses induce

'cross-tolerance' responses that are commonly associated with pathogen defense and reactions to wounding, promoting increased resistance to pathogens (Mintoff et al. 2015).

*In planta* functional validation of *AdEXLB8* was carried out in *G. max* composite plants since *A. hypogaea* stable transformation is currently limited and time-consuming (Brasileiro et al. 2014; Kuma et al. 2015). This approach allowed the successful *AdEXLB8* overexpression in hairy roots, leading to a remarkable reduction in the number of galls produced by *M. javanica*. Like *M. arenaria*, the *M. javanica* RKN causes important damages in host plants, reducing drastically the productivity of many tropical crops, and it is particularly aggressive in *G. max*, the most important commodity in the world (Beneventi et al. 2013). Since each hairy root obtained is an independent transformation event, a relative high number of eGFP-positive events could be analyzed, reinforcing that *G. max* composite plants are an efficient and fast system to validate *Arachis* RKN resistance candidates (Kuma et al. 2015). Moreover, the selection of transgenic events by eGFP fluorescence allowed the assessment of RKN infection uniquely in *AdEXLB8*-expressing roots, increasing the accuracy and speed of the analysis.

The literature is contradicting in terms of expansin roles in plant–microbe interactions, as the suppression of specific expansin genes reduced the infection by either pathogens, including nematodes (Ding et al. 2008; Gal et al. 2006; Marowa et al. 2016), or symbionts, as arbuscular mycorrhizal fungus (Dermatsev et al. 2010), whilst our results suggest that the overexpression of *AraEXLB8* gene mediates RKN defense in *G. max* roots. The extent to which these non-enzymatic proteins act as physical barriers against microbes or as a source of signals to activate defense responses remains unclear, albeit their induction by auxin could be an underlying mechanism that should be considered (Sasidharan et al. 2011). The preferential expression of the *Arachis EXLB* genes in rhizobium-induced nodules (Clevenger et al. 2016) besides their overall induction upon RKN infection may indicated that the *EXLB* genes are common components in the establishment of both parasitism and symbiosis and seem to play an important role in plant–microbe interactions. In fact, besides expansins, it has been demonstrated that rhizobia and RKN interactions with legumes share a significant number of target genes and display similar molecular mechanisms of symbiosis and pathogenesis to facilitate successful colonization (Damiani et al. 2012).

The effectiveness of *AdEXLB8* overexpression to reduce a RKN infection in stable transgenic plants of soybean and other legumes will enable the assessment of its potential to foster crop improvement for the mitigation of both, abiotic and biotic stresses, which is of utmost importance in the face of the predicted climate changes.

## Materials and methods

### Identification of the *Arachis* expansin superfamily and phylogenetic analysis

The protein sequences of *A. thaliana* and *G. max* expansins (http://www.personal.psu.edu/fsl/ExpCentral) were used to retrieve *A. duranensis* and *A. ipaënsis* expansin sequences from genome assemblies available in PeanutBase (http://peanutbase.org/) Genome Threader software with default parameters (Gremme et al. 2005). Additional sequences were obtained from this database using "expansin" as keyword and TBLASTN searches (e-value 1e−3) against the genomes. The presence of the canonical expansin domains DPBB (PF03330) and CBM63 (PF01357) (Cosgrove 2015) was examined using the CDART tool (http://ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) and only sequences containing both of the domains were retained.

For phylogenetic analysis, all of the putative *Arachis* expansin protein sequences were aligned using MAFFT software, including sequences from *P. vulgaris, M. truncatula* and *G. max* (https://phytozome.jgi.doe.gov/pz/portal.html) and *Arabidopsis thaliana* (https://www.arabidopsis.org). To provide a more accurate alignment and construction of the phylogenetic tree, sequences with over 90% similarity were trimmed with trimAl (Capella-Gutiérrez et al. 2009). Phylogenic trees were constructed using the Maximum-Likelihood method as implemented in the RAxML software (Stamatakis 2006), with an automatic search of the fittest model and parameters as well as a bootstrap search automatically stopped upon congruence. The nomenclature of *A. duranensis* and *A. ipaënsis* expansins was given according to the synteny between the species and classification in subfamilies accordingly to Kende et al. (2004).

### Genomic distribution, duplication pattern, and structure of expansin genes

The location of the putative expansin genes in *A. duranensis* and *A. ipaënsis* chromosomes was obtained from gene model annotations (http://peanutbase.org/) and all predicted protein sequences were submitted to self and all-against-all BLASTP analyses (e-value 1e− 10). The expansins distribution and respective gene duplication patterns in chromosomes were analyzed by the MCScanX software with default parameters (Wang et al. 2012) and the results were formatted with custom perl script to Circos format (http://circos.ca/) for graphical representation. To detect positive selection between genes, the ratio of *Ka*/*Ks* between the gene copies of expansin sequences was computed using the Nei-Gojobori statistics (Nei and Gojobori 1986), as implemented in McScanX.

*Arachis* expansin gene exon/intron organization retrieved from PeanutBase (http://peanutbase.org/) was submitted to GSDS software (http://gsds.cbi.pku.edu.cn/). InterPro and SignalP were used for protein domain and signal peptide prediction, respectively, and the isoelectric points and molecular weights of the proteins obtained from ExPASy (http://expasy.org/).

### RNA-seq expression profiling of *Arachis* expansin genes

Illumina RNA-seq data previously obtained by our group from roots of *A. duranensis* and *A. stenosperma* plants submitted to drought (available at http://peanutbase.org) or inoculated with RKN *M. arenaria* (collected at 3, 6 or 9 days after inoculation; DAI) (Guimaraes et al. 2015) or from leaves of *A. stenosperma* submitted to UV exposure (unpublished results) were used for the expansin genes expression analysis. The genome of *A. duranensis* was used to create an index with the Kallisto program (version 0.42.3) and only high quality reads from the transcriptome were then used for the quantification against the created index, with default settings (Bray et al. 2016). For the differential expression analysis, the edgeR package (Robinson et al. 2010) was employed. The log2 of fold-change value between control and treated samples in both species was used for the heatmap generation using heatmap2R package (Warnes et al. 2015).

### *AraEXLB8* diversity in *Arachis* genotypes

Total RNA was isolated from young leaves of 13 *Arachis* genotypes (Supplementary Table 9) and the corresponding cDNAs used for RT-PCR as described by Morgante et al. (2011), employing *AraEXLB8*-flanking primers (EXPF 5′-AGTGCTTCATCTCAAAATGGAACT-3′ and EXPR 5′-TCTATTCAAGCTGGATTTGAGTATCA-3′). Sequenced amplification products were aligned using the Sequencher software (Gene Codes, Ann Arbor, USA) with three additional available *AraEXLB8* sequences from *A. duranensis* and *A. ipaënsis* (http://peanutbase.org/gbrowse) and *A. hypogaea* 'Tifrunner' [GO326087.1; (http://www.ncbi.nlm.nih.gov)]. Nucleotide substitutions were defined by BLOSUM62 matrix (Henikoff and Henikoff 1992) and hydrophobicity (Kyte and Doolittle 1982) while DNA sequence variation by DnaSP software (http://www.ub.es/dnasp/).

Phylogenetic analysis was conducted using the above 16 *AraEXLB8* coding sequences and four additional outgroups (*G. max* GI:955303815; *P. vulgaris* GI:593787101; *C. arietinum* GI:828330493 and *M. truncatula* Medtr5g013440.1). The phylogeny was constructed by Bayesian Inference using MrBayes program (http://mrbayes.sourceforge.net/) with default parameters and performed with 20 million generations. The phylogenetic model (GTR + I) was selected as the fittest model according to ModelTest and the consensus trees were inspected in FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

### *AraEXLB8* qRT-PCR analysis and in situ transcripts detection

#### *Dry-down assay*

Normalized transpiration rates (NTR) were determined for 4-week-old plants of 13 Arachis genotypes (Supplementary Table 9) throughout a soil gradual water deficit assay (dry-down), as previously described (Leal-Bertioli et al. 2012). Roots were collected when each drought stressed (DS) individual reached a NTR value around 0.3 relative to well-watered (WW) controls. Samples from three individuals were assembled forming three independent biological replicates for each treatment (WW and DS), and per genotype.

#### *RKN inoculation assay*

*Arachis stenosperma, A. duranensis* and *A. hypogaea* 'Runner' 4-week-old plants (Supplementary Table 9) were inoculated with *M. arenaria* race 1 at juvenile stage 2 (J2) as described by Morgante et al. (2013). Root samples from five individuals were collected at zero (control), 3 DAI (stressed), and assembled to form two independent biological replicates per treatment and genotype.

#### *UV assay*

Young leaves were collected from 7-week-old plants of *A. stenosperma, A. duranensis* and *A. hypogaea* 'Runner' (Supplementary Table 9) and submitted to UV-C light treatment for 2 h and 30 min, essentially as described by Lopes et al. (2013). Samples from five individuals were collected before (control) and 24 h after (stressed) UV treatment and assembled to form three independent biological replicates for each treatment and genotype.

#### *Quantitative RT-PCR (qRT-PCR) analysis*

Total RNA was extracted from samples and the cDNA synthesized as previously described (Morgante et al. 2011). The qRT-PCR reactions were conducted in a 7300 Real-Time PCR System (Applied Biosystem) using primers of the *A. duranensis ELXB8* (*AdEXLB8*) gene, as previously described (Brasileiro et al. 2015). In accordance to Morgante et al. (2011), ACT1 and UBI2 were used as reference genes for samples subjected to dry-down whereas 60S and GAPDH to RKN inoculation and UV treatments. The relative quantification (RQ) of transcripts was determined as

described by Brasileiro et al. (2015) and statistically tested using one-way ANOVA followed by Tukey's test.

*In situ hybridization (ISH) analysis*

Root fragments were isolated and processed in accordance to (Morgante et al. 2013). A 659 bp cDNA sequence of *AdEXLB8* was amplified (EXPuniF: 5′-ACTGCCAGTCAC TTGGAACC-3′ and Exp464R: 5′-AGATCCATTCCGCCA TAGC-3′), cloned into pGEM®-T Easy (Promega, Madison, USA) and used to produce dig-labeled RNA probes (DIG RNA Labeling Kit SP6/T7, Roche). Slides were then hybridized with 2 ng of probe and hybridization sites were immunocytochemically detected as described (Morgante et al. 2013).

### *Meloidogyne javanica* bioassay in *G. max* composite plants overexpressing *AdEXLB8* gene

The coding region (753 bp) of *AdEXLB8* was cloned (Epoch Life Science, Texas, USA) under the control of the *A. thaliana* actin 2 promoter, at the *XhoI* restriction site of the binary vector pPZP-201BK-EGFP (Chu et al. 2014) containing an enhanced Green Fluorescent Protein (eGFP) to form the pPZP-AdEXLB8 vector. *Agrobacterium rhizogenes* strain 'K599', harboring pPZP-201BK-EGFP (empty) or pPZP-AdEXLB8, vectors was then used to generate *G. max* 'Williams 82' composite plants employing fresh bacterial paste inoculum (Chu et al. 2014), basically as described before (Kuma et al. 2015). Transgenic roots were identified one week later through eGFP fluorescence under GFP1 filter set on a M205 stereomicroscope (Leica Microsystems, Wetzlar, Germany) and the eGFP-negative roots were excised before plant transfer to a 3:1 sand:plant gel mixture (v:v) and acclimation in greenhouse conditions. Three-week-old composite plants were challenged with approximately 1000 J2 of *M. javanica* and the number of galls determined at 60 DAI. The qRT-PCR analysis of *AdEXLB8* expression in eGFP-positive roots was conducted as described above. Outliers in galls counting data were removed by Grubbs' test and means compared by unpaired t-test using Graphpad software (http://graphpad.com/quickcalcs/ttest1/).

**Acknowledgements** We would like to thank Drs. Marília Santos, Márcio Moretzsohn, Roberto Togawa, Marcos Mota, and Regina Carneiro (Embrapa Cenargen; Brazil) and MSc. Paula Vasconcelos (UNESP; Brazil) for helpful contribution on bioassays and data analysis and Peggy Ozias-Akins (University of Georgia, USA) for providing pPZP-201BK-EGFP vector. The authors gratefully acknowledge support from 'Empresa Brasileira de Pesquisa Agropecuária' (Embrapa); 'Conselho Nacional de Desenvolvimento Científico e Tecnológico' (CNPq); and 'Fundação de Apoio à Pesquisa do DF' (FAPDF). EGJD is a recipient of a CNPq/PVE (Special Guest Researcher) grant from the Brazilian program 'Science without Borders'; LAG is a recipient of fellowships from CNPq and APZM, BMP and RBS from 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES).

**Author Contributions** Conceived and designed the experiments: LAG, ACGA, PMG and ACMB; Performed the experiments: BMP, MAPS and RBS; Analyzed the data: LAG, APZM, ACGA, LFAF and ACMB; Wrote the paper: LAG, APZM, ACGA, LFAF, EGJD, PMG and ACMB; Contributed reagents/materials/analysis tools: ACGA, EGJD, PMG and ACMB; Critically revised the manuscript: All authors.

**Compliance with ethical standards**

**Conflict of interest** The authors have declared that no competing interests exist.



## References

Barkley NA, Dean RE, Pittman RN, Wang ML, Holbrook CC, Pederson GA (2007) Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing. Genet Res 89:93–106. doi:10.1017/S0016672307008695

Baxter A, Mittler R, Suzuki N (2014) ROS as key players in plant stress signalling. J Exp Bot 65:1229–1240

Bechara MD et al (2010) Phylogenetic relationships in genus *Arachis* based on ITS and 5.8 S rDNA sequences. BMC Plant Biol 10:255–255

Beneventi M et al (2013) Transcription profile of soybean-root-knot nematode interaction reveals a key role of phythormones in the resistance reaction. BMC Genomics 14:322

Bertioli DJ, Seijo G, Freitas FO, Valls JFM, Leal-Bertioli SCM, Moretzsohn MC (2011) An overview of peanut and its wild relatives. Plant Genet Resour 9:134–149 doi:10.1017/S1479262110000444

Bertioli DJ et al. (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48:438

Brasileiro ACM, Araujo ACG, Leal-Bertioli SC, Guimaraes PM (2014) Genomics and genetic transformation in *Arachis*. Int J Plant Biol Res 2:1017

Brasileiro ACM et al. (2015) Transcriptome profiling of wild *Arachis* from water-limited environments uncovers drought tolerance candidate genes. Plant Mol Biol Rep. doi:10.1007/s11105-015-0882-x

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973

Che J, Yamaji N, Shen RF, Ma JF (2016) An Al-inducible expansin gene, OsEXPA10 is involved in root cell elongation of rice. Plant J 88:132. doi:10.1111/tpj.13237

Chen Q, Han Z, Jiang H, Tian D, Yang S (2010) Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives. J Mol Evol 70:137–148

Chen X et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci USA. doi:10.1073/pnas.1600899113

Cho HT, Cosgrove DJ (2004) Expansins as agents in hormone action. In: Davies PJ (ed) Plant hormones. 3rd edn. Kluwer Academic, Dordrecht, pp 262–281

Chu Y, Guimaraes L, Wu C, Timper P, Holbrook C, Ozias-Akins P (2014) A technique to study *Meloidogyne arenaria* resistance in *Agrobacterium rhizogenes* -transformed peanut. Plant Dis 98:1292–1299

Clevenger JPC, Chu Y, Scheffler B, Ozias-Akins P (2016) A developmental transcriptome map for allotetraploid *Arachis hypogaea*. Front Plant Sci 7:1446

Cosgrove DJ (2015) Plant expansins: diversity and interactions with plant cell walls. Curr Opin Plant Biol 25:162–172

Cosgrove DJ (2016) Catalysts of plant cell wall loosening. F1000Research 5:F1000 Faculty Rev-1119. doi:10.12688/f1000research.7180.1

Cunha FB, Nobile PM, Hoshino AA, Moretzsohn MdC, Lopes CR, Gimenes MA (2008) Genetic relationships among *Arachis hypogaea* L. (AABB) diploid Arachis species with AA BB genomes. Genet Resour Crop Evol 55:15–20. doi:10.1007/s10722-007-9209-6

Dal Santo S, Vannozzi A, Tornielli GB, Fasoli M, Venturini L, Pezzotti M, Zenoni S (2013) Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. PLoS ONE 8:e62206

Damiani I et al (2012) Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. New Phytol 194:511–522

Dermatsev V et al (2010) Microarray analysis and functional tests suggest the involvement of expansins in the early stages of symbiosis of the arbuscular mycorrhizal fungus *Glomus intraradices* on tomato (*Solanum lycopersicum*). Mol Plant Pathol 11:121–135

Ding X, Cao Y, Huang L, Zhao J, Xu C, Li X, Wang S (2008) Activation of the indole-3-acetic acid–amido synthetase GH3-8 suppresses expansin expression and promotes salicylate-and jasmonate-independent basal immunity in rice. Plant Cell 20:228–240

Ding H, Zhang ZM, Qin FF, Dai LX, Li CJ, Ci DW, Song WW (2014) Isolation and characterization of drought-responsive genes from peanut roots by suppression subtractive hybridization.Electron J Biotechnol 17:304–310

Ding A, Marowa P, Kong Y (2016) Genome-wide identification of the expansin gene family in tobacco (*Nicotiana tabacum*). Mol Genet Genomics 291:1891–1907

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 60:433–453. doi:10.1146/annurev.arplant.043008.092122

Gal T, Aussenberg E, Burdman S, Kapulnik Y, Koltai H (2006) Expression of a plant expansin is involved in the establishment of root knot nematode parasitism in tomato. Planta 224:155–162. doi:10.1007/s00425-005-0204-x

Gremme G, Brendel V, Sparks ME, Kurtz S (2005) Engineering a software tool for gene structure prediction in higher organisms. Inf Softw Technol 47:965–978

Guimaraes PM et al. (2010) A study of gene expression in the nematode resistant wild peanut relative, *Arachis stenosperma*, in response to challenge with *Meloidogyne arenaria*. Trop Plant Biol. doi:10.1007/s12042-010-9056-z

Guimaraes PM et al (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genomics 13:387

Guimaraes PM et al (2015) Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS ONE 10:e0140937

Han Y, Chen Y, Yin S, Zhang M, Wang W (2015) Over-expression of TaEXPB23, a wheat expansin gene, improves oxidative stress tolerance in transgenic tobacco plants. J Plant Physiol 173:62–71

Harb A, Krishnan A, Ambavaram MM, Pereira A (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. Plant Physiol 154:1254–1271

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919

Janila P et al. (2016) Genomic tools in groundnut breeding program: status and perspectives Front Plant Sci 7:289

Kende H et al (2004) Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol 55:311–314

Krapovickas A, Gregory W (1994) Taxonomia del genero *Arachis* (Leguminosae). Bonplandia 8:1–186

Krishnamurthy P, Hong JK, Kim JA, Jeong M-J, Lee Y-H, Lee SI (2014) Genome-wide analysis of the expansin gene superfamily reveals *Brassica rapa* -specific evolutionary dynamics upon whole genome triplication. Mol Genet Genomics 290:521–530

Kuma KM et al (2015) A high efficient protocol for soybean root transformation by *Agrobacterium rhizogenes* and most stable reference genes for RT-qPCR analysis. Plant Cell Rep 34:1987–2000. doi:10.1007/s00299-015-1845-2

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Leal-Bertioli SCM et al (2012) The effect of tetraploidization of wild *Arachis* on leaf morphology and other drought-related traits. Environ Exp Bot 84:17–24

Liu Y, Zhang J, Li W, Guo C, Shu Y (2015) In silico identification, phylogeny and expression analysis of expansin superfamily in *Medicago truncatula*. Biotechnol Biotechnol Equip 30:1–7

Lopes RM, Silveira D, Gimenes MA, Vasconcelos PAS, Alves RBN, Silva JP, Agostini-Costa TS (2013) Characterization of resveratrol content in ten wild species of section *Arachis*, genus *Arachis*. Genet Resour Crop Evol 60:2219–2226

Lu Y, Liu L, Wang X, Han Z, Ouyang B, Zhang J, Li H (2015) Genome-wide identification and expression analysis of the expansin gene family in tomato. Mol Genet Genomics. doi:10.1007/s00438-015-1133-4

Lü P, Kang M, Jiang X, Dai F, Gao J, Zhang C (2013) RhEXPA4, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to Arabidopsis. Planta. doi:10.1007/s00425-013-1867-3

Marowa P, Ding A, Kong Y (2016) Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep. doi:10.1007/s00299-016-1948-4

McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall extension in plants. Plant Cell 4:1425–1433. doi:10.1105/tpc.4.11.1425

Mintoff S, Rookes J, Cahill D (2015) Sub-lethal UV-C radiation induces callose, hydrogen peroxide and defence-related gene expression in *Arabidopsis thaliana*. Plant Biol 17:703–711

Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111:113–126. doi:10.1093/aob/mcs237

Morgante C, Guimaraes P, Martins A, Araujo A, Leal-Bertioli S, Bertioli D, Brasileiro A (2011) Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. BMC Res Notes 4:339

Morgante CV et al (2013) A survey of genes involved in *Arachis stenosperma* resistance to *Meloidogyne arenaria* race 1. Funct Plant Biol 40:1298–1309

Murat F et al. (2014) Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16 K protogenes. Genome Biol Evol 6:12–33

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ozalvo R et al (2014) Two closely related members of *Arabidopsis* 13-lipoxygenases (13-LOXs), LOX3 and LOX4, reveal distinct functions in response to plant-parasitic nematode infection. Mol Plant Pathol 15:319–332

Proite K, Carneiro R, Falcao R, Gomes A, Leal-Bertioli S, Guimaraes P, Bertioli D (2008) Post-infection development and histopathology of *Meloidogyne arenaria* race 1 on *Arachis* spp. Plant Pathol 57:974–980. doi:10.1111/j.1365-3059.2008.01861.x

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Rodgers-Melnick E et al (2012) Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. Genome Res 22:95–105. doi:10.1101/gr.125146.111

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242

Sampedro J, Lee Y, Carey RE, DePamphilis C, Cosgrove DJ (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. Plant J 44:409–419. doi:10.1111/j.1365-313X.2005.02540.x

Sampedro J, Carey RE, Cosgrove DJ (2006) Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. J Plant Res 119:11–21

Sampedro J, Guttman M, Li L-C, Cosgrove DJ (2015) Evolutionary divergence of β-expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. Plant J 81:108–120. doi:10.1111/tpj.12715

Santini L, Munhoz CdF, Bonfim MF Jr, Brandão MM, Inomoto MM, Vieira MLC (2016) Host transcriptional profiling at early and later stages of the compatible interaction between *Phaseolus vulgaris* and *Meloidogyne incognita*. Phytopathology 106:282–294

Sasidharan R, Voesenek LA, Pierik R (2011) Cell wall modifying proteins mediate plant acclimatization to biotic and abiotic stresses. Crit Rev Plant Sci 30:548–562

Seader VH, Thornsberry JM, Carey RE (2016) Utility of the *Amborella trichopoda* expansin superfamily in elucidating the history of angiosperm expansins. J Plant Res 129:199–207

Shirasawa K et al (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of arachis and divergence of the legume genomes. DNA Res 20:173–184. doi:10.1093/dnares/dss042

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Tirumalaraju SV, Jain M, Gallo M (2011) Differential gene expression in roots of nematode-resistant and -susceptible peanut (*Arachis hypogaea*) cultivars in response to early stages of peanut root-knot nematode (*Meloidogyne arenaria*) parasitization. J Plant Physiol 168:481–492

Wang Y et al (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49–e49

Warnes G, Bolker B, Bonebakker L, Gentleman R, Liaw W, Lumley T (2015) gplots: Various R Programming Tools for Plotting Data. R package version 2.16. 0. 2015.

Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ (2006) Crystal structure and activities of EXPB1 (Zea m 1), a β-expansin and group-1 pollen allergen from maize. Proc Natl Acad Sci USA 103:14664–14671

Zhang X-Q, Wei P-C, Xiong Y-M, Yang Y, Chen J, Wang X-C (2011) Overexpression of the Arabidopsis α-expansin gene AtEXPA1 accelerates stomatal opening by decreasing the volumetric elastic modulus. Plant Cell Rep 30:27–36

Zhang J-Y et al (2014a) Global reprogramming of transcription and metabolism in Medicago truncatula during progressive drought and after rewatering. Plant Cell Environ 37:2553–2576

Zhang S, Xu R, Gao Z, Chen C, Jiang Z, Shu H (2014b) A genome-wide analysis of the expansin genes in *Malus* × *Domestica*. Mol Genet Genomics 289:225–236

Zhang W et al (2014c) Genome-wide identification and characterization of maize expansin genes expressed in endosperm. Mol Genet Genomics 289:1061–1074

Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, Hu Y (2014) Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biol 14:93

## 3. CONCLUSIONS

Several studies have characterized genes responding to individual stresses in plants. Nonetheless, considering the predicted climate changes and its impact on plant resistance, there is a strong need to identify genes responding to multiple stresses simultaneously, in order to produce more adapted crops (Mittler, 2006). This challenge can be met by the exploitation of cultivated species wild relatives. These wild relatives are genetically diverse and were selected during evolution, to resist to a range of abiotic and biotic stresses.

In the present study, the peanut wild relative *A. stenosperma* was used as source of resistance genes against root knot nematodes (RKN) and drought. The analysis of its transcriptome under drought (Chapter I), nematode infection (Chapter I) and both stresses simultaneously (Chapter II), led to the identification of several candidate genes to *Meloidogyne* spp. resistance and drought tolerance (Chapter III).

In Chapter I, universally conserved genes responsive to root-knot nematodes were identified, across four species, A. *stenosperma*, *C. arabica*, *G. max* and *O. glaberrima*, all inoculated with different species of RKN, which allowed the identification of 17 commonly regulated orthogroups. Among these orthogroups the majority of the genes were assigned to receptor kinases, which are known as the first layer of defense in plants (Tang et al., 2017). In addition, genes assigned to cell wall proteins and modifiers, seemed to be conserved among evolutionary distant plant species. Also in this chapter (Chapter I), the comparison of the transcriptome of two contrasting *Arachis* species for RKN resistance, *A. stenosperma* highly resistant to *M. arenaria*, and *A. duranensis* moderately susceptible led to the identification of key genes in the resistance to this RKN, including those from the NBS-LRR family, which are strongly regulated in the resistant species. Exclusive genes from *A. stenosperma* were found to be next to important QTL for nematode resistance and appear to be physically close, forming a cluster of genes. This cluster formation was previously described, and are often required for an effective resistance in plants (Nieri et al., 2017; Sinapidou et al., 2004). Also, the transcriptome of these two species was compared in response to drought stress. *A. duranensis* is considered a more drought-tolerant species while *A. stenosperma* is a sensitive species (Leal-Bertioli et al., 2012). Many differentially expressed genes (DEGs) in *A. duranensis* were identified, disclosing important genes in response to drought stress.

In Chapter II the molecular responses of *A. stenosperma* submitted to both drought and RKN simultaneously were analyzed, and we demonstrated that the responses to individual stresses is not the same as the response to combined stresses, trigging specific transcriptomic responses. We also showed that the drought stress is dominant over the biotic response and that the combined treatment showed a great transcriptional reprograming, with a general upregulation of genes from the ethylene and abscisic acid pathways and a downregulation of genes from the jasmonic acid pathway. These results are in accordance with the central role of the ABA hormone, with ERF and MYC as central transcription factors in the response to combined response (Bostock et al., 2014).

The use of innovative approaches in the present study proved to be a powerful tool for the prospection of stress-related genes. The pipelines generated by this study, as well as the databases of orthologs genes from several species, and the transcriptome of RKN-resistant species *A. stenosperma,* will be used in future projects and will also be available to the scientific community. Also, the present study allowed the identification of core conserved genes across different species and multiple stresses which can be exploited for other plant-pathogen interactions.

The characterization of three important plant gene families related to defense, NBS-LRR, dehydrin and expansin added knowledge in their function and response to biotic and abiotic stresses.

Ultimately, the production of transgenic plants overexpressing the candidate genes found in this study, have the potential to enhance the effectiveness and durability of the resistance/tolerance of important crops to drought and RKN infection.

## ARTIGOS RELACIONADOS À TESE

**Comparative root transcriptome of wild *Arachis* reveals NBS-LRR genes related to nematode resistance**

Published in BMC Plant Biology - 2018

Ana Paula Zotta Mota, Bruna Vidigal, Etienne G.J. Danchin, Roberto Coiti Togawa, Soraya C. M. Leal-Bertioli, David John Bertioli, Ana Claudia Guerra Araujo, Ana Cristina Miranda Brasileiro, Patricia Messenberg Guimaraes

Mota et al. BMC Plant Biology (2018) 18:159
https://doi.org/10.1186/s12870-018-1373-7

BMC Plant Biology

RESEARCH ARTICLE Open Access

# Comparative root transcriptome of wild *Arachis* reveals NBS-LRR genes related to nematode resistance

Ana Paula Zotta Mota[1,2], Bruna Vidigal[1], Etienne G. J. Danchin[3], Roberto Coiti Togawa[1], Soraya C. M. Leal-Bertioli[4], David John Bertioli[4], Ana Claudia Guerra Araujo[1], Ana Cristina Miranda Brasileiro[1] and Patricia Messenberg Guimaraes[1*]

## Abstract

**Background:** The Root-Knot Nematode (RKN), *Meloidogyne arenaria*, significantly reduces peanut grain quality and yield worldwide. Whilst the cultivated species has low levels of resistance to RKN and other pests and diseases, peanut wild relatives (*Arachis* spp.) show rich genetic diversity and harbor high levels of resistance to many pathogens and environmental constraints. Comparative transcriptome analysis can be applied to identify candidate resistance genes.

**Results:** Transcriptome analysis during the early stages of RKN infection of two peanut wild relatives, the highly RKN resistant *Arachis stenosperma* and the moderately susceptible *A. duranensis*, revealed genes related to plant immunity with contrasting expression profiles. These included genes involved in hormone signaling and secondary metabolites production and also members of the NBS-LRR class of plant disease resistance (R) genes. From 345 NBS-LRRs identified in *A.duranensis* reference genome, 52 were differentially expressed between inoculated and control samples, with the majority occurring in physical clusters unevenly distributed on eight chromosomes with preferential tandem duplication. The majority of these NBS-LRR genes showed contrasting expression behaviour between *A. duranensis* and *A. stenosperma*, particularly at 6 days after nematode inoculation, coinciding with the onset of the Hypersensitive Response in the resistant species. The physical clustering of some of these NBS-LRR genes correlated with their expression patterns in the contrasting genotypes. Four NBS-LRR genes exclusively expressed in *A. stenosperma* are located within clusters on chromosome Aradu. A09, which harbors a QTL for RKN resistance, suggesting a functional role for their physical arrangement and their potential involvement in this defense response.

**Conclusion:** The identification of functional novel R genes in wild *Arachis* species responsible for triggering effective defense cascades can contribute to the crop genetic improvement and enhance peanut resilience to RKN.

**Keywords:** *Meloidogyne*, resistance genes, peanut, Root Knot Nematode, transcriptome

* Correspondence: patricia.guimaraes@embrapa.br
[1]EMBRAPA Genetic Resources and Biotechnology, Brasilia, DF, Brazil
Full list of author information is available at the end of the article

## Background

Root-Knot Nematode (RKN), *Meloidogyne arenaria*, affects peanut production in the US, Africa and Asia and can result in significant yield losses [1]. Whilst the cultivated species (*Arachis hypogaea*) has low levels of RKN resistance, peanut wild relatives (*Arachis* spp.) show rich genetic diversity harboring high levels of resistance to many pathogens and environmental constraints [2–5]. All *A. hypogaea* cultivars with improved RKN resistance were developed through the introgression of two segments of a single chromosome from the wild relative *A. cardenasii* [6–9], thus making the identification of additional resistance sources critical to avoid resistance breakdown and assure breeding advances.

The wild species *A. stenosperma* harbors high levels of resistance against the peanut RKN *M. arenaria* and various foliar fungi [10–14]. Overall, the penetration and development of the RKN in the resistant species was reduced in comparison to the susceptible, with dark blue cytoplasm and altered organelle structures observed in the central cylinder, indicating a hypersensitive-like response (HR). In the moderately susceptible *A. duranensis*, the nematode reproduction occurs, albeit at lower levels and with a development delay when compared to the susceptible *A. hypogaea* [12].

In response to RKN infection, our previous studies showed that *A. stenosperma* bares a mechanism of resistance known as the Hypersensitive Response (HR) [15, 16], which is often triggered by Resistance genes (R) [17]. In addition, the recent identification of four QTLs in *A. stenosperma* reducing RKN galling and egg production [12], reinforces the importance of this species as a new source of resistance.

Plant R genes are key to many plant-pathogen interactions, as they enable plants to recognize pathogens and activate inducible defenses which often culminate in rapid HR response [18]. The vast majority of plant R genes are NBS-LRR, as they encode proteins with an amino-terminal variable domain, a central Nucleotide Binding Site (NBS) and a carboxy-terminal Leucine Rich Repeats (LRR) domain [19]. Both classes of NBS-LRR genes (TIR-type and CC-type/non-TIR) are commonly present in multigene clusters in plant genomes and can occur as true alleles across naturally variant genetic backgrounds [17]. Although many plant genomes have been sequenced and thousands of putative R genes, the Resistance Gene Analogs (RGAs) identified, only a relatively small number of R genes associated with nematodes resistance have been isolated and fully characterized [20, 21].

In *Arachis*, the first survey of RGAs using degenerate primers targeting the NBS domain revealed 78 NBS-LRR encoding sequences with unknown function [22]. Later, hundreds of RGAs were isolated from different peanut cultivars using the same strategy and genome BAC sequencing [23–25]. More recently, a genome-wide analysis of NBS-LRR genes in the peanut progenitor wild species, *A. duranensis* and *A. ipaënsis*, identified over 300 representatives classified in four NBS-LRR family types [26, 27]. However, only a relatively small number of NBS-LRR involved in the responses to pathogens has been unveiled [27, 28]. The analysis of *A. stenosperma* transcriptome identified several candidate genes involved in the defense signaling and response at the early stages of its incompatible interaction with RKN, including NBS-LRR genes [16], but the lack of a reference genome hampered the accurate identification and characterization of members of NBS-LRR and other complex gene families. The recent availability of the sequenced genomes of *A. duranensis* and *A. ipaënsis* [26], has greatly facilitated genome-wide studies in the genus [27, 29–31]. Nonetheless, no studies have yet contemplated the set of NBS-LRR genes expressed upon nematode infection.

The comparative transcriptome analysis of resistant and susceptible genotypes to different stresses has provided new insights into plant response mechanisms and can identify new sets of candidate resistance genes [32–38]. This approach has not yet been explored in wild *Arachis* species, but can be particularly fruitful, as these species have relatively limited transcriptome data available, still lack publicly available microarrays, reference transcriptomes and comprehensive transcripts datasets.

In this study, we investigated the expression profiles of candidate resistance genes to *M. arenaria* in two wild *Arachis* species with contrasting responses, the highly resistant *A. stenosperma* and the moderately susceptible *A. duranensis*, with the focus on the NBS-LRR class of R genes. Considering the narrow genetic base of peanut and the single RKN resistance source available in the crop cultivars, the identification and characterization of new resistance genes in wild genotypes will substantially contribute to expand the repertoire of resistances and secure their durability.

## Methods

### Plant material and Illumina sequencing

*Arachis duranensis* (accession K7988) roots challenged with *M. arenaria* race 1 were obtained as previously described [14], using four-week-old plants inoculated with 20,000 *M. arenaria* juveniles (J2). RNA was extracted from whole roots collected at 3, 6, and 9 days after inoculation (DAI) and non-inoculated plants as a control (Ctrl), using a modified lithium chloride protocol [39], and purified with Invisorb Plant RNA Mini Kit (Invitek, Berlin, Germany). Two independent biological replicates were produced by pooling equal amounts of total RNA per collecting point and cDNA was produced using the Super Script II enzyme and oligo (dT) 20 primer (Invitrogen, Carlsbad, CA, USA), according to manufacturer's

instructions. Eight paired-end libraries were constructed corresponding to the control and each of the three time points of the interaction: *A. duranensis* (DCtrl), 3DAI (DN3), 6DAI (DN6) and 9DAI (DN9) in biological duplicates. Libraries were sequenced in Hi-Seq 2000 at FASTERIS (www.fasteris.com), employing the mRNA-Seq and TruSeq (TM) SBS v5 protocols (Illumina, San Diego, CA). For *A. stenosperma* (accession V10309), we used Illumina transcript reads produced earlier using the same above conditions [16].

### Gene expression analysis

Illumina raw reads from *A. duranensis* and *A. stenosperma* were trimmed by Trimmomatic version 0.33 [40] and their quality checked by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Cleaned high quality reads were mapped to the annotated reference genome of *A. duranensis* (accession V14167) [26] (https://peanutbase.org) using the default settings of GMAP/GSNAP package [41]. The reads from each species were counted by HTSeq-Count [42] and the differential expression determined by the R-based statistical DESeq [43].

Mapped genes were considered as differentially expressed genes (DEGs) if their relative gene expression levels showed an adjusted *p*-value (FDR) < 0.05 and an amplitude of differential expression of at least 4-FC (log2FC > 2.0 or < -2.0) between RKN samples and controls. Due to typically low levels of expression in plants, NBS–LRR genes were considered as differentially expressed genes with low fold-change ($_L$DEG) [44], if their relative gene expression levels showed an adjusted *p*-value (FDR) < 0.05, regardless of the fold change (FC) amplitude of differential expression values. The visualization of commonly expressed genes among the libraries was conducted by UpSetR (https://gehlenborglab.shinyapps.io/upsetr/).

### Functional classification

Functional annotation and classification of the DEGs into categories via Gene Ontology (GO) terms was based on *A. duranensis* gene models annotation (http://peanutbase.org/). The Hypergeometric test for overrepresentation from FUNC package [45] was used to test for significantly enriched GO categories among *A. duranensis* DEGs using default parameters, and only genes with a FWER < 0.05 for overrepresentation were selected for further analysis. Transcription factors DEGs were identified based on the classification of the Plant TF database (http://planttfdb.cbi.pku.edu.cn/).

### MapMan ontology

To include NBS-LRR genes and obtain an as comprehensive as possible analysis [44], we used all *A. stenosperma* and *A. duranensis* differentially expressed genes with low fold-change ($_L$DEG), which include all genes with significant differential relative gene expression between nematode inoculated and control samples at an adjusted *p*-value (FDR) < 0.05, regardless of its expression magnitude (FC). To identify their biological functions and involvement in biotic stress pathways, gene models from *A. duranensis* reference genome (https://peanutbase.org/) covered by the above selected RNA-Seq reads of *A. stenosperma* and *A. duranensis* were submitted to Mercator [46] against the *Arabidopsis thaliana* database, using default settings. The results were submitted to MapMan to visualize the expression of the genes in the biotic pathway [47].

### NBS-LRR physical clustering, expression profile and phylogenetic analysis

To identify physical gene clusters of NBS-LRR in *A. duranensis* reference genome, the definition of Richly et al., [48] was used; with two or more NBS-LRR genes occurring within a maximum of eight ORFs and less than 250 kb apart.

The *in silico* expression profile of these NBS-LRR genes was carried out by mapping the RNA-Seq data onto their previously predicted classification in the *A. duranensis* reference genome [26]. Those NBS-LRR genes differentially expressed with low fold-change ($_L$DEG) were used to construct the expression clustering.

Clustering analysis of significant genes based on common expression patterns was conducted using the cutree function of gplots package from CRAN [49].

For the phylogenetic analysis of the *Arachis* NBS-LRR family, the NB-ARC domain (PF00931) was used for the HMMsearch against the 345 *A. duranensis* NBS-LRR predicted protein sequences described by Bertioli et al., [26]. Only the sequences with more than 50% of the full-length NB-ARC domain were kept and aligned using MAFFT software [50], using --auto parameter to select the best alignment strategy. We eliminated all columns of the alignment that contained more than 10% of gaps by using trimAl, to provide a more accurate alignment and construction of the phylogenetic tree [51]. Construction of the phylogenetic tree was conducted using RAxML software [52] with an automatic detection of the fittest evolutionary model and an estimated gamma distribution of rates of evolution. For bootstrap replicates, the '-autoMRE' option to automatically stop RAxML upon convergence was used. Duplication analysis was conducted with McscanX (http://chibba.pgml.uga.edu/mcscan2/) as previously described for *A. duranensis* [29].

### Expression analysis by qRT-PCR

The expression analysis of candidate genes was conducted by qRT-PCR using inoculated samples gathered in pools and the respective Ctrl samples. For this, total RNA from three individuals at each collecting point (3, 6, and 9 DAI) were pooled at equal amounts to constitute a biological

replicate of inoculated samples. Two independent replicates of inoculated (STR) and control samples were thus formed for each species (*A. stenosperma* and *A. duranensis*) and used for cDNA synthesis as described above.

Reactions were carried out using three technical replicates for each sample using the Platinum® SYBR® Green qPCR Super Mix-UDG w/ROX kit (Invitrogen, Carlsbad, CA, USA) according to manufacturer's recommendations on StepOne Plus Real-Time PCR System (Applied Biosystem Foster City, CA, USA). The qRT-PCR analysis and specific primer pairs design were conducted for 17 NBS-LRR genes (Additional file 1: Table S1) as previously described [14]. Average cycle threshold (Cq) values were estimated using the online real-time PCR Miner tool [53] and normalized to two reference genes (60S and ACT1), as previously established [39]. Expression ratios of transcripts from the inoculated pool relative to Ctrl pool were determined and statistically tested using REST 2009 software [54].

## Results

### RNA-Seq data for both *A. stenosperma* and *A. duranensis* mapped to the *A. duranensis* reference genome

On average, 97% of *A. duranensis* cleaned RNA-Seq reads from the three time-points studied could be mapped to the *A. duranensis* reference genome where they illuminated around 25,000 gene models (Table 1). In comparison, over 95% of the *A. stenosperma* cleaned reads were cross-mapped to the *A. duranensis* reference genome where they illuminated over 22,000 gene models. The intersection of the two datasets shows that 25,000 gene models are supported by both *A. duranensis* and *A. stenosperma* reads which represent 68% of the approximate 36,000 *A. duranensis* gene models [26] (Table 1). Hence, a large set of genes could be compared regarding differential expression between the two closely related *Arachis* species upon infection by *M. arenaria*.

### Different genes at different time points are regulated during upon nematode infection in the two *Arachis* species

We identified 189 and 657 DEGs after RKN infection in *A. duranensis* and *A. stenosperma* respectively, which showed statistical significance at adjusted *p*-value (FDR) <0.05 and two-fold change [log2 ratio of (control/stress) >2 or <-2]. For the highly resistant *A. stenosperma* most DEGs were modulated at 6 DAI (472) and least at 3 DAI (79), whereas in the moderately susceptible *A. duranensis* the majority of responsive genes occurred at 3 DAI (124), with a steady decrease in the numbers to 9 DAI (46) (Fig. 1a). This suggests that nematode resistance in *A. stenosperma* involves not only the modulation of a larger (3.5 fold) set of genes but also different and specific time points when compared to *A. duranensis*.

Only 63 DEGs were found to be shared between the two species (Fig. 1b) with the majority occurring between *A. duranensis* 3 DAI and *A. stenosperma* 6 DAI/9DAI (23), reinforcing distinct defense responses displayed by the two species against the nematode infection. Also, most exclusive genes appeared at the more relevant time point for the onset of the resistance response in each species with *A. stenosperma* harboring 294 exclusive genes at 6 DAI, while *A. duranensis* exhibited 57 unique genes at 3 DAI (Fig. 1b). Within *A. stenosperma*, the highest number of shared DEGs occurred between 6 and 9 DAI (73), followed by 3 and 6 DAI (28), indicating that the main specific molecular responses take place around the sixth day resulting in bursts of gene expression. In *A. duranensis*, most genes at 3 DAI are shared between the three time points and also with *A. stenosperma* at 6 and 9 DAI (23). Overall, the number of genes shared between all the time points in *A. stenosperma* is superior to that in *A. duranensis*, 19 and 6, respectively (Fig. 1b). This suggests that for the onset of the HR in the resistant species, a considerable number of genes must be triggered around the sixth post-infection

**Table 1** Statistics of *Arachis* spp. reads mapped onto the *A. duranensis* reference genome and analysis of the differentially expressed genes (DEGs) under RKN infection against the control

| Library[a] | Number of mapped genes Replicate 1 | % of mapped reads | Number of mapped genes Replicate 2 | % of mapped reads | Number of $_{L}$DEG (FDR < 0.05) | Number of DEG (FDR < 0.05) (Log2FC >2 or < -2) |
|---|---|---|---|---|---|---|
| DCTR | 25,096 | 97.58 | 25,023 | 97.82 | N/A | N/A |
| DN3 | 25,338 | 97.42 | 25,389 | 97.95 | 2,078 | 124 |
| DN6 | 25,466 | 98.01 | 24,999 | 97.97 | 791 | 61 |
| DN9 | 24,969 | 97.83 | 24,564 | 97.76 | 685 | 46 |
| SCTR | 22,914 | 96.39 | 22,375 | 96.89 | N/A | N/A |
| SN3 | 21,602 | 95.4 | 22,237 | 96.37 | 472 | 79 |
| SN6 | 21,337 | 94.81 | 21,914 | 95.88 | 2,314 | 472 |
| SN9 | 22,193 | 96.39 | 21,615 | 96.31 | 479 | 128 |

[a] Different time points after *M. arenaria* inoculation in *A. duranensis* DCTR – control; DN3 – 3 DAI; DN6 – 6 DAI; DN9 – 9 DAI; and *A. stenosperma* SCTR – control; SN3 – 3 DAI; SN6 – 6 DAI; SN9 9 DAI. N/A – Not Applied

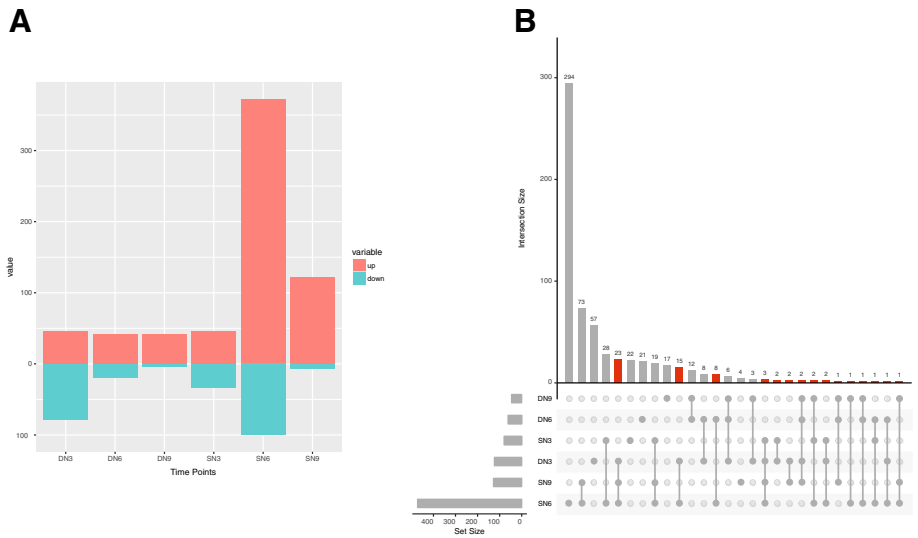



**Fig. 1:** Differentially expressed genes (DEGs) in *A. duranensis* (DN) and *A. stenosperma* (SN) during *M. arenaria* infection at 3, 6 and 9 days after inoculation. **a** Barplot representing the number of DEGs up (red) and down (green) regulated at different time points of the analysis in *A. duranensis* (DN3, DN6, DN9) and *A. stenosperma* (SN3, SN6, SN9) infected roots. **b** Graphic representation of all the intersections between the DEGs in *A. duranensis* (DN3, DN6, DN9) and *A. stenosperma* (SN3, SN6, SN9) infected with *M. arenaria*. The red bars indicate the intersections between *A. duranensis* and *A. stenosperma*. The set size represents the number of DEGs at each condition (genotype/DAI) and the black dots, their intersections

day (6 DAI), with few remaining differentially expressed up to 9 DAI (Fig. 1b). There are no common DEGs across the three treatments and the two species.

The contrasting expression behavior of some genes in response to RKN between these two species (Fig. 1) was further corroborated by the distinct expression profiles of 14 candidate genes previously identified as candidates involved in *A. stenosperma* resistance to RKN [14, 16] (Additional file 2: Table S2). These genes are involved in pathogen perception, signal transduction, protein ubiquitination, hormone signaling and secondary metabolites production (*AsCHI2*, *AsGH3*, *AsTIR-NBS*, *AsBger*, *AsIOMT*, *AsKel*, *AsTAT*, *AsALKBH2*, *AsSLP*, *AsBTB*, *AsAraH8*, *AsTMV*, *AsPRR37* and *AsSAG*), which are crucial steps for the triggering and accomplishment of plant defense response. Accordingly, the majority of these genes showed not only upregulation in *A. stenosperma* (6 DAI), but also a significant downregulation in *A. duranensis* at 3 DAI (Additional file 3: Figure S1), reinforcing the relevance of these early time points and strengthening their roles in this defense response.

### The differentially expressed genes encompass different functional categories in the two *Arachis* species

Gene ontology (GO) was applied to categorize the function of DEGs identified in both species and to analyze the enrichment of these categories in each of the three time points after nematode inoculation (3, 6, 9 DAI) (Fig. 2). In the molecular function category, many GO terms related to catalytic activities, oxygen reduction and scavenging of ROS products such as peroxidases, oxireductases, pectinesterases and other antioxidant enzymes were significantly enriched in *A. stenosperma*, at 3 and 6 DAI, when the HR response occurs (Fig. 2) [55]. There is a clear enrichment for protein kinase activity in *A. duranensis* roots at 3 DAI when most genes are regulated in this genotype in response to the nematode. This is probably a result of intense cell activity, as kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction. In the biological process category, an enrichment of terms related to protein phosphorylation, compatible with kinase activity occurs at 3 DAI in *A. duranensis*, while the response to oxidative stress is a frequent GO term in *A. stenosperma*. Abundance of terms related to DNA replication and cellular components movement is also observed in the last two stages of *A. duranensis* interaction (6 and 9 DAI) suggesting the beginning of feeding sites formation (Fig. 2).

There was also enrichment for GO terms corresponding to localization or activity in the cell wall, inner cell membrane, and cell periphery. In *A. duranensis* at 6 and 9 DAI, there was an enrichment of GO terms corresponding to minichromosome maintenance complex (MCM), which is involved in both the initiation and the elongation step of eukaryotic DNA replication. Because of its role in genome duplication in proliferating cells, deregulation of the MCM function can result in chromosomal defects that may contribute to tumorigenesis. This coincides with the initial phase of cell gall formation in the moderately susceptible *A. duranensis* [15] and was not identified in *A. stenosperma* (Fig. 2).

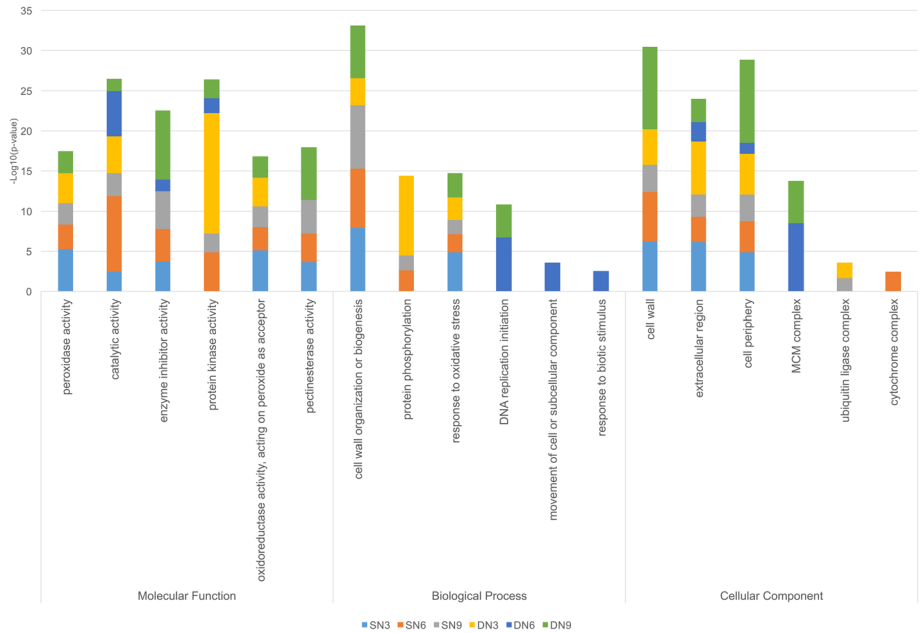



**Fig. 2:** Enrichment of Gene Ontology (GO) terms in DEGs of *A. duranensis* (DN3, DN6, DN9) and *A. stenosperma* (SN3, SN6, SN9) roots infected with *M. arenaria* at different time points (3, 6, 9 DAI). DEGs are distributed in all three functional categories

### RKN infection triggers expression of pathogen-defense related pathways

In the depiction "biotic response to stress", responsive genes from both species belonging to families known to be involved in plant defense towards nematode infection (R genes, PR-proteins, secondary metabolites) were identified (Fig. 3). These genes were distributed in two main pathways (central and laterals) (Fig. 3), which include genes modulated in response to RKN mechanism of parasitism. First, as RKN penetrate the host roots, they cause mechanical wounding which initiates Jasmonic Acid (JA) and Ethylene (ET) production, leading to the activation of signaling molecules, Transcription Factors (TFs), cell wall enzymes, and secondary metabolites, as shown in both lateral sides of Fig. 3. Then, as RKN produce elicitors to establish the feeding sites, they also trigger a specific defense response, initiated by R genes and mediated by signaling molecules, including kinases (MAPK), culminating with the production of PRs and other defense proteins, as visualized in the central region of the MapMan illustration (Fig. 3).

The distribution of transcripts ($_{L}$DEG) into defense pathways showed clearly the differences between the expression patterns of suites of genes activated in two contrasting species (Fig. 3). We theorize that in the resistant *A. stenosperma*, the immune response might be triggered by proteins that recognize specific effectors, such as resistance (R) proteins, as R genes seem to be most upregulated at 6 DAI and to a lesser extent at 9 DAI (Fig. 3). This is in clear contrast with *A. duranensis* transcripts assigned to the same function, that showed a general downregulation at 3 and 6 DAI. Following R genes activation, a respiratory burst and an enrichment of signaling molecules were observed mainly in *A. stenosperma* which might contribute to the control of the pathogen spread, leading to a striking rise in the expression of defense proteins (PRs) at 6 DAI, which coincides with HR in this species [55] (Fig. 3).

Pathogen-related proteins (PRs) were detected as the most represented defense protein, with 91 representatives modulated in both species, albeit with opposite expression trends. In *A. duranensis* 33 exclusive PR genes were mostly downregulated (3 DAI), while 21 unique to *A. stenosperma* were predominantly upregulated (6 DAI). Transcripts associated with pathogen cell wall breakdown (ß-glucanases), plant hormonal balance and cell wall modification and production of secondary metabolites were also differentially regulated in these two species (Fig. 3).

It is notable that the major signaling pathways triggered in *A. stenosperma* in response to *M. arenaria* seems to involve JA and ET, especially at 6 DAI (Fig. 3), although other hormones involved in plant development, such as auxins, abscisic acid (ABA), and brassinosteroids, also seem to participate in this defense response. In contrast, the activation of the Salicylic Acid (SA) signaling pathway seems to be more prevalent in *A. duranensis* than in *A. stenosperma*, which might be related to the well described jasmonate-salicylate antagonism occurring in other plant-pathogen interactions [56].

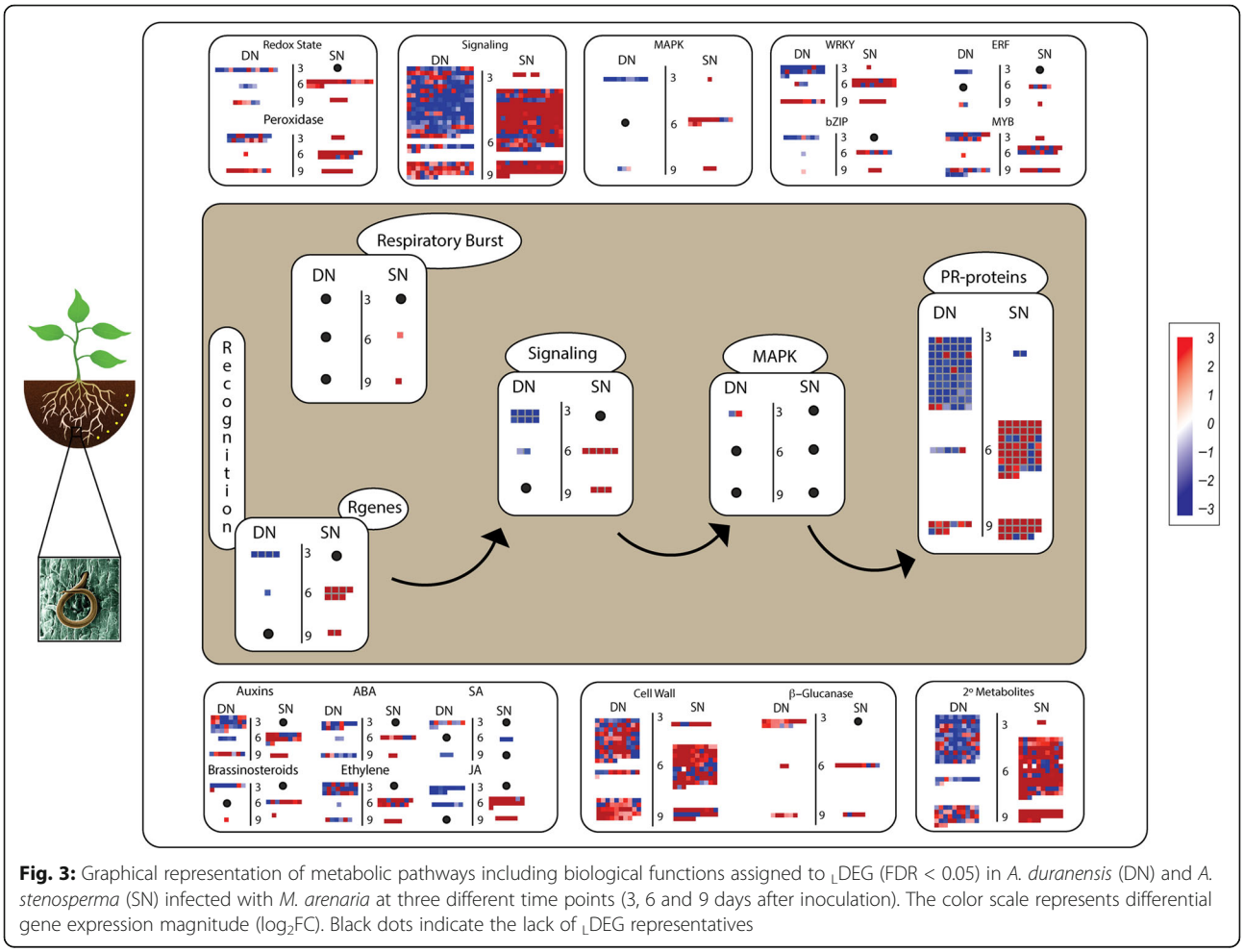



**Fig. 3:** Graphical representation of metabolic pathways including biological functions assigned to $_{L}$DEG (FDR < 0.05) in *A. duranensis* (DN) and *A. stenosperma* (SN) infected with *M. arenaria* at three different time points (3, 6 and 9 days after inoculation). The color scale represents differential gene expression magnitude ($log_2$FC). Black dots indicate the lack of $_{L}$DEG representatives

### Transcription factors induced upon RKN infection

In the large-scale transcriptional reprogramming observed in both species in response to RKN infection, four TF families WRKY, MYB, ERF and bZIP played a critical role (Fig. 3). Therefore, a more detailed expression analysis of 105 DEGs belonging to these families was conducted. Overall, the expression of these TFs varied during nematode infection according to the species and time point observed, with most genes showing a contrasting expression behavior between the two species (Additional file 4: Figure S2).

A subgroup of 11 representatives of the WRKY family showed upregulation in *A. stenosperma*, especially at 6 DAI, while downregulation in *A. duranensis*, mainly at 3 DAI (Additional file 4: Figure S2A), which suggests they might play a role in their contrasting resistance response. This is expected, as JA and SA plant defense pathways activated in this interaction, require large scale transcriptional reprogramming, including *WRKY* genes [57]. Likewise, four *MYB* genes showed strong upregulation in *A. stenosperma* at 6 DAI and downregulation in *A. duranensis* at 3 DAI (Additional file 4: Figure S2B). Members of *MYB* family are modulated by wounding [58], and in turn regulate some flavonoid genes involved in plant defense response, as observed in this study.

Five *bZIP* representatives showed a contrasting expression behavior in the two species studies, being upregulated in *A. stenosperma* mainly at 3 and 6 DAI and downregulated in *A. duranensis* at 3 DAI [59] (Additional file 4: Figure S2C). The same trend was observed amongst *ERF* members that are responsive to salt, cold, drought, wounding and fungi [60] and were, in their majority, strongly upregulated in *A. stenosperma* and downregulated in *A. duranensis* (Additional file 4: Figure S2D).

### NBS-LRR expression profiling upon RKN infection

The overall transcriptional reprogramming outline of transcripts associated to *A. stenosperma* resistance to RKN (Fig. 3) and the occurrence of the HR response strongly suggests that R genes, especially those encoding NBS-LRR proteins, play a pivotal importance in this incompatible interaction. Therefore, they were further characterized in this study in terms of expression behavior, phylogeny and physical clustering aiming to dissect their

functional roles in the RKN response in both *Arachis* species.

From the 345 NBS-LRR-encoding genes of *A. duranensis* [26], 52 were differentially expressed genes with low fold-change ($_L$DEG) (FDR<0.05) in at least one condition (genotype/DAI). Among these, 27 were modulated between infected and control roots in *A. stenosperma* and 37 in *A. duranensis*, of which only 12 were identified in both species.

Most of these 52 NBS-LRR genes showed a contrasting expression profile between A. duranensis and A. stenosperma upon nematode infection, with six expression clusters identified (Fig. 4). Genes comprising clusters A, B and C tend to show downregulation in *A. stenosperma* and upregulation in *A. duranensis*, whilst those in cluster D, E and F showed opposite expression profile. Cluster D comprised all the NBS-LRR that were upregulated in *A. stenosperma* at 6 DAI and most of those downregulated in *A. duranensis* at 3 DAI, which are the critical points of the onset of the defense response in each species [15].

In *A. duranensis*, most NBS-LRR genes were modulated at 3 DAI (34) with 20 exclusives to this time point (Additional file 5: Figure S3). On the other hand, in *A. stenosperma*, 6 DAI was the time point with most representatives (25) with 12 exclusives genes modulated (Additional file 5: Figure S3; Fig. 4). All the genes that are differentially regulated in both, *A. stenosperma* and *A. duranensis* are in the cluster D (Fig. 4). We suggest that the contrasting expression behaviour of the majority of NBS-LRR genes between the two species (Fig. 4) contribute to their distinctive defense responses, as *A. stenosperma* shows a strong Hypersensitive Response (HR), being practically immune to the nematode, whilst *A. duranensis* response propels a delay in the parasite penetration and development of the feeding cell [15].

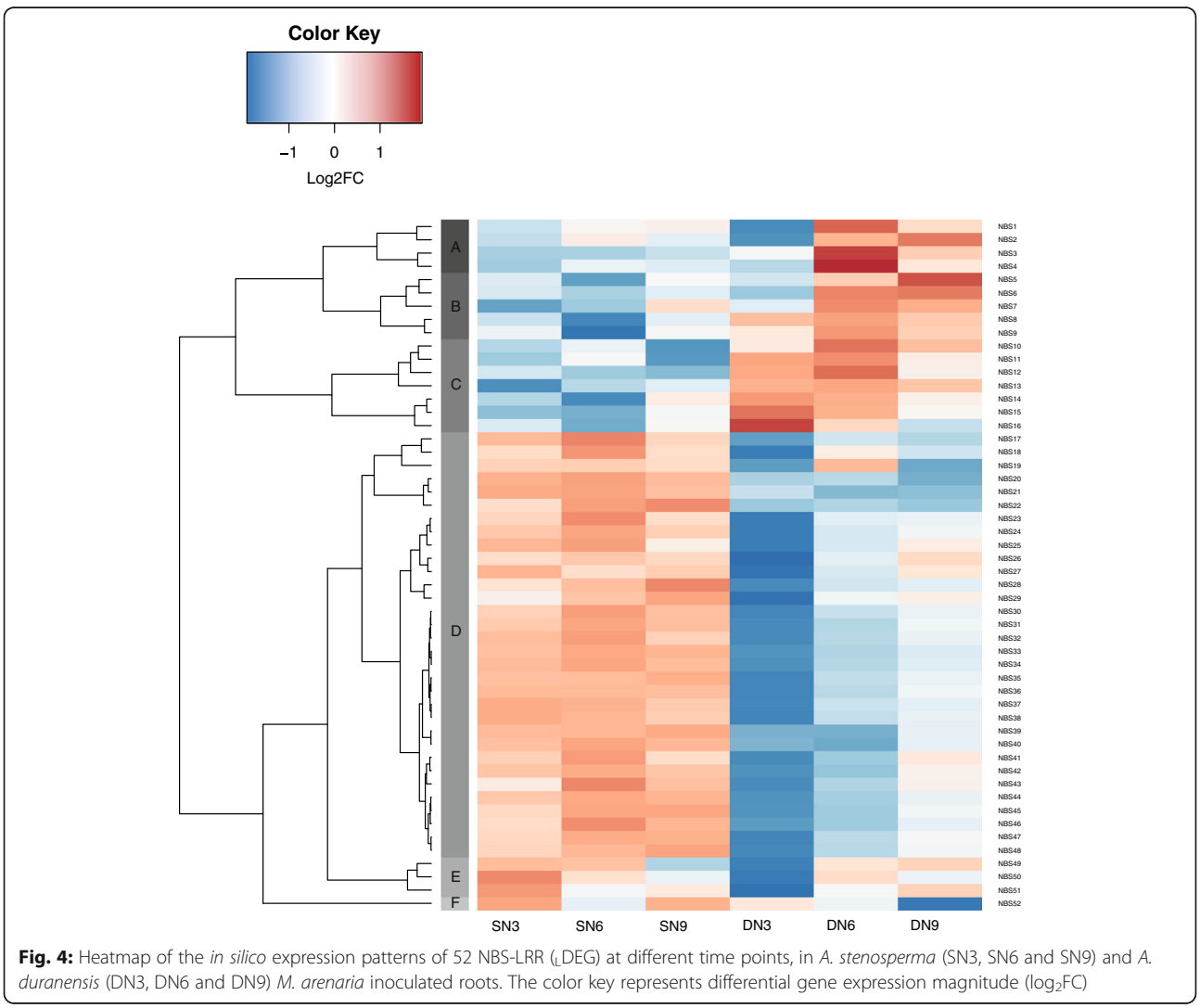



**Fig. 4:** Heatmap of the *in silico* expression patterns of 52 NBS-LRR ($_L$DEG) at different time points, in *A. stenosperma* (SN3, SN6 and SN9) and *A. duranensis* (DN3, DN6 and DN9) *M. arenaria* inoculated roots. The color key represents differential gene expression magnitude ($log_2$FC)

### Chromosomal distribution of NBS-LRRs

The 37 NBS-LRR genes identified as $_{L}$DEG in *A. duranensis* were distributed in all chromosomes, except for chromosome Aradu.A07 (http://peanutbase.org/), while the 27 *A. stenosperma* NBS-LRR ($_{L}$DEG) were placed on fewer chromosomes with no representatives on chromosomes Aradu.A04, A06 and A07 (Additional file 6: Table S3).

All *Arachis* NBS-LRR ($_{L}$DEG) were divided into two classes according to the presence or absence of the TOLL/interleukin-1 receptor (TIR), being named TIR-NBS-LRR and CC-NBS-LRR [19] and further classified into four subfamilies based on the presence of the above domains alone or in combination (N, TNL, TN, NL) [19]. Among the 37 *A. duranensis* NBS-LRR ($_{L}$DEG), the great majority (92%) belonged to NL and TNL subclasses, whilst only three to TN (8.1%). A similar distribution of the 27 *A. stenosperma* NBS-LRR ($_{L}$DEG) occurred with 15 in the NL subclass (55.5%), eight in TNL (29.6%) and only four TN (14.8%) (Additional file 7: Figure S4). Interestingly, no representatives of subclass N were identified as differentially expressed in response to RKN infection in neither of the two *Arachis* species studied.

### NBS-LRR phylogenetic analysis

For the phylogenetic analysis, we kept only the 314 *A. duranensis* proteins that had more than 50% of the full-length NB-ARC domain (PF00931). The midpoint-rooted phylogenetic tree exhibited a basal separation into two major groups supported by high bootstrap values (93) (Fig. 5). The first group was mainly composed of TIR-type NBS-LRRs, whilst the second of CC-type NBS-LRRs. (Fig. 5). A small monophyletic group of CC-type NBS-LRRs, supported by high bootstrap value (99) was outgroup and more closely related to the TIR-types than to the rest of the CC-types. This suggest that the CC-type might be the ancestral NBS-LRR structure and that the TIR-type evolved secondarily by losing the CC domain and gaining a TIR domain. We also noted that in the clade containing most of CC-types, a few proteins with both a TIR and a CC domain were observed (Fig. 5; green and orange circles). Interestingly, all these cases belong to one single monophyletic, albeit not highly supported clade (bootstrap=23). This suggests that a subgroup of CC-type NBS-LRRs might have secondarily gained a TIR domain.

Among the 314 *A. duranensis* corresponding genes, 46 were identified as $_{L}$DEGs and are highlighted in the phylogenetic tree (colored names - Fig. 5). These genes were distributed across different branches of the tree, showing that their expression was not restricted to a specific subfamily of NBS-LRR or even to a CC or TIR type. However, genes located in the same chromosome tend to group in subclades in the phylogenetic tree. This is consistent with tandem duplication being the most common type of duplication (46% of duplicated NBS-LRR genes), while proximal and dispersed duplication represented only 28 and 21%, respectively (Additional file 6: Table S3). The duplication pattern did not correlate with the NBS-type, being equally distributed between the CC and TIR-types. Genes not harboring a sufficiently complete NB-ARC domain were not integrated in this tree; however, they showed a similar expression pattern to the rest of their paralogous copies (Figs. 4 and 5).

### NBS-LRR genes physical clustering

From all NBS-LRR genes identified, 55% (172) were in clusters located on all *A. duranensis* chromosomes (except for A06 and A07), which are represented by dots inside the chromosomes in Fig. 6. From these, 29 were $_{L}$DEGs (red dots) with no representatives in chromosomes Aradu.A06, A07 and A10. (Figure 6). The majority of these NBS-LRR clusters contained only one $_{L}$DEG representative; however, in chromosomes Aradu.A01, A02, A04, and A09 some of these low expressed genes ($_{L}$DEG) (12) appeared in pairs and belonged to the same species. All these five clusters containing the $_{L}$DEG are homogeneous, where the same type of NBS (CC or TIR) is observed, with three clusters assigned to TIR-type and two to CC-type (Fig. 6). Homogeneous clusters are expected due to the predominant type of duplication being tandem.

In chromosomes Aradu.A01 and A02, we identified NBS-LRR clusters constituted respectively of four and two $_{L}$DEGs in both *A. stenosperma* and *A. duranensis* (Fig. 6). These six genes belong to the expression group D (Fig. 4), characterized by a strong upregulation in *A. stenosperma* (6 DAI) and downregulation in *A. duranensis* (3 DAI) (Figs. 4 and 6). Another cluster containing two $_{L}$DEGs was found in chromosome Aradu.A04. However, these genes appeared exclusively in *A. duranensis* and belonged to different expression groups (A and D) with opposite expression trends upon nematode infection (Figs. 4 and 6).

Chromosome Aradu.A09 harbored the largest number of differentially expressed genes with low fold-change ($_{L}$DEG) NBS-LRRs in physical clusters (Fig. 6), and contained two NBS clusters with genes exclusively regulated in *A. stenosperma*. Fittingly, a significant QTL for *M. arenaria* resistance has been found in this chromosome [12]. Moreover, most of the genes in these two clusters (except by NBS14) belong to expression cluster D, which showed denotable upregulation in *A. stenosperma* (6 DAI) (Fig. 4).

All the NBS-LRR ($_{L}$DEG) grouped in clusters were compared against the Plant Resistance Genes database PRGdb (http://prgdb.crg.eu) using BlastP to verify their homology to known resistance genes. Remarkably, four genes in chromosome Aradu.A09 (NBS22, NBS14, NBS19 and the non-$_{L}$DEG Aradu.Y6X63), which were exclusively regulated in *A. stenosperma*, showed homology to the

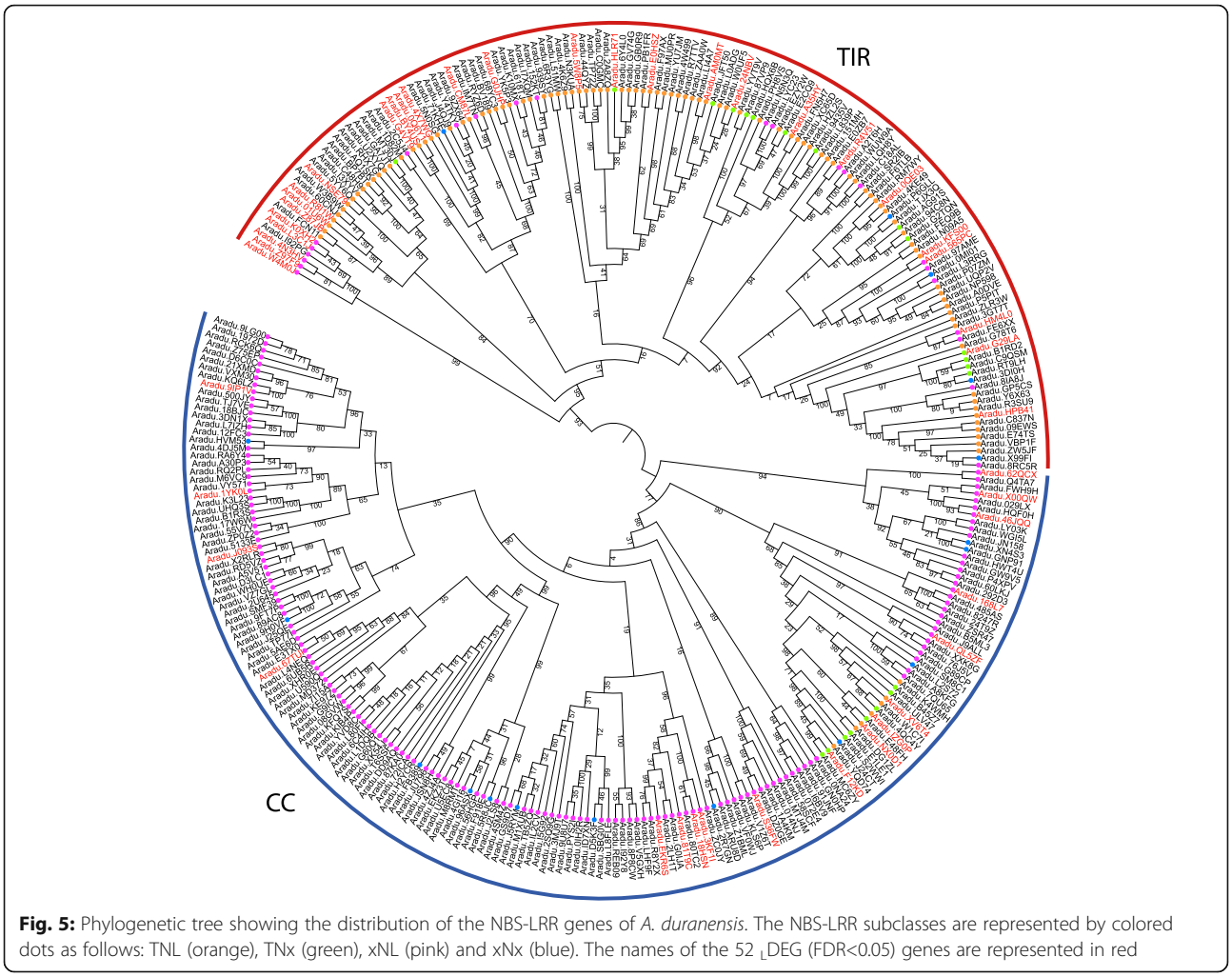



**Fig. 5:** Phylogenetic tree showing the distribution of the NBS-LRR genes of *A. duranensis*. The NBS-LRR subclasses are represented by colored dots as follows: TNL (orange), TNx (green), xNL (pink) and xNx (blue). The names of the 52 $_{L}$DEG (FDR<0.05) genes are represented in red

resistance gene *Gro*1-4 against the yellow potato cyst nematode *Globodera rostochiensis* [61]. In contrast, NBS-LRR ($_{L}$DEG) showing no contrasting expression behavior between the two species within the same clusters lacked homology to known nematode resistance genes, with higher similarity to resistance genes against virus or fungi.

### Expression analysis by qRT-PCR

The expression behavior of seven representatives of the major NBS-LRR expression group (D) (Fig. 4) was validated in both *Arachis* species by qRT-PCR using specific primers (Additional file 1: Table S1) and RNA pooled from the three time points for both species. Expression analysis showed a contrasting behavior between the two species for all NBS-LRR transcripts tested (Fig. 7a) in accordance with *in silico* analysis (Fig. 4), i.e., an upregulation (ranging from 1.168 to 8.89 FC) in *A. stenosperma* and downregulation (ranging from 0.883 to 0.466 FC) in *A. duranensis* (Fig. 7a).

The expression profiles of ten NBS-LRR ($_{L}$DEGs) genes located in clusters (Fig. 6) was also evaluated by qRT-PCR using specific primers (Additional file 1: Table S1) and RNA pooled from the three time points studied in *A. duranensis* and *A. stenosperma* (Fig. 7b). The qRT-PCR analysis corroborated the induction of all the $_{L}$DEGs belonging to clusters on chromosomes Aradu.A01, A02 and A09 in *A. stenosperma*. Nevertheless, in the moderately susceptible species (*A. duranensis*) only six genes were amplified, as four $_{L}$DEGs were *A. stenosperma* exclusive (NAS) (Fig. 7b).

## Discussion

The genomic similarity between *A. stenosperma* and *A. duranensis* allowed the mapping of 95% of *A. stenosperma* RNA-Seq reads, which lacks a complete sequenced genome, to the *A. duranensis* reference genome with no significant loss of information, as already seen for other related species [62–64]. We identified over 26% *A. stenosperma* genes as significantly differently expressed (FDR < 0.05) with at least four fold (DEGs) when infected by RKN, compared to control in one or more time points. The smaller number of DEGs

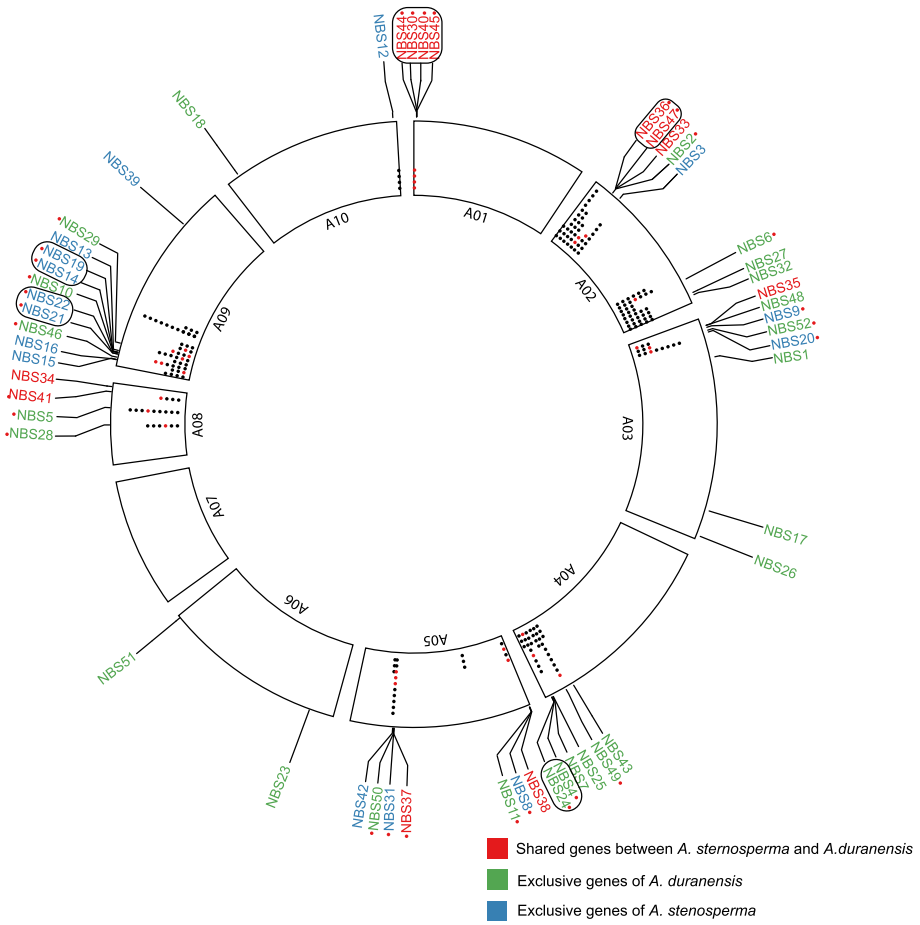



**Fig. 6:** Distribution of *A. stenosperma* and *A. duranensis* NBS-LRR genes in clusters, on the *A. duranensis* chromosomes. Black dots show NBS-LLR genes in clusters; red dots show $_{L}$DEGs in clusters. Genes shared between *A. duranensis* and *A. stenosperma* are named in red, *A. stenosperma* exclusive genes in blue and *A. duranensis*, in green. Red dots next to gene names highlight the $_{L}$DEGs. Physical clusters containing $_{L}$DEGs in pairs from the same species are circled

found, in comparison to our previous works using *de novo* assembly [16], is probably due to a less redundant, albeit more comprehensive, reference gene set coming from the genome assembly. Genome-based strategies are known to provide less chimeric contigs, better filtration of contaminations and a greater representation of the low-abundant transcripts [35].

According to Proite et al., [15], the development of *M. arenaria* in the resistant (*A. stenosperma*) and in the moderately susceptible (*A. duranensis*) differed along the interaction. In the resistant genotype, no giant cells were formed and the HR response was triggered, blocking the nematode development, while in the moderately susceptible species, the giant cell were formed around the 19th day after the inoculation (J3 to J4), albeit in smaller numbers. Likewise, in this study, a distinct downstream gene reprogramming upon nematode infection was observed in the two *Arachis* species, with most genes being upregulated in the resistant, whilst downregulated in the more susceptible species. The timing of gene expression responses was also different, as in *A. duranensis* most DEGs appear at the very beginning of the interaction (3 DAI) which is compatible to the early onset of the PTI (Pattern-Triggered Immunity) [65–67], whilst in *A. stenosperma* most genes are upregulated at 6 DAI, resembling an ETI (Effector-Triggered Immunity) type of response. In addition, a higher number of upregulated genes in *A. stenosperma* assigned to the categories of R genes, PR proteins, TFs and secondary metabolites were upregulated at 6 and 9 DAI. These genes expression patterns coincide with the occurrence of the Hypersensitive Response (HR) displayed by this genotype [15]. The distinctive expression behavior of candidate genes previously identified as being involved in HR response to *M. arenaria* [16] herein observed between the two *Arachis* species, corroborates their different mechanisms of defense responses.

Despite JA being classically more related to resistance to necrotrophic and SA to biotrophic pathogens,

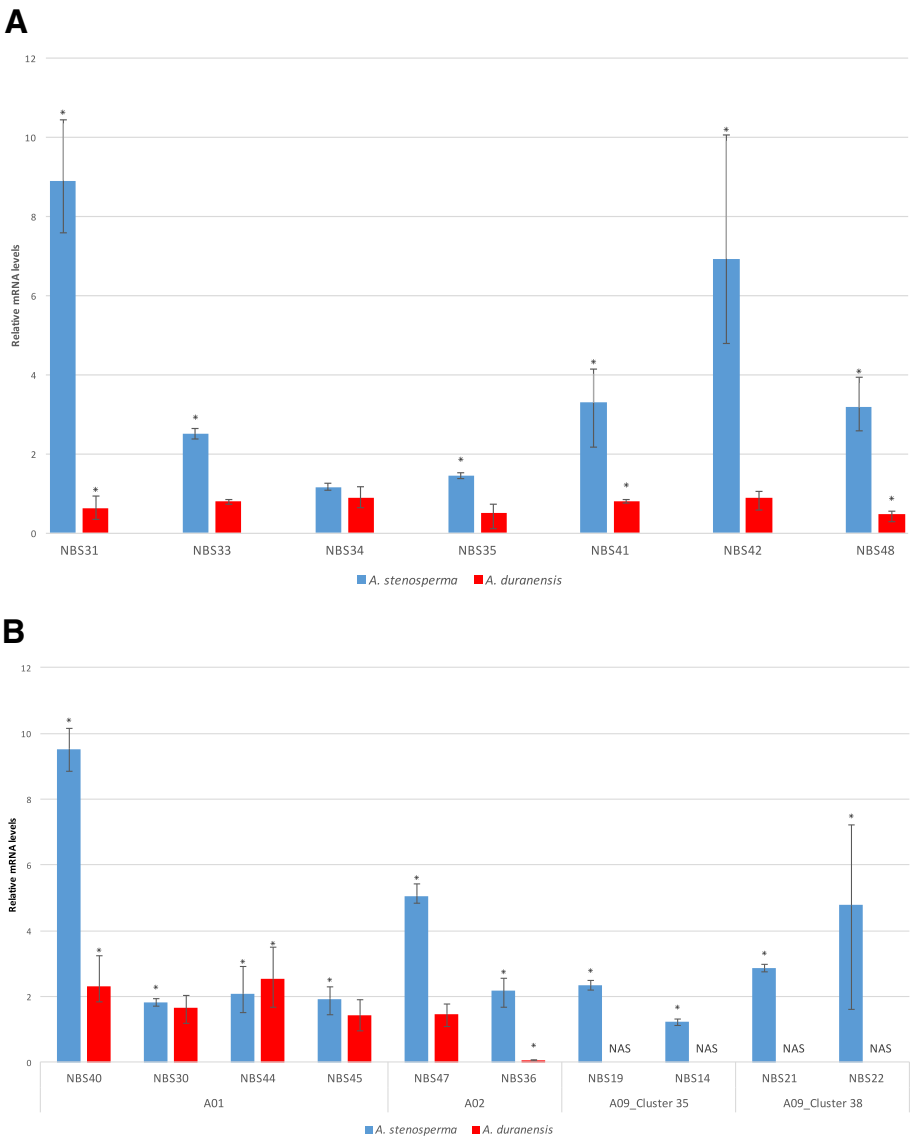



**Fig. 7:** Relative mRNA levels of the NBS-LRR genes in *A. duranensis* and *A. stenosperma*. **a** Transcripts relative quantification of ten $_{L}$DEGs of *A. duranensis* (red) and *A. stenosperma* (blue) distributed in four clusters, on chromosomes Aradu.A01, A02, and A09. The expression levels of the $_{L}$DEGs in *A. duranensis* and *A. stenosperma* relative to control. Non-amplified sequences (NAS) corresponding to exclusive genes of *A. stenosperma*. **b** Expression levels of seven $_{L}$DEGs in *A. stenosperma* (blue) and *A. duranensis* (red) inoculated with *M. arenaria* relative to control samples. Statistically significant regulated genes (*) and error bars

Jasmonate–Ethylene (ET/JA) pathways have also been found to be essential to the resistance to some biotrophic pathogens [68–70]. In *A. stenosperma*, a clear upregulation of genes involved in the ET/JA pathways was observed, particularly at 6 and 9 DAI, while genes related to salicylic acid biosynthesis were mostly downregulated. Although a reciprocal inhibition between SA and JA is well documented [71, 72], many examples show that the interplay between these hormones pathways is not always antagonistic [73, 74], with some evidences that SA can induce JA synthesis and promote ETI responses [75]. This seems to be the case here, with the suppression of SA possibly being triggered by the enrichment for other genes involved in the ET/JA pathways, such as auxins induced by the RKN nematode [76].

Typically, resistance to RKN in both wild and domesticated plants is conferred by R genes mostly from the NBS-LRR family, which sense nematode effectors and trigger the first specific response (ETI) to the nematode infection initiating a cascade of defense genes [77]. However, a typical plant genome contains hundreds of NBS-LRR genes, located in chromosomal clusters, which makes the dissection of the targeted locus and its isolation by functional or genetic recombination rather

difficult [78]. In this study, the analysis of NBS-LRR genes expression in contrasting genotypes enabled the identification of 27 differentially expressed genes with low fold-change ($_L$DEG) in *A. stenosperma* RKN infected roots, of which 19 showed upregulation, while their orthologs in *A. duranensis* were downregulated (Fig. 4), pointing to their involvement in the HR response displayed by the resistant species. The phylogenetic analysis of the NBS-LRR family shows a basal separation in two highly supported clades, one mainly composed of TIR-type and the other of CC-type. This corroborates previous studies in *A. duranensis* [26, 27] and other plants species [79, 80]. However, the different phylogenetic groups do not match the different NBS-LRRs subclasses that are based on the arrangement of the different main domains (CC, TIR, NBS, LRR). As our phylogenetic tree is based on the widely conserved NB-ARC domain, this strongly suggests promiscuous domain shuffling in the NBS-LRR family, resulting in multiple convergent emergences of the same arrangements rather than common ancestral inherited arrangements.

The occurrence of physical clusters of NBS-LRR genes has been reported in different plant species, including *A. duranensis* [26, 27] and can be associated to the co-expression of these genes, the type of duplication that occurred during evolution of the plant species or their biological role [81–83]. In our study, most of the NBS–LRR ($_L$DEG) belonged to the tandem duplication type in both *Arachis* species. This kind of duplication is well known to contribute to the expansion of genes involved in stress responses such as those responding to pathogens [27, 84, 85]. Also, as seen in other studies [26, 86], most NBS-LRR clusters found here are accumulated in the hot recombination regions of the distal chromosomal regions, which favor recombination reshuffling and R loci which require rapid diversification to overcome the emergence of new pathogen races.

Over a third (32%) of the NBS-LRRs responding to *M. arenaria* infection in both species were found in pairs with the same expression pattern within these clusters, which is often necessary for an effective resistance [87]. Clustering has frequently been related to co-expression of functionally related genes [23, 85, 88], and suitably, a larger number of these NBS-LRR pairs occurred in the resistant species (*A. stenosperma*) than in the more susceptible species (*A. duranensis*).

Remarkably, on chromosome 9, two NBS-LRR clusters are composed exclusively of *A. stenosperma* $_L$DEGs. Among them, three genes (NBS14, NBS19 and NBS22) show high similarity to the potato *Gro1-4* gene, which confers resistance to another endoparasitic nematode, the cyst nematode *G. rostochiensis* [61], and belong to the TIR-type class which is often associated with resistance to different types of biotic stresses [27, 89, 90]. Also on chromosome Aradu.A02, another two NBS-LRRs (NBS36 and NBS47) showed high similarity to *Gro1-4*, and although belonging to gene clusters shared between both *Arachis* species, they exhibited contrasting gene expression behavior. Some genes in this same cluster with no significant differential expression were assigned to other R-gene categories, showing that, albeit genes in a cluster are physically close, they might not all show significant differential expression or bear the same function [23].

Despite differences on the parasitic process and feeding site structures between RKN and the cyst nematode *Globodera* spp. [91], both are obligate sedentary endoparasites, relying on juveniles (J2) to penetrate the roots and induce feeding sites. Therefore, the similarity between the NBS-LRRs identified here, especially the RKN-responsive in *A. stenosperma*, and known nematode R genes, might pose a functional role for these genes in *Arachis*. In addition, the fact that these NBS-LRR clusters are located on chromosome Aradu.A02 and A09, near two major QTLs identified in a population derived from a cross between the above species, which are associated with the reduction of *M. arenaria* root galling and egg production [12], reinforces their role in this strong resistance response. Thus, the genes located in the above clusters will be prioritized for further functional investigation, either by their overexpression in susceptible genotypes or genome editing.

Plant immune signaling network is rather complex. Nonetheless, this complexity is necessary, as pathogens continuously evolve a new repertoire to overcome host immune responses and plant adaptation is much slower than pathogen evolution [92]. Therefore, plants should carefully balance this complex network, as it might have a negative impact on their fitness. The use of contrasting wild *Arachis* species with different levels of resistance to *M. arenaria* has enabled the identification of expressed NBS-LRR clusters containing genes with potential functional relevance in this defense response, reinforcing the usefulness of comparative genomic analysis of NBS genes as an efficient means of mining functional R genes. Also, studies showing that rapidly evolved NBS genes from different species are capable of conferring defense against the same pathogen in *Gramineae* [93], strengthening their use in breeding programs.

## Conclusion

The further isolation and characterization of the R genes responsible for triggering effective defense cascades found in this study and their careful selection and stacking in order to balance the strength and specificity of this immune response, can contribute for a more durable resistance in peanut and other RKN affected crops.

## Additional files

**Additional file 1: Table S1.** Primers used for qRT-PCR analysis. (XLSX 11 kb)

**Additional file 2: Table S2.** *Arachis stenosperma* defense candidate genes against the RKN *M. arenaria*. (XLSX 27 kb)

**Additional file 3: Figure S1.** Expression profiles of 14 nematode responsive candidate genes in *A. stenosperma* and *A. duranensis* at 3,6 and 9 DAI with *M. arenaria*. (PDF 46 kb)

**Additional file 4: Figure S2.** Transcriptional profile of members of four TF families: **(A)** WRKY **(B)** MYB **(C)** bZIP **(D)** ERF in *A. stenosperma* (SN3, SN6, SN9) and *A. duranensis* (DN3, DN6, DN9) in response to RKN infection. (PDF 1356 kb)

**Additional file 5: Figure S3.** Intersections between the NBS-LRR ($_L$DEG) (FDR<0.05) in *A. duranensis* (DN3, DN6, DN9) and *A. stenosperma* (SN3, SN6, SN9) infected with *M. arenaria*. The set size represents the number of RKN- responsive genes in each condition (genotype/DAI) and the black dots their intersections. (PDF 8 kb)

**Additional file 6: Table S3.** $_L$DEG NBS-LRR genes (FDR < 0.05) in *Arachis duranensis* and *Arachis stenosperma*. (XLSX 34 kb)

**Additional file 7: Figure S4.** Distribution of *A. stenosperma* and *A. duranensis* expressed NBS-LRR subclasses in *A. duranensis* chromosomes (http://peanutbase.org/). (PDF 1178 kb)

### Abbreviations

CC: Coiled-Coil; DAI: Days After Inoculation; DEG: Differentially Expressed Genes; FC: Fold-Change; FDR: False Discovery Rate; GO: Gene Ontology; HR: Hypersensitive Response; LRR: Leucine Rich Repeats; NBS: Nucleotide Binding Site; qRT-PCR: quantitative Reverse Transcription-Polymerase Chain Reaction; QTL: Quantitative Trait Locus; RGA: Resistance gene analogs; RKN: Root-Knot nematode; TF: Transcription Factors; TIR: Toll/interleukin receptor

### Acknowledgements

We thank Dr. José Francisco Montenegro Valls for supplying the wild Arachis (*A. stenosperma* and *A. duranensis*) seeds and Dr. Regina M.D.G.Carneiro for providing the RKN nematode *M. arenaria*.

### Funding

This work was financially supported by grants from CNPq- Brazilian National Council for Scientific and Technological Development; CAPES- Coordination for the Improvement of Higher Education Personnel and FAP-DF- Distrito Federal Research Foundation. Each of the funding bodies granted the funds based on a research proposal. They had no influence over the experimental design, data analysis or interpretation, or writing the manuscript.

### Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. Sequences were deposited in SRA (BioProject Accession: PRJNA284674).

### Authors' contributions

APZM contributed to the design of the work, data analysis and interpretation and drafting the article; BV contributed to the data collection and data analysis; EGJD contributed to the data analysis and interpretation and critical revision of the article; RT contributed to the data analysis and interpretation; ACMB contributed to the design of the work, data analysis and interpretation and critical revision of the article; ACGA data analysis and interpretation and critical revision of the article; SCMLB and DJB contributed to data interpretation and critical revision of the article; PMG contributed to the conception and design of the work, data analysis and drafting the article; All the authors have approved the final version of the paper to be published.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

## Publisher's Note



### Author details

[1]EMBRAPA Genetic Resources and Biotechnology, Brasilia, DF, Brazil. [2]Universidade Federal do Rio Grande do Sul, Campus do Vale, Porto Alegre, RS, Brazil. [3]INRA, Université Côte d'Azur, CNRS, ISA, Paris, France. [4]Center for Applied Genetic Technologies, University of Georgia, Athens, Georgia, USA.

Received: 12 October 2017 Accepted: 26 July 2018
Published online: 06 August 2018

### References

1. Dong W, Holbrook CC, Timper P, Brenneman TB, Mullinix BG. Comparison of methods for assessing resistance to *Meloidogyne arenaria* in peanut. J Nematol. 2007;39:169–75.
2. Nelson SC, Simpson CE, Starr JL. Resistance to *Meloidogyne arenaria* in *Arachis* spp. germplasm. J Nematol. 1989;21:654–60.
3. Dwivedi SL, Bertioli DJ, Crouch JH, Valls JF, Upadhyaya HD, Fárvero A, et al. Peanut. In: Kole C, Genome mapping and molecular breeding in plants, editors. Oilseeds. Berlin Heidelberg: Springer-Verlag; 2007. p. 115–51.
4. Stalker HT, Moss JP. Cytogenetics and utilization of *Arachis* Species. Adv Agron. 1987;41:1–40.
5. Stalker HT, Tallury SP, Ozias-Akins P, Bertioli D, Leal-Bertioli SCM. The value of diploid peanut relatives for breeding and genomics. Peanut Sci. 2013;40:70–88.
6. Chu Y, Gill R, Timper P, Holbrook CC, Ozias-Akins P. Identification of rare recombinants leads to tightly linked markers for nematode resistance in peanut. Peanut Sci. 2016;197:88–93.
7. Clevenger J, Chu Y, Arrais Guimaraes L, Maia T, Bertioli D, Leal-Bertioli S, et al. Gene expression profiling describes the genetic regulation of *Meloidogyne arenaria* resistance in *Arachis hypogaea* and reveals a candidate gene for resistance. Sci Rep. 2017;7:1317.
8. Nagy ED, Chu Y, Guo Y, Khanal S, Tang S, Li Y, et al. Recombination is suppressed in an alien introgression on chromosome 5A of peanut harboring Rma, a dominant root-knot nematode resistance gene. Mol Breed. 2010;26:357–70.
9. Simpson CE, Starr JL. Registration of "COAN" peanut. Crop Sci. 2001;41:918.
10. Guimarães PM, Brasileiro ACM, Morgante CV, Martins ACQ, Pappas G, Silva OB, et al. Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genomics. 2012;13:387.
11. Leal-Bertioli SCMDM, De Farias MP, Silva PÍTIT, Guimarães PM, Brasileiro ACM, Bertioli DJ, et al. Ultrastructure of the initial interaction of *Puccinia arachidis* and *Cercosporidium personatum* with leaves of *Arachis hypogaea* and *Arachis stenosperma*. J Phytopathol. 2010;158:792–6.
12. Leal-Bertioli SCM, Moretzsohn MC, Roberts PA, Ballén-Taborda C, Borba TCO, Valdisser PA, et al. Genetic mapping of resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: a new source of nematode resistance for peanut. G3: Genes, Genomes. Genetics. 2016;6:377–90.
13. Michelotto MD, Barioni W, De Resende MDV, De Godoy IJ, Leonardecz E, Fávero AP. Identification of fungus resistant wild accessions and interspecific hybrids of the genus *Arachis*. PLoS One. 2015;10:1–17.
14. Morgante CV, Brasileiro ACM, Roberts PA, Guimaraes LA, Araujo ACG, Fonseca LN, et al. A survey of genes involved in *Arachis stenosperma* resistance to *Meloidogyne arenaria* race 1. Funct Plant Biol. 2013;40:1298–309.
15. Proite K, Carneiro R, Falcao R, Gomes A, Leal-Bertioli S, Guimaraes P, et al. Post-infection development and histopathology of *Meloidogyne arenaria* race 1 on *Arachis* spp. Plant Pathol. 2008;57:974–80.
16. Guimaraes PM, Guimaraes LA, Morgante CV, Silva OB Jr, Araujo ACG, Martins ACQ, et al. Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS One. 2015;10:e0140937.
17. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. Science. 2013;341:746–51.
18. Kourelis J, van der Hoorn RAL. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. Plant Cell. 2018;30:285–99.
19. Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. Plant disease resistance genes encode members of an ancient

and diverse protein family within the nucleotide-binding superfamily. Plant J. 1999;20:317–32.
20. Li R, Rashotte AM, Singh NK, Weaver DB, Lawrence KS, Locy RD. Integrated signaling networks in plant responses to sedentary endoparasitic nematodes: a perspective. Plant Cell Rep. 2015;34:5–22.
21. Barbary A, Djian-Caporalino C, Marteu N, Fazari A, Caromel B, Castagnone-Sereno P, et al. Plant genetic background increasing the efficiency and durability of major resistance genes to root-knot nematodes can be resolved into a few resistance QTLs. Front Plant Sci. 2016;7:1–9.
22. Bertioli DJ, Leal-Bertioli SCM, Lion MB, Santos VL, Pappas G, Cannon SB, et al. A large scale analysis of resistance gene homologues in *Arachis*. Mol Gen Genomics. 2003;270:34–45.
23. Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, et al. Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. New Phytol. 2011;192:164–78.
24. Wang H, Penmetsa RV, Yuan M, Gong L, Zhao Y, Guo B, et al. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol. 2012;12:10.
25. Yuksel B, Estill J, Schulze S, Paterson A. Organization and evolution of resistance gene analogs in peanut. Mol Gen Genomics. 2005;274:248–63.
26. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48:438–46.
27. Song H, Wang P, Li C, Han S, Zhao C, Xia H, et al. Comparative analysis of NBS-LRR genes and their response to *Aspergillus flavus* in *Arachis*. PLoS One. 2017;12:e0171181.
28. Zheng YX, Li CJ, Liu Y, Yan CX, Zhang TT, Zhuang WJ, Shan SH. Cloning and characterization of an NBS-LRR resistance gene from peanuts (*Arachis hypogaea* L.). Physiol Mol Plant Pathol. 2013;84:70–5.
29. Guimaraes LA, Mota APZ, Araujo ACG, de Alencar Figueiredo LF, Pereira BM, de Passos Saraiva MA, et al. Genome-wide analysis of expansin superfamily in wild *Arachis* discloses a stress-responsive expansin-like B gene. Plant Mol Biol. 2017;94:79–96.
30. Wang P, Song H, Li C, Li P, Li A, Guan H, et al. Genome-wide dissection of the heat shock transcription factor family genes in *Arachis*. Front Plant Sci. 2017;8:106.
31. Song H, Wang P, Lin J-Y, Zhao C, Bi Y, Wang X. Genome-wide identification and characterization of WRKY gene family in peanut. Front Plant Sci. 2016;7:1–15.
32. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015;53:474–85.
33. Cui J, Luan Y, Jiang N, Bao H, Meng J, Science L, et al. Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lncRNA16397 conferring resistance to *Phytophtora infestans* by co-expressing glutaredoxin. The Plant J. 2016;89:577–89.
34. Draffehn AM, Li L, Krezdorn N, Ding J, Lübeck J, Strahwald J, et al. Comparative transcript profiling by SuperSAGE identifies novel candidate genes for controlling potato quantitative resistance to late blight not compromised by late maturity. Front Plant Sci. 2013;4:423.
35. Jain S, Chittem K, Brueggeman R, Osorno JM, Richards J, Nelson BD. Comparative transcriptome analysis of resistant and susceptible common bean genotypes in response to soybean cyst nematode infection. PLoS One. 2016;11:1–21.
36. Li R, Rashotte AM, Singh NK, Lawrence KS, Weaver DB, Locy RD. Transcriptome analysis of cotton (*Gossypium hirsutum* L.) genotypes that are susceptible, resistant, and hypersensitive to reniform nematode (*Rotylenchulus reniformis*). PLoS One. 2015;10:e0143261.
37. Xing X, Li X, Zhang M, Wang Y, Liu B, Xi Q, et al. Transcriptome analysis of resistant and susceptible tobacco (*Nicotiana tabacum*) in response to root-knot nematode *Meloidogyne incognita* infection. Biochem Biophys Res Commun. 2016;482:1114–21.
38. Wu J, Zhu J, Wang L, Wang S. Genome-wide association study identifies NBS-LRR-encoding genes related with anthracnose and common bacterial blight in the common bean. Front Plant Sci. 2017;8:1–15.
39. Morgante CV, Guimaraes PM, Martins A, Araujo ACG, Leal-Bertioli SCM, Bertioli DJ, et al. Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. BMC Res Notes. 2011;4:339.
40. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
41. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010;26:873–81.
42. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013;8:1765–86.
43. Love M, Anders S, Huber W. Differential analysis of count data–the DESeq2 package. Genome Biol. 2014;15:550.
44. Jain N, Thatte J, Braciale T, Ley K, O'Connell M, Lee JK. Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. Bioinformatics. 2003;19:1945–51.
45. Prüfer K, Muetzel B, Do H-H, Weiss G, Khaitovich P, Rahm E, et al. FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinformatics. 2007;8:41.
46. Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, et al. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 2014;37:1250–8.
47. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, et al. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004;37:914–39.
48. Richly E, Kurth J, Leister D. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol Biol Evol. 2002;19:76–84.
49. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data. R Package version 2. 16.0; 2015. p. 1.
50. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
51. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
52. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90.
53. Zhao S, Fernald RD. Comprehensive algorithm for quantitative real-time polymerase chain reaction. J Comput Biol. 2005;12:1047–64.
54. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST{©}) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 2002;30:36.
55. Proite K, Carneiro R, Falcão R, Gomes A, Leal-Bertioli S, Guimarães P, et al. Post-infection development and histopathology of *Meloidogyne arenaria* race 1 on *Arachis* spp. Plant Pathol. 2008;57:974–80.
56. Kunkel BN, Brooks DM. Cross talk between signaling pathways in pathogen defense. Curr Opin Plant Biol. 2002;5:325–31.
57. Pandey SP, Somssich IE. The role of WRKY transcription factors in plant immunity. Plant Physiol. 2009;150:1648–55.
58. Cheong YH, Chang H-SS, Gupta R, Wang X, Zhu T, Luan S. Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol. 2002;129:661–77.
59. Hwang I, Manoharan RK, Kang J, Chung M, Kim Y, Nou I. Genome-wide identification and characterization of bZIP transcription factors in *Brassica oleracea* under cold stress. Biomed Res Int. 2016; https://doi.org/10.1155/2016/4376598.
60. Jisha V, Dampanaboina L, Vadassery J, Mithöfer A, Kappara S, Ramanan R. Overexpression of an AP2/ERF type transcription factor OsEREBP1 confers biotic and abiotic stress tolerance in rice. PLoS One. 2015;10:1–24.
61. Paal J, Henselewski H, Muth J, Meksem K, Menéndez CM, Salamini F, et al. Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. Plant J. 2004;38:285–97.
62. Collins LJ, Biggs PJ, Voelckel C, Joly S. An approach to transcriptome analysis of non-model organisms using short-read sequences. Genome Inform. 2008;21:3–14.
63. Salzberg SL, Sommer DD, Puiu D, Lee VT. Gene-boosted assembly of a novel bacterial genome from very short reads. PLoS Comput Biol. 2008;4:e1000186.
64. Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, et al. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. Science. 2007;318:441–4.
65. Bigeard J, Colcombet J, Hirt H. Signaling mechanisms in pattern-triggered immunity (PTI). Mol Plant. 2015;8:521–39.

66. Li B, Meng X, Shan L, He P. Transcriptional regulation of pattern-triggered immunity in plants. Cell Host Microbe. 2016;19:641–50.
67. Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell. 2006;124:803–14.
68. Robert-Seilaniantz A, Navarro L, Bari R, Jones JDG. Pathological hormone imbalances. Curr Opin Plant Biol. 2007;10:372–9.
69. Yamada S, Kano A, Tamaoki D, Miyamoto A, Shishido H, Miyoshi S, et al. Involvement of OsJAZ8 in jasmonate-induced resistance to bacterial blight in rice. Plant Cell Physiol. 2012;53:2060–72.
70. Berrocal-Lobo M, Molina A, Solano R. Constitutive expression of Ethylene-Response-Factor1 in arabidopsis confers resistance to several necrotrophic fungi. Plant J. 2002;29:23–32.
71. Van der Does D, Leon-Reyes A, Koornneef A, Van Verk MC, Rodenburg N, Pauwels L, et al. Salicylic acid suppresses jasmonic acid signaling downstream of SCFCOI1-JAZ by targeting GCC promoter motifs via transcription factor ORA59. Plant Cell. 2013;25:744–61.
72. Spoel SH, Johnson JS, Dong X. Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. Proc Natl Acad Sci. 2007;104:18842–7.
73. Tamaoki D, Seo S, Yamada S, Kano A, Miyamoto A, Shishido H, et al. Jasmonic acid and salicylic acid activate a common defense system in rice. Plant Signal Behav. 2013;8:e24260.
74. Mur LAJ. The Outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. Plant Physiol. 2005;140:249–62.
75. Liu L, Sonbol F-M, Huot B, Gu Y, Withers J, Mwimba M, et al. Salicylic acid receptors activate jasmonic acid signalling through a non-canonical pathway to promote effector-triggered immunity. Nat Commun. 2016;7:13099.
76. Robert-Seilaniantz A, Grant M, Jones JDG. Hormone crosstalk in plant disease and defense: more than just JASMONATE-SALICYLATE antagonism. Annu Rev Phytopathol. 2011;49:317–43.
77. Williamson VM, Kumar A. Nematode resistance in plants: the battle underground. Trends Genet. 2006;22:396–403.
78. Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat Biotechnol. 2016;34:652–5.
79. Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, et al. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 2012;13:75.
80. Andolfo G, Jupe F, Witek K, Etherington GJ, Ercolano MR, Jones JDG. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biol. 2014;14:120.
81. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet. 2004;20:116–22.
82. Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics. 2008;91:243–8.
83. Yi G, Sze S-H, Thon MR. Identifying clusters of functionally related genes in genomes. Bioinformatics. 2007;23:1053–60.
84. Rodgers-melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, et al. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. Genome Res. 2012:95–105.
85. Michelmore RW, Meyers BC, Young ND. Cluster of resistance genes in plants evolveby divergent selection and a birth-and-death process. Curr Opin Plant Biol. 1998;8:1113–30.
86. Nieri D, Di Donato A, Ercolano MR. Analysis of tomato meiotic recombination profile reveals preferential chromosome positions for NB-LRR genes. Euphytica. 2017;213:206.
87. Sinapidou E, Williams K, Nott L, Bahkt S, Tör M, Crute I, et al. Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. Plant J. 2004;38:898–909.
88. Christie N, Tobias PA, Naidoo S, Külheim C. The *Eucalyptus grandis* NBS-LRR gene family: physical clustering and expression hotspots. Front Plant Sci. 2016;6:1238.
89. Wan H, Zhao Z, Malik A, Qian C, Chen J. Identification and characterization of potential NBS-encoding resistance genes and induction kinetics of a putative candidate gene associated with downy mildew resistance in *Cucumis*. BMC Plant Biol. 2010;10:186.
90. Wan H, Yuan W, Ye Q, Wang R, Ruan M, Li Z, et al. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. BMC Genomics. 2012;13:502.
91. Fuller VL, Lilley CJ, Urwin PE. Nematode resistance. New Phytol. 2008;180:27–44.
92. Tsuda K, Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Curr Opin Plant Biol. 2010;13:459–65.
93. Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J, et al. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Natl Acad Sci. 2013;110:18572–7.

**Early responses to dehydration in contrasting wild *Arachis* species**

Published in PlosOne – 2018

Christina Cleo Vinson, Ana Paula Zotta Mota, Thais Nicolini de Oliveira, Larissa Arrais Guimaraes, Soraya C. M. Leal-Bertioli, Thomas Christopher Rhys Williams, Alexandre Nepomuceno, Mario Alfredo Passos Saraiva, Ana Claudia Guerra Araujo, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

PLOS ONE

RESEARCH ARTICLE

# Early responses to dehydration in contrasting wild *Arachis* species

**Christina Cleo Vinson[1,2], Ana Paula Zotta Mota[1,3], Thais Nicolini Oliveira[1,3], Larissa Arrais Guimaraes[1¤a], Soraya Cristina Macedo Leal-Bertioli[1¤b], Thomas Christopher Rhys Williams[3], Alexandre Lima Nepomuceno[4], Mario Alfredo Passos Saraiva[1], Ana Claudia Guerra Araujo[1], Patricia Messenberg Guimaraes[1], Ana C. M. Brasileiro[1]***

**1** Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP, Final W5 Norte, Brasília, DF–Brazil, **2** Universidade de Brasília, Campus Darcy Ribeiro, Brasília, DF–Brazil, **3** Universidade Federal do Rio Grande do Sul, Campus do Vale, Porto Alegre, RS—Brazil, **4** Embrapa Soja, CP, Londrina, PR—Brazil

¤a Current address: Department of Horticulture and Institute of Plant Breeding, University of Georgia, Tifton, Georgia, United States of America
¤b Current address: Center for Applied Genetic Technologies, University of Georgia, Athens, Georgia, United States of America
* ana.brasileiro@embrapa.br





**Citation:** Vinson CC, Mota APZ, Oliveira TN, Guimaraes LA, Leal-Bertioli SCM, Williams TCR, et al. (2018) Early responses to dehydration in contrasting wild *Arachis* species. PLoS ONE 13(5): e0198191. https://doi.org/10.1371/journal.pone.0198191

**Editor:** Prasanta K. Subudhi, Louisiana State University College of Agriculture, UNITED STATES

**Received:** December 22, 2017

**Accepted:** May 14, 2018

**Published:** May 30, 2018



**Data Availability Statement:** All data generated or analyzed during this study are included in the published article and its Supporting Information files. Sequences were deposited in SRA (BioProject ID: PRJNA284674).

**Funding:** This work was financially supported by grants from Embrapa - Brazilian Agricultural Research Corporation; CNPq- Brazilian National Council for Scientific and Technological Development; CAPES- Coordination for the Improvement of Higher Education Personnel and

## Abstract

Wild peanut relatives (*Arachis* spp.) are genetically diverse and were selected throughout evolution to a range of environments constituting, therefore, an important source of allelic diversity for abiotic stress tolerance. In particular, *A. duranensis* and *A. stenosperma*, the parents of the reference *Arachis* A-genome genetic map, show contrasting transpiration behavior under limited water conditions. This study aimed to build a comprehensive gene expression profile of these two wild species under dehydration stress caused by the withdrawal of hydroponic nutrient solution. For this purpose, roots of both genotypes were collected at seven time-points during the early stages of dehydration and used to construct cDNA paired-end libraries. Physiological analyses indicated initial differences in gas exchange parameters between the drought-tolerant genotype of *A. duranensis* and the drought-sensitive genotype of *A. stenosperma*. High-quality Illumina reads were mapped against the *A. duranensis* reference genome and resulted in the identification of 1,235 and 799 Differentially Expressed Genes (DEGs) that responded to the stress treatment in roots of *A. duranensis* and *A. stenosperma*, respectively. Further analysis, including functional annotation and identification of biological pathways represented by these DEGs confirmed the distinct gene expression behavior of the two contrasting *Arachis* species genotypes under dehydration stress. Some species-exclusive and common DEGs were then selected for qRT-PCR analysis, which corroborated the *in silico* expression profiling. These included genes coding for regulators and effectors involved in drought tolerance responses, such as activation of osmosensing molecular cascades, control of hormone and osmolyte content, and protection of macromolecules. This dataset of transcripts induced during the dehydration process in two wild *Arachis* genotypes constitute new tools for the understanding of the distinct gene regulation processes in these closely related species but with contrasting drought responsiveness. In addition, our findings provide insights into the nature of drought

FAP-DF- Distrito Federal Research Foundation. Each of the funding bodies granted the funds based on a research proposal. They had no influence over the experimental design, data analysis or interpretation, or writing the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

tolerance in wild germoplasm, which might be explored as novel sources of diversity and useful wild alleles to develop climate-resilient crop varieties.

## Introduction

Climate change models suggest a trend towards increasing risks in the agricultural sector worldwide, leading to the emergence of less productive or even inappropriate areas for cultivation in the tropics [1]. In addition, given the massive demand for crop irrigation, intense competition for freshwater is expected to develop and constitute a critical issue in the near future global bioeconomy. Research therefore needs to focus on plant adaptations to changing environmental constraints by developing technologies to increase crop productivity and sustainability in drought-prone areas [2,3]. Indeed, breeding programs are currently concerned with the adaptation of crops to the global climatic fluctuations by increasing their resilience to the impact of changes in rainfall distribution, heat stress intensification, and water scarcity.

Upon perception of a reduction in water availability, well adapted plants rapidly react to minimize water loss, maintain cellular osmotic adjustment, control water flux, and protect cells against oxidative stress and the damaging effects of dehydration. As part of this response, orchestrated molecular networks are activated, and interconnected responses at physiological, morphological, and biochemical levels are triggered to direct the plant to cope with water scarcity [4]. These molecular regulatory mechanisms include the production of osmoprotectants, antioxidants, and hormones, with abscisic acid (ABA) and ethylene (ET) acting as the principal signaling molecules that link transcriptional networks to environmental adaptation. Three signaling cascades are known to act following the onset of the drought stress [5–7]: (1) ABA-dependent signaling, mediated by the ABA-responsive element (ABRE) present in the promoter of various abiotic stress-related genes; (2) ABA-independent signaling towards the dehydration-responsive element (DRE); and (3) ethylene-signaling pathways that are associated with ERE (ethylene-responsive element) sequences in promoters.

Progress in genomics has facilitated the understanding of the mechanisms controlling abiotic stress responses in plants and led to the discovery of hundreds of genes involved in drought tolerance [6,8]. The generation of transgenic crops with enhanced drought tolerance is now a reality due to the use of such genes which activate or modulate specific pathways that improve plant ability to endure water-limited field conditions. However, the biological role of many drought-induced genes that participate in mechanisms underlying plant responses to water deficit is still unknown.

Peanut (*Arachis hypogaea*) is an important source of protein and oil and is, after soybean, the second most cultivated grain legume in the world, with an annual production of about 46 million tons [9]. A number of constraints affects peanut production, including susceptibility to certain biotic and abiotic stresses, and the breeding of improved cultivars has been hindered by its narrow genetic base [10–12]. Conversely, wild *Arachis* species have high genetic diversity and are mainly associated with the savannah-like Cerrado biome, where they have evolved mechanisms to adapt and survive under conditions of limited water availability [10,13]. Thus, the gene pool that involves wild *Arachis* species is a rich source of desirable agronomical traits, including drought tolerance, that could be used to expand the genetic basis of cultivated peanut through the introgression of wild genome segments [11,12,14]. The use of wild germplasm is one of the most promising alternatives to mobilize genetic variability, creating a strategic link between plant genetic resources and breeding programs.

The recent advances in *Arachis* genome sequencing [15,16] have opened new opportunities for the discovery of useful agronomic alleles harbored by wild germplasm. Indeed,

transcriptome studies of wild *Arachis* made possible the identification of candidate genes responsive to several biotic and abiotic stresses, including root-knot nematode attack, fungal infection, and drought, to be deployed in peanut breeding programs[12,17–22]. More recently, the global gene expression profiling of wild *A*. *duranensis* and *A*. *magna* subjected to gradual water deficit enabled the identification of candidate genes associated with major processes underlying plant tolerance to drought conditions, such as those involved in signal transduction, primary metabolism, hormone homeostasis, and protection/adaptation of cellular structures [19,23].

In the present study, a comprehensive transcriptome characterization of two accessions of wild *Arachis* species (*A*. *duranensis* and *A*. *stenosperma*), contrasting in their transpiration behavior during water deficit imposition [24], was conducted aiming to improve the current understanding of the mechanisms deployed by these genotypes to withstand water deficit. *A*. *duranensis* and *A*. *stenosperma* are also the parents of the mapping populations used to construct the diploid genetic map for the AA genome of *Arachis* [25], facilitating their future use in breeding programs. Comparative transcriptome studies in plants have profited from the vast amount of genomic information produced in recent years to compare gene expression by a developmental approach in a single species or by an evolutionary approach across multiple species [26]. Herein, a comparative transcriptomic analysis of these two contrasting *Arachis* genotypes subjected to dehydration revealed a number of common and exclusively regulated genes coding for regulators and effectors involved in water deficit responses. Further qRT-PCR validation revealed promising candidate genes for drought tolerance in wild germoplasm.

## Materials and methods

### Plant material and dehydration treatment

Seeds of *Arachis duranensis* (accession K7988) and *A*. *stenosperma* (accession V10309) were obtained from the Brazilian Arachis Active Germplasm Bank, maintained at Embrapa Genetic Resources and Biotechnology (Brasilia, Brazil). Seeds were germinated on germitex paper with 2% (w/v) Ethrel (2-chloroethylphosphonic acid) and 0.05% (w/v) Thiram (tetramethylthiuram disulfide), at 25 ± 1˚C and 65 ± 5% relative humidity. Fifteen-day-old seedlings were transferred to a hydroponic system in a randomized block experimental design, with 30 individual biological triplicates for *Arachis duranensis* and 30 for *A*. *stenosperma*. Individuals were placed in polystyrene supports (S1 Fig) that allowed seedling roots to be completely immersed in the aerated pH 6.6 balanced nutrient solution [27]. Plantlets were grown in a controlled environment chamber, at 25 ± 2˚C and 60 ± 5% relative humidity, under a photosynthetic photon flux density of 1.5 x $10^3$ μmoles $m^{-2}$ $s^{-1}$, equivalent to 8.93 x $10^4$ lux, and a 12 h day length, as described previously [27]. After four weeks, the dehydration treatment was initiated by withdrawing the nutrient solution (air-drying on chamber temperature; S1 Fig), and roots for both genotypes were collected after zero (T0 or control); 25 (T25); 50 (T50); 75 (T75); 100 (T100); 125 (T125) and 150 (T150) min. Four root samples were pooled per biological replicate for each time point, immediately frozen in liquid nitrogen, and stored at -80˚C until RNA extraction.

### Physiological analysis

The photosynthetic rate (*A*), stomatal conductance ($g_s$), transpiration rate (TR), vapor pressure deficit based on leaf temperature (VpdL), intercellular carbon dioxide ($CO_2$) concentration ($C_i$), and leaf temperature, were measured as previously described [28], using a portable Photosynthesis System (LI-COR, model LI-6400) set at a light intensity of 1,000 μmol $m^{-2}$ $s^{-1}$. Data were collected at the first three time points (T0; T25 and T50), corresponding to the control

and the first 50 min of dehydration from the first three quadrifoliate leaves. Measurements were made on the middle leaflet of completely expanded leaves. Data obtained for each genotype were compared using the t-test ($p \leq 0.05$).

## cDNA library sequencing and *in silico* expression profiling

Root total RNA extraction, purification, and integrity checking were performed according to [29]. For cDNA library construction, RNA samples were pooled for each genotype: a control group (CTR), formed from samples collected at T0, and a stressed group (STR), formed from samples collected throughout the dehydration treatment (T25 to T150) and pooled in equal amounts. cDNA library construction and high-throughput sequencing were conducted at Fasteris SA (Plan-les-Ouates, Switzerland), according to the Illumina HiSeq2000 protocol (paired-end 2x100 bp).

The Illumina raw reads were trimmed using the Trimmomatic software version 0.33 [30] and the quality checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). For mapping to the reference genome, the high-quality reads were submitted to the default settings of the GMAP/GSNAP package [31] against the annotated *A. duranensis* genome available at PeanutBase (http://peanutbase.org/). The mapped reads were counted using the HTSeqCount software version 0.9.1 [32] using the resolution mode union and the analysis of differential gene expression between CTR and STR samples was conducted using the edgeR package in R [33]. Mapped genes were considered as significantly Differentially Expressed Genes (DEGs) when their relative gene expression levels had an adjusted *p-value* (false discovery rate; FDR) <0.05 and at least 4-fold change (FC) value between STR and CTR samples ($\log_2$FC > 2 or < -2). The distribution of the DEGs in the *Arachis duranensis* chromosomes (http://peanutbase.org/) was made using the Circa software (http://omgenomics.com/circa/). Volcano analysis was conducted using the Trinity Differential Expression scripts (https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Differential-Expression).

## Functional analysis

For the comparison with other legumes species, the OrthoFinder algorithm [34] was used to analyze the complete proteome of four legume species (*Arachis duranensis*; *Arachis ipaënsis*; *Glycine max*; and *Phaseolus vulgaris*). OrthoFinder performs a blast all-against-all and uses the reciprocal best hit for each species to cluster the Ortholog Groups (OGs). Family-companion (https://bbric-pipelines.toulouse.inra.fr/family-companion) was used to generate the binary matrix of the OGs, which was submitted to UpSetR [35] for comparison visualization.

The functional annotation of *A. duranensis* (http://peanutbase.org/) gene models was used to classify the DEGs identified in both *Arachis* species for the Gene Ontology (GO) and the Transcription Factors (TFs). The Hypergeometric Test for overrepresentation from the FUNC package [36] was used to test for enriched GO categories among the DEGs. Genes that were differentially expressed were coded "1", whereas mapped genes received the code "0". The results were then analyzed by the REVIGO method to decrease redundancy [37]. DEGs coding for putative TFs were classified according to their respective family according to the Plant TF database (http://planttfdb.cbi.pku.edu.cn/).

To assign functional terms to candidate genes, sequences of all *A. stenosperma* and *A. duranensis* DEGs were submitted to the Mercator software (http://mapman.gabipd.org/web/guest/app/Mercator) against the *Arabidopsis thaliana* reference database, using the default settings. The mapping file predicted by Mercator was further used as input for the MapMan software (http://mapman.gabipd.org/) for visualization of the gene expression data from the two *Arachis* species in different abiotic stress pathways.

### qRT-PCR analysis

Expression profile of selected DEGs was conducted by quantitative reverse transcription PCR (qRT-PCR) analysis using RNA samples from roots of *A*. *duranensis* and *A*. *stenosperma* collected at each time-point (T0 to T150). Total RNA was treated with DNase and converted into cDNA, as previously described [29], and used as the template for qRT-PCR reactions carried out on a StepOne Plus Real-Time PCR System (Applied Biosystems, Foster City, USA). Specific primers for the 20 selected DEGs (S1 Table) were designed using previously described parameters [29], and qRT-PCR reactions were conducted in technical triplicates for each sample, using No Template (NTC) and No Amplification (NAC) as negative controls. Average cycle threshold (Cq) values were estimated using the online real-time PCR Miner tool [38] and normalized with two reference genes (ACT1 and UBI2), in accordance with [39]. For a first series of qRT-PCR analysis, expression ratios of DEG transcripts were validated using STR and CTR sample pools, as described above for cDNA library construction. A more detailed qRT-PCR analysis was then conducted to determine the expression ratios of 15 DEGs at each of the six stressed points (T25 to T150) relative to control point (T0), in independent biological triplicates. The relative quantification of transcripts was determined and statistically tested using the REST 2009 v. 2.0.13 software [40].

## Results

### Wild *Arachis* physiological response to dehydration stress

Physiological analyses were carried out at the beginning of the experiment (0, 25, and 50 min) in order to compare the two wild *Arachis* genotypes (Fig 1; S2 Table). *A*. *duranensis* showed a higher photosynthetic rate, stomatal conductance, and transpiration rate, together with a lower leaf temperature, and vapor pressure difference than *A*. *stenosperma* (Fig 1). The intercellular $CO_2$ concentration was the only physiological parameter for which no statistically significant difference between the two species was observed (Fig 1F). Any changes in these parameters were detected over the course of the experiment (data not shown), most likely due to the fact that these measurements were made at the beginning of the stress treatment which, while a long enough period to introduce a pronounced stress in roots of these species, as demonstrated by the alterations in gene expression presented below, was not long enough to affect leaf gas exchange [27]. Importantly, previous work with the same accessions of these species showed a more conservative whole plant transpiration response to water deficit in *A*. *duranensis* compared to *A*. *stenosperma* [24]. On this basis, the genotypes of *A*. *duranensis* and *A*. *stenosperma* used in this study will refer as 'tolerant' and 'sensitive', respectively, a designation supported by the gene expression results presented below.

### Wild *Arachis* transcriptome response to dehydration stress

Illumina HiSeq2000 sequencing of *A*. *duranensis* and *A*. *stenosperma* transcriptome libraries generated a total of 17.5 and 24.8 million raw reads (Table 1). After removing adapters and low base quality reads, more than 94% high-quality reads ($\geq$ Q30) were used for downstream analyses. The sequencing data was deposited in the NCBI-SRA database under the BioProject number PRJNA284674.

High-quality reads of both species were mapped onto the reference *A*. *duranensis* genome (http://peanutbase.org/) to allow a comparative transcriptome analysis of their water deficit responsiveness. The majority of the reads from roots of *A*. *duranensis* (98.71%) and *A*. *stenosperma* (95.46%) could be mapped to the reference genome, respectively resulting in the

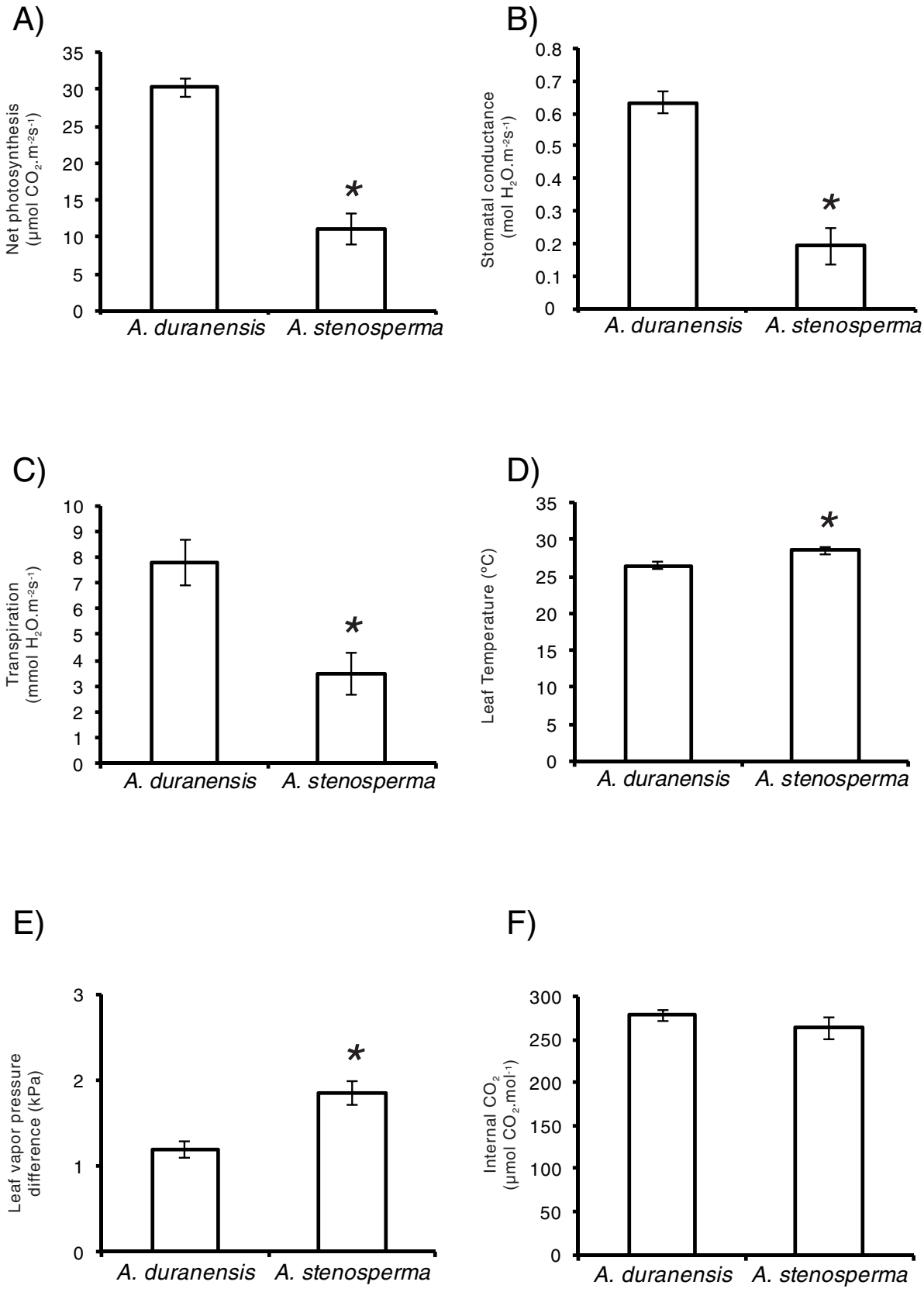

A)
Net photosynthesis (µmol $CO_2$.m$^{-2}$s$^{-1}$)
35
30
25
20
15
10
5
0
*
A. duranensis
A. stenosperma
B)
Stomatal conductance (mol $H_2O$.m$^{-2}$s$^{-1}$)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
*
A. duranensis
A. stenosperma
C)
Transpiration (mmol $H_2O$.m$^{-2}$s$^{-1}$)
10
9
8
7
6
5
4
3
2
1
0
*
A. duranensis
A. stenosperma
D)
Leaf Temperature (°C)
35
30
25
20
15
10
5
0
*
A. duranensis
A. stenosperma
E)
Leaf vapor pressure difference (kPa)
3
2
1
0
*
A. duranensis
A. stenosperma
F)
Internal $CO_2$ (µmol $CO_2$.mol$^{-1}$)
300
250
200
150
100
50
0
A. duranensis
A. stenosperma

**Fig 1. Physiological analysis of *Arachis* spp. plants during the early stages of dehydration.** A) Photosynthetic rate; B) Stomatal conductance; C) Transpiration rate; D) Leaf temperature; E) Leaf vapor pressure deficit; and F) Internal $CO_2$ concentration of *A. duranensis* and *A. stenosperma* plants subjected to dehydration. Data represent averages calculated over the early stages (0, 25 and 50 min) of the drought imposition and standard error. Statistically significant differences (n = 3, Student's t-test, $p<0.05$) are indicated by asterisk.

https://doi.org/10.1371/journal.pone.0198191.g001

identification of 21,125 and 21,502 gene models (Table 1; S3 and S4 Tables), from the 36,734 predicted protein-coding genes previously delineated for *A. duranensis* [15].

*In silico* analysis of differential gene expression, using only FDR< 0.05 as parameter, 1,257 genes could be assigned as significantly expressed between the two *A. duranensis* libraries (AdSTR and AdCTR; Table 1). A similar number (1,725) of significantly expressed genes was also identified between the two *A. stenosperma* libraries (AsSTR and AsCTR; Table 1). When the fold change parameter ($log_2FC > 2$ or $< -2$) was applied to select among them the Differentially Expressed Genes (DEGs), a total of 1,235 DEGs was assigned to *A. duranensis* and 799 to *A. stenosperma* (Table 1; S5 and S6 Tables). In *A. duranensis* (Fig 2A; S5 Table), the great majority of DEGs (1,205; 97.6%) were present in both STR and CTR libraries, with 676 upregulated and 529 downregulated in response to dehydration, whereas only 21 (1.7%) were exclusive to the STR library and nine (0.7%) to the CTR library. Likewise, *A. stenosperma* STR and CTR libraries shared a total of 766 DEGs, with the majority (640) being positively regulated, while 25 DEGs were exclusive to the STR library and eight to the CTR library (Fig 2B; S6 Table).

Comparing the overall responses to dehydration of the two species, a total of 1,602 DEGs were identified in at least one of the species (Fig 2C; S7 Table), with 432 being commonly regulated in both *A. duranensis* and *A. stenosperma* during the dehydration treatment. The majority of common DEGs were induced (369) whilst fewer were repressed (62). Only one DEG, coding for a putative SENESCENCE-ASSOCIATED PROTEIN (SAP; S1 Table). displayed contrasting regulation between the species (induced in *A. duranensis* and repressed in *A. stenosperma*). Regarding the exclusive DEGs, *A. duranensis* showed a higher number (803; 50.1%) than *A. stenosperma* (367; 22.9%) (Fig 2C; S7 Table). The number of upregulated *A. duranensis*-specific DEGs (327) is lower than those downregulated (476) whereas the number of upregulated *A. stenosperma*-specific (296) DEGs clearly exceeded that of those downregulated (71; Fig 2C; S7 Table). These results suggested that most of the genes responsive to the dehydration treatment were species-specific and that the drought responsiveness in the

**Table 1. Summary of sequencing data.** Illumina HiSeq2000 sequencing data of *Arachis* spp. transcriptome and mapping of reads to the reference *A. duranensis* genome.

| Libraries | AdCTR | AdSTR | AsCTR | AsSTR |
|---|---|---|---|---|
| **Yield (Mb)** | 1,817 | 1,691 | 2,444 | 2,518 |
| **%PF[a]** | 97.10 | 96.90 | 97.61 | 96.79 |
| **Total number of reads** | 9,085,802 | 8,453,190 | 12,219,081 | 12,587,737 |
| **≥ Q30(%)[b]** | 96.08 | 95.97 | 94.87 | 94.88 |
| **Mapping in the reference** | ***A. duranensis*** | | ***A. stenosperma*** | |
| **Total number of mapped gene models** | 21,125 | | 21,502 | |
| **Significantly expressed genes[c]** | 1,257 | | 1,725 | |
| **Differentially expressed genes (DEGs)[d]** | 1,235 | | 799 | |

[a] PF stands for "passed filter" i.e. clusters that fulfill the default Illumina quality criteria.

[b] A Q score of 30 (Q30) means that there is a 1 in 1,000 chance that base call is wrong.

[c] FDR< 0.05

[d] FDR< 0.05 and Log2FC (> 2 or < -2).

https://doi.org/10.1371/journal.pone.0198191.t001

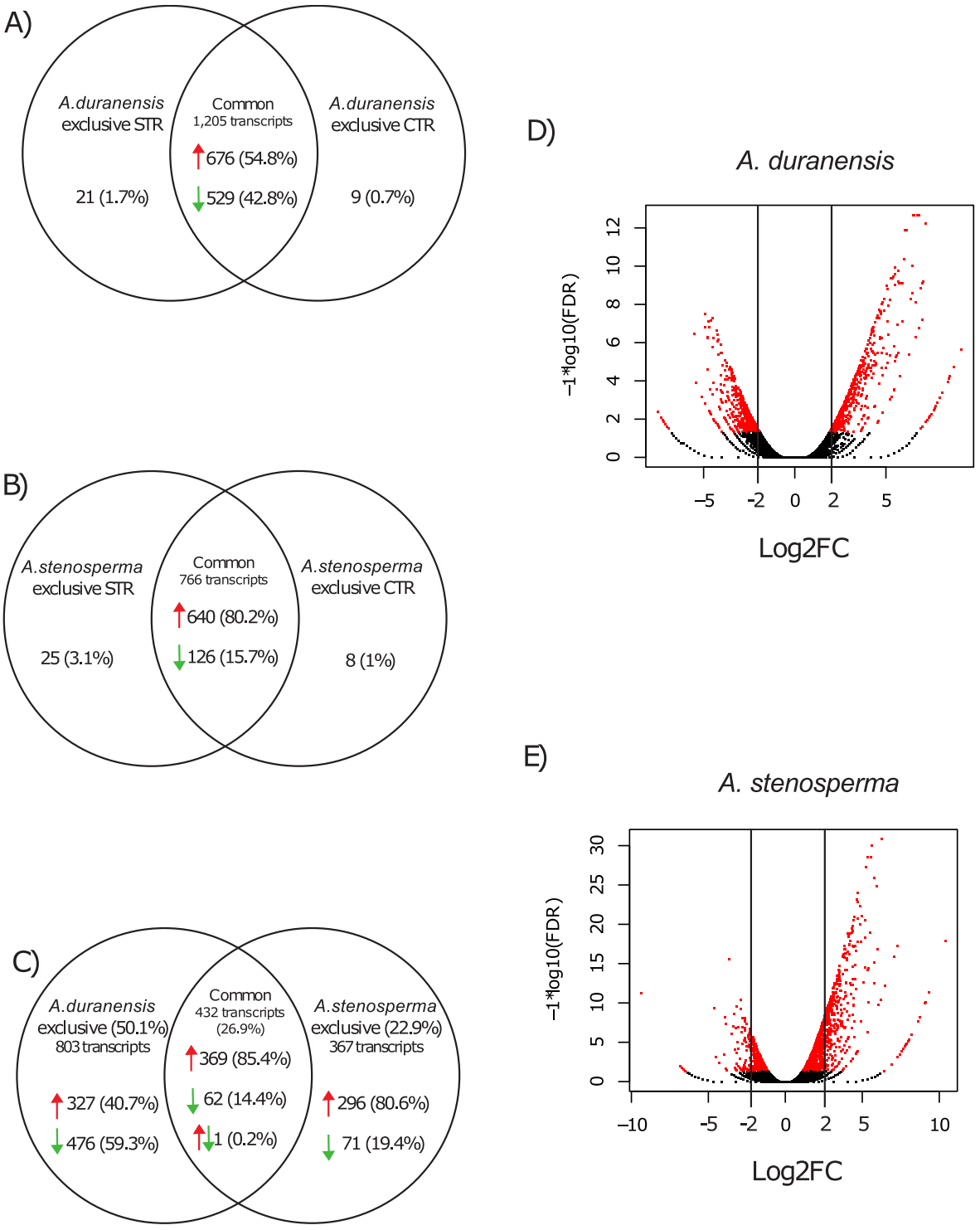



**Fig 2. Expression patterns of *Arachis* spp. transcripts in response to dehydration.** Venn diagrams showing the number of common and exclusive Differentially Expressed Genes (DEGs) in STR and CTR libraries of *A. duranensis* (A); *A. stenosperma* (B); and in both species (C). STR indicates the library from stressed (T25 to T150) samples and CTR the library from control (T0) samples. Red arrows indicate DEGs significantly upregulated during dehydration and green the downregulated DEGs. Volcano plots showing the comparison of gene expression profiles between STR and CTR samples in *A. duranensis* (D) and *A. stenosperma* (E), with each gene represented by one dot (DEGs in red and not significant in black).

https://doi.org/10.1371/journal.pone.0198191.g002

tolerant *Arachis* genotypes was associated with the modulation of two-times more specific transcripts when compared to the contrasting sensitive genotype.

Gene expression profiles *in silico* represented by volcano plots (Fig 2D and 2E) support the idea that the accessions of the two *Arachis* species had a different pattern of expression under dehydration stress. Despite the similar number (1,257 genes for *A*. *duranensis* and 1,725 for *A*. *stenosperma*; Table 1) of significantly expressed genes (FDR$<$ 0.05), after application of the fold change parameter ($Log_2FC > 2$ or $< -2$), it drastically decreased to half for *A*. *stenosperma* (799 genes) whereas it was almost the same for *A*. *duranensis* (1,235 genes; Table 1; Fig 2D and 2E). Besides to have 1.5-times as many DEGs, the overall expression magnitude of genes in response to dehydration is greater in the tolerant genotype than in the sensitive, in particular for downregulated genes.

The chromosomal distribution analysis of the 1,602 DEGs identified in at least one of the species revealed an uneven distribution across all ten chromosomes in the diploid *A*. *duranensis* reference A-genome (Fig 3). Chromosome A03 comprised the largest number of DEGs (13.5%) whilst chromosome A02 had the lowest (6.9%). In accordance with previous studies [15,23], DEGs, as putative protein-coding genes, were more frequent in the distal regions of all *A*. *duranensis* chromosomes, with the gene-rich, abnormally small, chromosome A08 showing the highest density (Fig 3).

## Functional analysis

**Functional comparison of DEGs.** A comparative analysis of the *Arachis duranensis* transcripts with their orthologs from three other legumes species (*Arachis ipaënsis*, *Glycine max*, and *Phaseolus vulgaris*) was conducted in order to identify common Ortholog Groups (OGs). *A*. *ipaënsis* and *A*. *duranensis* are closely related wild species and the most probable ancestors of the cultivated peanut [10], whilst *G*. *max* and *P*. *vulgaris* are major tropical grain legumes which, as is the case for peanut, are produced in drought-prone areas [9]. *A*. *stenosperma* was not included in this OGs analysis since the complete sequence of its genome is not yet available. A total of 17,610 OGs were identified in one or more legume species in which a large core set of proteins, consisting of 10,395 OGs (59%), were common to all four species (Fig 4). Additionally, 3,908 OGs were unique to *Arachis*, whereas only 2,062 OGs were exclusive to the non-*Arachis* species. A high number of OGs were specific to one single species, with *G*. *max* showing more exclusive OGs (814) followed by *A*. *ipaënsis* (696), *A*. *duranensis* (558), and *P*. *vulgaris* (164). The 1,602 DEGs identified in the present study as dehydration-responsive in *A*. *duranensis* and *A*. *stenosperma* (Fig 2C) were distributed within these defined clusters of legume OGs (Fig 4). In accordance with the OGs distribution, a large number of DEGs were shared among all the four species (1,444), corresponding to 90% of the DEGs, which was, surprisingly, higher than the number of DEGs common to *A*. *duranensis* and *A*. *stenosperma* (432; Fig 2C). Approximately 4% of the DEGs (58) were shared between the *Arachis* species whereas a similar number (54) were unique to *A*. *duranensis*. Consistent with the divergence of *Arachis* from the Phaseoloids sub-clade within the legume family [10], the two wild *Arachis* species showed a similar number of OGs and DEGs in common with *G*. *max* (287 and 17, respectively) and with *P*. *vulgaris* (341 and 22, respectively).

**Functional annotation of DEGs.** To assess the enriched Gene Ontology (GO) categories in the biological networks involved in the response to dehydration of both *Arachis* species, GO analysis was performed using the previously identified 1,602 DEGs. Although the comparison between *A*. *duranensis* and *A*. *stenosperma* showed similarities in the most abundant functional categories, important differences among the less frequent categories were found, giving insights into how each species responds specifically towards dehydration (Fig 5). Regarding

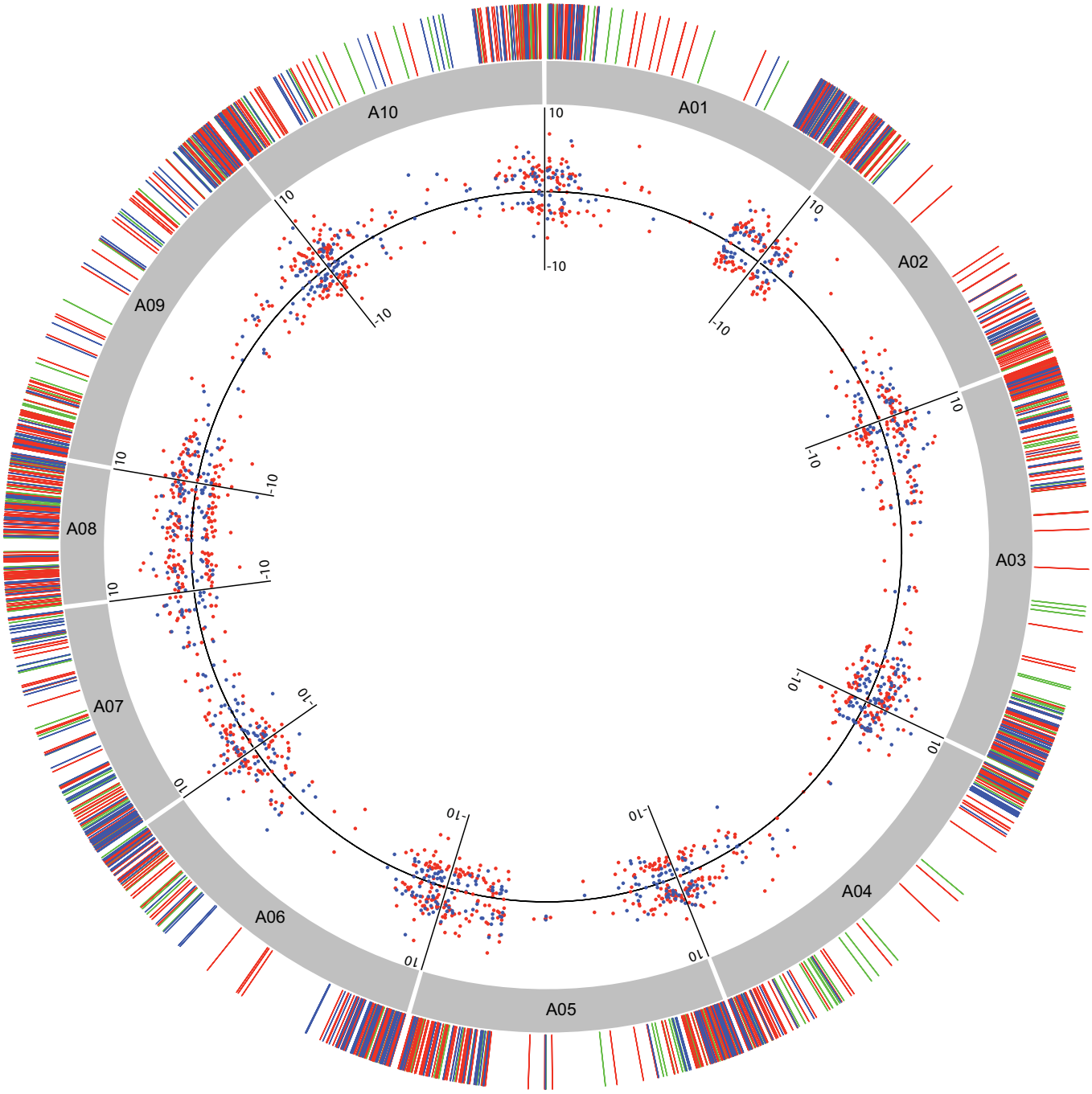



**Fig 3. Circos plot detailing chromosome distribution of DEGs.** Distribution of the 1,602 DEGs in the ten chromosomes of *A. duranensis* (A01 to A10). The outer lines represent exclusive *A. duranensis* (red) and *A. stenosperma* (blue) and common (green) DEGs. The inner dots represent the distribution of Log2FC values for each up- and downregulated DEG in *A. duranensis* (red) and *A. stenosperma* (blue), with the line indicating Log2FC = 0.

https://doi.org/10.1371/journal.pone.0198191.g003

cellular location, GO categories such as "membrane", "extracellular region", and "cell wall" showed similar levels of abundance between the two species, whereas the transcripts of the tolerant species were specific in the "cytosol" and "membrane part" categories and those from the

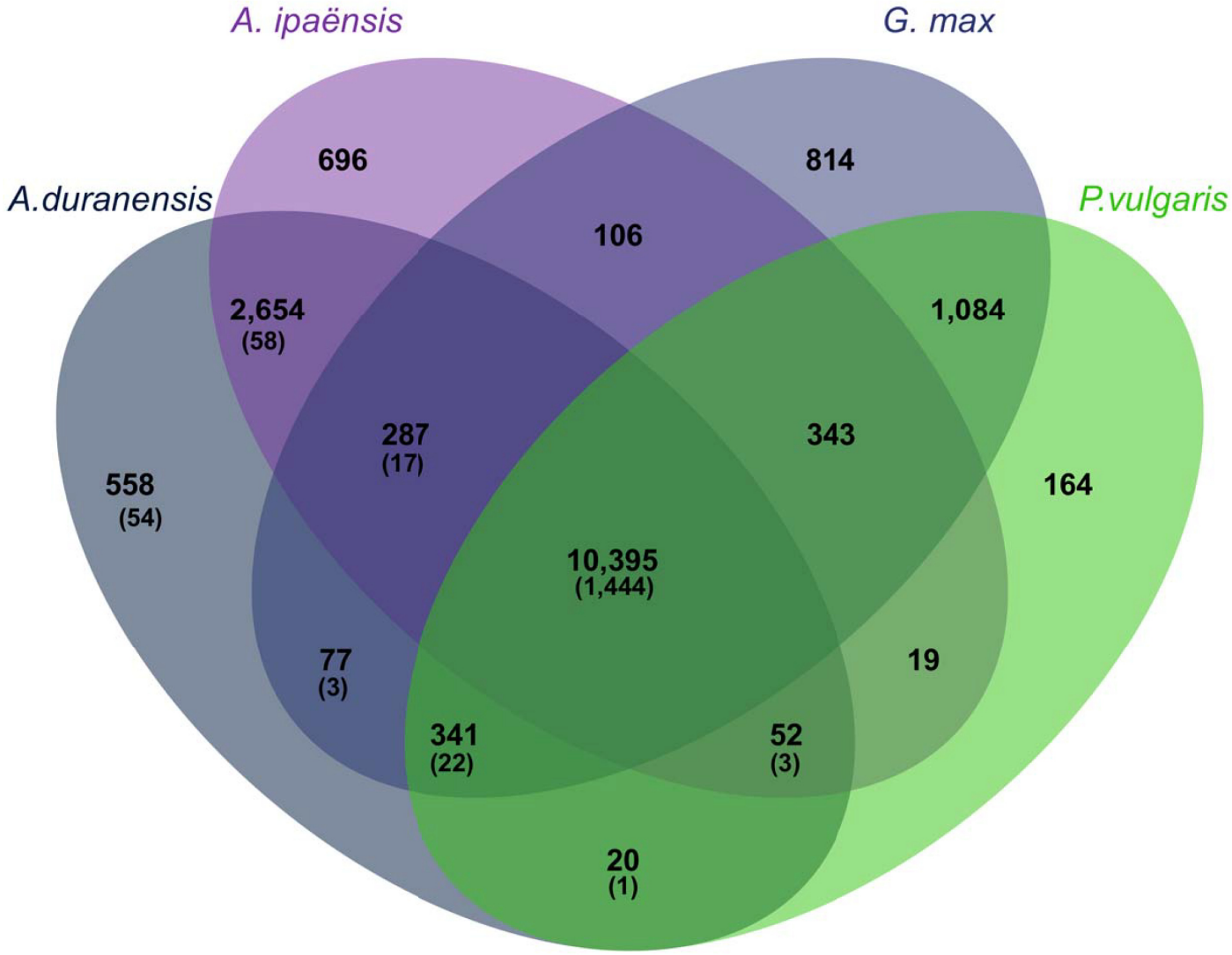



**Fig 4. Clusters of Ortholog Groups (OGs) from four legume species.** Venn diagram indicating the number of Ortholog Groups (OGs) among *Arachis duranensis; Arachis ipaënsis; Glycine max; and Phaseolus vulgaris*. The number of DEGs in the different groups is in parentheses.

https://doi.org/10.1371/journal.pone.0198191.g004

sensitive species were specific in the "cell cortex" and "plant-type vacuole". In the molecular function, both species had most of their transcripts similarly distributed in the "iron, heme- and tetrapyrrole binding", and "catalytic and oxidoreductase activities" categories. Species-specific categories included "transporter, polysaccharide binding, kinase and inositol activities" for *A. duranensis* and "terpene, dioxygenase, lipase, transcription factor and lyase activities" for *A. stenosperma*. In the biological process, the majority of the transcripts were categorized under "oxidation reduction process" and some in the "proline catabolic process" both important processes in plant responses towards dehydration and drought stresses. A specific detail of the biological process categories comparison was that *A. duranensis* had exclusive transcripts in the "response to abiotic stimulus" category, whereas *A. stenosperma* had specific transcripts in the "responses to oxidative stress" category.

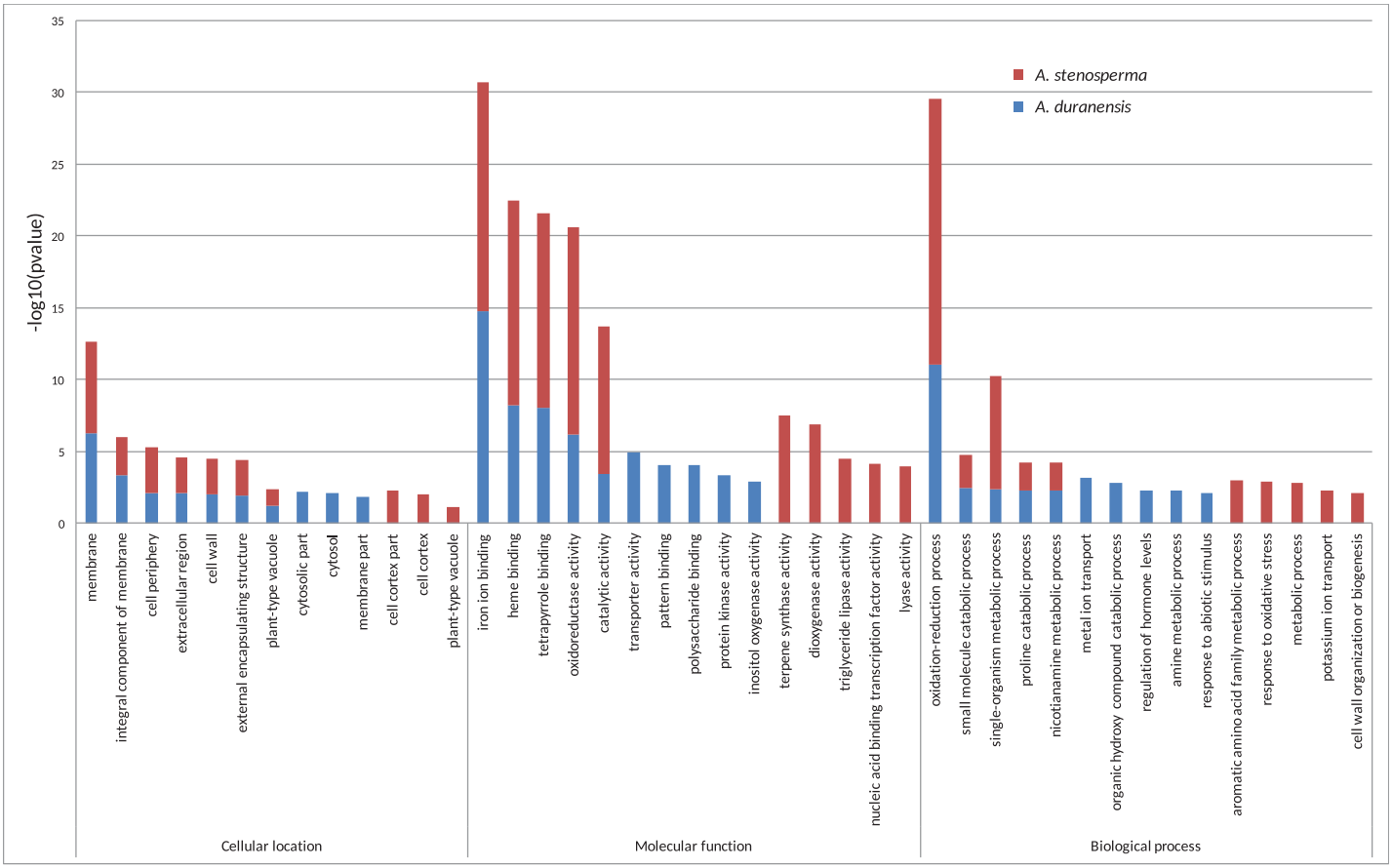



**Fig 5. Gene ontology enrichment of *A. duranensis* and *A. stenosperma* DEGs during dehydration.**

https://doi.org/10.1371/journal.pone.0198191.g005

**Functional analysis of Transcription Factor families.** The temporal regulation of Transcription Factor (TF) expression is a key process in the downstream signaling pathways that initiate protective drought stress responses. Here, the analysis of *A. duranensis* and *A. stenosperma* transcriptome identified 142 DEGs as members of 26 TF families (S2 Fig). Of these, 66 TFs were assigned as exclusive to *A. duranensis*, and 26 to *A. stenosperma* and 50 shared between the species. Genes belonging to the bHLH (BASIC/HELIX-LOOP-HELIX), MYB (MYELOBLASTOSIS), WRKY (WRKY DOMAIN), ERF (ETHYLENE RESPONSIVE FACTOR), and NAC (NAM, ATAF, and CUC) families represented the most abundant TFs that exhibited modulations in transcript levels as part of the response to dehydration, regardless of the *Arachis* species, as previously observed [19,22,41]. Regarding species-specific TF families, seven were exclusive to *A. duranensis*: LBD (LOB DOMAIN), RAV (RELATED TO ABI3/VP1), DBB (DOUBLE B-BOX), ARF (AUXIN RESPONSE FACTOR), SRS (SHI RELATED SEQUENCE), SBP (SQUAMOSAL PROMOTER BINDING PROTEIN), and WOX (WUS HOMEOBOX-CONTAINING). However, only two families were found exclusively in *A. stenosperma*: ZF-HD (ZINC-FINGER HOMEODOMAIN) and NF-YA (NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT ALPHA) (S2 Fig). These species-specific TF-coding genes belong, in general, to novel and less-characterized classes of TFs, and some of them have been described as important regulators in the environmental stimuli and hormone responses, such as LBD, RAV, ARF, SRS, and NF-YA (http://planttfdb.cbi.pku.edu.cn/) [42,43].

## Biological pathways in response to dehydration

To better understand the pathways that may be involved in the differential dehydration responsiveness of the tolerant and sensitive genotypes of *Arachis* species, all 1,235 and 799 DEGs (Table 1; S5 and S6 Tables) identified in *A*. *duranensis* and *A*. *stenosperma*, respectively, were submitted to Mercator and MapMan analyses. The putative role of the products of those transcripts that were significantly up- or downregulated in response to dehydration stress was then visualized for both species (Fig 6). An overview of the general regulation patterns of genes-coding for proteins revealed the major biochemical pathways involved in dehydration responses in the tolerant and sensitive species, including hormone metabolism and signaling; regulation of transcription; oxidative stress protection; and metabolism of carbohydrates and amino acids.

In the drought perception response (Fig 6A), the tolerant genotype of *A*. *duranensis* showed a higher number and more negatively regulated genes coding for receptor kinases than the sensitive genotype of *A*. *stenosperma*. The three signaling pathways activated upon drought perception and mediated by receptor kinases [5,6] are represented in the hormones box (Fig 6A). The responses of the jasmonic acid (JA) and ABA pathways appeared similar in the two species, both in terms of quantity of transcripts and expression pattern. The ET pathway, however, showed a different response, where *A*. *duranensis* had more negatively regulated representatives than *A*. *stenosperma*. Following the signaling cascade, the overall pattern of genes coding for TFs in response to dehydration was similar in the two species, with *A*. *duranensis* exhibiting again more downregulated transcripts. The same was observed for the proteins involved in redox regulation where the drought-tolerant genotype of *A*. *duranensis* also presented more genes with a negative modulation than the drought-sensitive genotype of *A*. *stenosperma*.

In the cellular metabolism categories that were analyzed (Fig 6B), the magnitude of gene expression was generally smaller than those in the stress perception categories (Fig 6A). In carbohydrate and amino acids metabolism, overall expression behavior was similar in the two species with *A*. *duranensis* showing a higher number of regulated transcripts. In particular, in the cell wall category more than twice as many transcripts were regulated by dehydration in *A*. *duranensis* than in *A*. *stenosperma*, being the majority negatively modulated. The exception was in the minor carbohydrate category where the sensitive genotype showed more transcripts regulated than the tolerant.

An overview of the general response to "stress categories" (S3 Fig) showed that most of the DEGs were assigned to signaling, protein degradation, and secondary metabolism, whereas only three categories were exclusive to *A*. *duranensis* (hormone metabolism/salicylic acid, stress biotic receptors, and kinases) and no category was exclusive to *A*. *stenosperma*. Likewise, for metabolism (S4 Fig), 47% were assigned as exclusive to *A*. *duranensis* with none exclusive to *A*. *stenosperma*.

Based on the expression profiling revealed by Mapman analysis (Fig 6) and their functional GO categories, 20 candidate genes exhibiting differential expression behavior between *A*. *duranensis* and *A*. *stenosperma* in response to dehydration were selected for further expression analysis by qRT-PCR. These candidate genes are representative of the main pathways previously identified as being involved in drought and dehydration responses in plants, including genes acting as regulators and effectors (Fig 6). Regulator-coding candidates include: genes involved in the metabolism and signaling of the hormones ABA [*9-CIS-EPOXYCAROTENOID DIOXYGENASE* (*NCED*) and *ABSCISIC ACID 8′-HYDROXYLASE* (*ABA HYDROXYLASE*)]; ethylene [*1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE* (*ACC SYNTHASE*)]; and cytokinins [*ISOPENTENYLTRANSFERASE* (*IPT*)] or involved in signaling and transcriptional regulation as

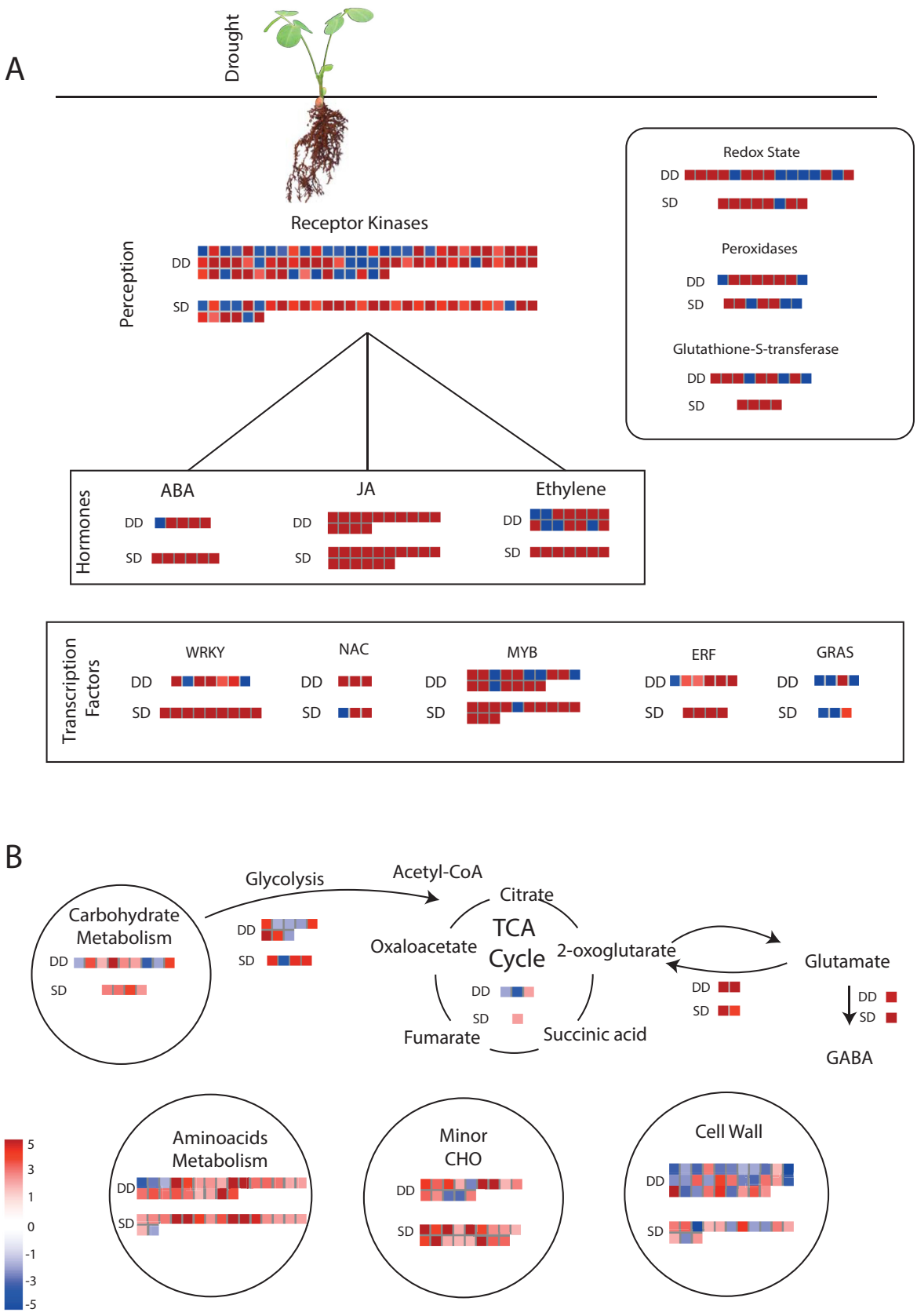

A
Drought
Receptor Kinases
Perception
DD
SD
Redox State
Peroxidases
Glutathione-S-transferase
Hormones
ABA
JA
Ethylene
Transcription Factors
WRKY
NAC
MYB
ERF
GRAS
B
Glycolysis
Acetyl-CoA
Carbohydrate Metabolism
Citrate
TCA Cycle
Oxaloacetate
2-oxoglutarate
Glutamate
Fumarate
Succinic acid
GABA
Aminoacids Metabolism
Minor CHO
Cell Wall
5
3
1
0
-1
-3
-5

**Fig 6. Overview of DEGs expression patterns.** MapMan analysis showing molecular functional categories of DEGs expression patterns in roots of dehydration-stressed plants of *A. duranensis* (DD) and *A. stenosperma* (SD) relative to control. Squares show the different genes encoding proteins related to drought perception and oxidative responses (A) and carbohydrate and amino acid metabolism steps (B). Upregulated genes are indicated by red squares and downregulated by blue squares.

https://doi.org/10.1371/journal.pone.0198191.g006

transcription factors [*DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN* (*DREB*), *WRKY*, and *SCARECROW*), ion transporter [*PLASMA MEMBRANE CALCIUM ATPASE* (*PMCA*)], and protein kinases (*KINASE* and *MITOGEN*). Effector-coding candidates include: genes associated with osmoregulation processes as sugar starvation [*ASPARAGINE SYNTHETASE* (*ASN SYNTHETASE*) and *FATTY ACID DESATURASE* (*FAD*)]; carbohydrate metabolism [*GALACTINOL SYNTHASE* (*GOLS*) and *PLASMA MEMBRANE MANNITOL TRANSPORTER* (*MAT*)], and cell wall biosynthesis and growth [*EXPANSIN A* (*EXPA*) and *AQUAPORIN*). In addition, other effector genes are involved in general abiotic stress responses, such as macromolecular protection [*DEHYDRIN* (*DHN*)] and senescence [*CYSTEINE PROTEASE* (*CYS PROTEASE*), *SAP*, and *F-BOX/KELCH-REPEAT PROTEIN* (*F-BOX*) were also analyzed.

## Validation of expression profile by qRT-PCR

In order to validate the RNA-Seq data, qRT-PCR analysis was carried out with the 20 selected candidate genes. All genes showed specificity of transcript amplification with high amplification efficiencies (S1 Table). Although all the primers were designed based on the sequence of *A. duranensis* gene-models, the majority also efficiently amplified *A. stenosperma* samples and produced single-peaked melting curves. The exceptions were for five genes (*ACC SYNTHASE*; *DREB*; *SCARECROW*; *MAT*; and *AQUAPORIN*) which appeared to be species-specific (Fig 7). In the case of the *SCARECROW* and *MAT* genes, they did not show significant FDR ($< 0.05$) when comparing the stressed to the control samples. Overall the high levels of similarity between *A. duranensis* and *A. stenosperma* orthologs emphasize the close association between AA genomes of *Arachis* [10,25]. The *in vitro* expression patterns of the 20 candidate genes estimated by qRT-PCR was consistent with those predicted by *in silico* analysis for both species using pooled CTR and STR samples (Fig 7), except for the *NCED* and *DREB* genes. Also, in accordance with the *in silico* analysis, *SAP* was the only gene which showed a different expression profile between the two species in qRT-PCR analysis, being upregulated in *A. duranensis* and downregulated in *A. stenosperma* (Fig 7). In addition to their almost identical expression behaviors, most genes presented very similar *in vitro* and *in silico* fold change magnitudes. This indicates that both *in vitro* and *in silico* analyses resulted in comparable expression profiles in wild *Arachis* and reinforced the validity of estimating gene abundance using RNA-Seq data with posterior validation using more thorough spatial and temporal qRT-PCR analysis [19,21]. Accordingly, based on the *A. hypogaea* expression atlas [44], weak or no basal expression of the 20 *A. duranensis* candidate genes homologs could be observed in non-stressed peanut roots (S8 Table). Considering that cultivated peanut (*A. hypogaea*) and wild *A. stenosperma* are more sensitive to drought than *A. duranensis* [24], and that the 20 selected genes showed a contrasting expression behavior between the two wild genotypes, it could be suggested that these candidates contribute, to different extents, to the tolerance of *A. duranensis* to dehydration stress.

In order to track changes in expression of the candidate genes throughout the development of the dehydration stress, qRT-PCR was carried out in the tolerant *A. duranensis* genotype, for all time points following the beginning of the treatment, i.e., T25, T50, T75, T100, T125, and T150. In this point-by-point analysis, the expression of the majority of candidate genes in response to the dehydration was in accordance with the previous *in silico* and *in vitro*

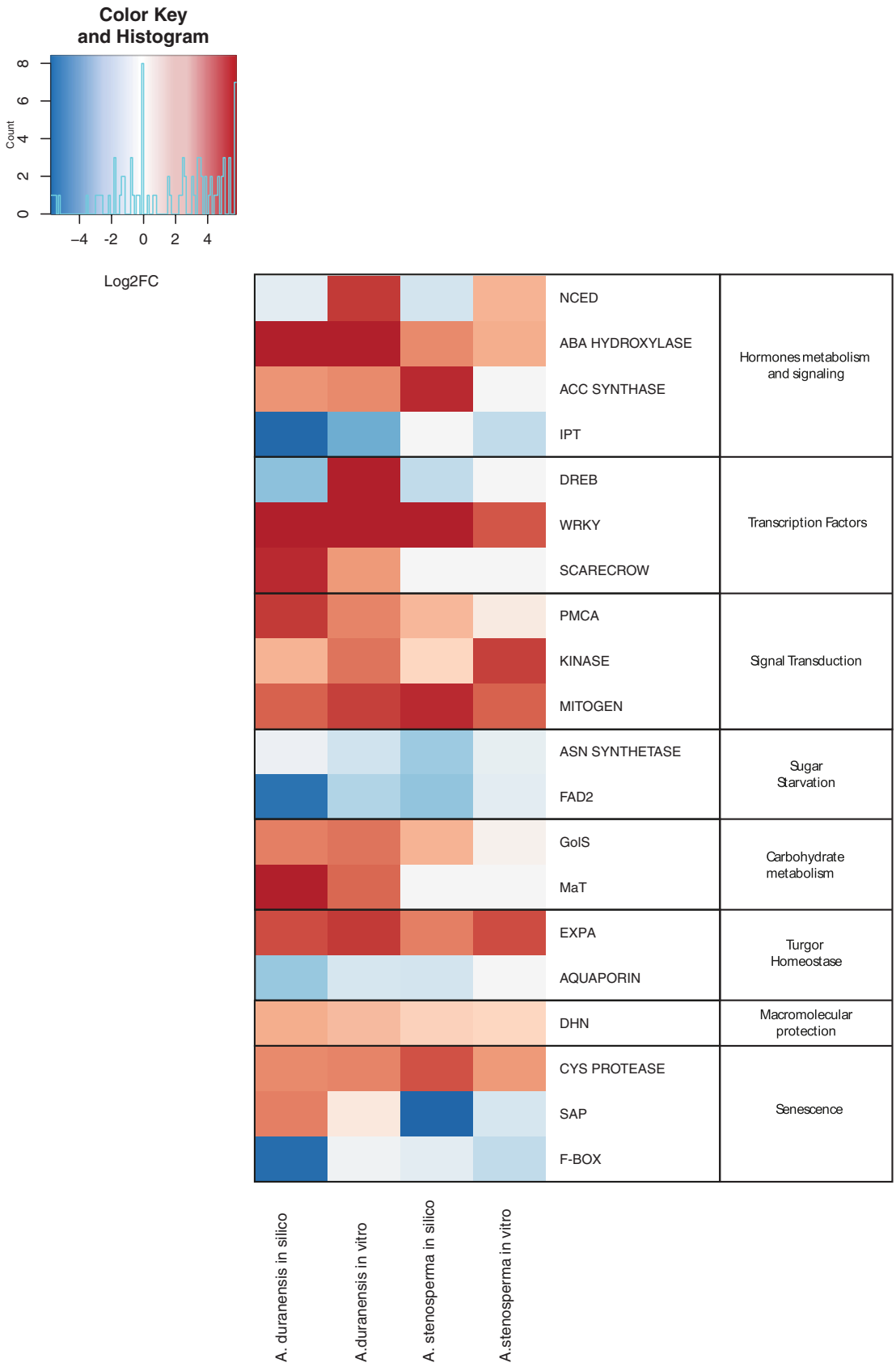

Color Key
and Histogram
Count
0
2
4
6
8
−4
-2
0
2
4
Log2FC
NCED
ABA HYDROXYLASE
ACC SYNTHASE
IPT
DREB
WRKY
SCARECROW
PMCA
KINASE
MITOGEN
ASN SYNTHETASE
FAD2
GolS
MaT
EXPA
AQUAPORIN
DHN
CYS PROTEASE
SAP
F-BOX
Hormones metabolism and signaling
Transcription Factors
Signal Transduction
Sugar Starvation
Carbohydrate metabolism
Turgor Homeostase
Macromolecular protection
Senescence
Regulators
Efectors
A. duranensis in silico
A.duranensis in vitro
A. stenosperma in silico
A.stenosperma in vitro

**Fig 7. Heatmap of the relative expression of the 20 candidate genes in *Arachis duranensis* and *A. stenosperma* after *in silico* (RNA-Seq data) and *in vitro* (qRT-PCR) analyses.** Normalized log2FC values are shown in a red–blue scale, with darker red the most upregulated and the darker blue the most downregulated DEGs.

https://doi.org/10.1371/journal.pone.0198191.g007

expression analyses conducted with pooled samples (Figs 7 and 8). Five candidates (*IPT*, *SCARECROW*, *FAD*, *MAT*, and *F-BOX*), for which the general expression profile did not agree with the previous results, were not considered for further analysis. The expression profiles of the three candidate genes involved in hormone metabolism and signaling (*NCED*; *ABA HYDROXYLASE* and *ACC SYNTHASE*) were quite similar, showing a gradual increase in transcript levels through the 150 min of the dehydration (Fig 8A–8C). In particular, the *ABA HYDROXYLASE* gene (Fig 8B), involved in ABA catabolism, was remarkably responsive to water deficit in *A. duranensis* roots, with an upregulation above 72-fold at the end of the treatment (125 and 150 min). As expected for transcriptional regulons, the two TF-coding genes were highly induced in response to the imposed dehydration stress. The *DREB* gene showed a strong upregulation of 178-fold immediately upon stress perception (25 min) followed by a rapid decline in its expression (Fig 8D), whereas the *WRKY* gene reached the peak of its expression (15-fold) at 75 min and maintained this level until the end of the treatment (Fig 8E). Similarly, other candidate genes involved in signal transduction also displayed an increase in their expression in response to dehydration, reaching the highest expression at the end (*PMCA*; Fig 8F) or the beginning (*KINASE* and *MITOGEN*; Fig 8G and 8H, respectively) of the treatment. The opposite profile, consisting of the repression of gene expression as a consequence of water withdrawal, was observed for only two genes (*ASN SYNTHETASE* and *AQUAPORIN*; Fig 8I and 8J, respectively). The remaining candidate genes, involved in general cell protection and adaptive mechanisms of drought response (*GOLS*, *DHN*, *EXPA*, *CYS PROTEASE*, and *SAP*) showed a similar overall profile, with a gradual induction as the dehydration progressed with a peak at the middle of the treatment followed by a decrease in their expression at the end (Fig 8K–8O).

## Discussion

Cultivated peanut (*A. hypogaea*), an allotetraploid AABB-genome, typically exhibits greater drought sensitivity than its wild diploid AA-genome relatives, possibly due to differences in morphological and physiological features relevant to drought adaptation [24]. Due to their varied levels of drought tolerance, wild relatives may offer novel genes and alleles that can be of use in the improvement of the drought tolerance of cultivated peanut [10–12,45,46]. In order to understand the multiplicity of transcriptional modulation in response to drought, two wild diploid species, *A. duranensis* and *A stenosperma*, were selected. These genotypes of the two species exhibit contrasting transpirational behavior, with *A. duranensis* showing a conservative transpiration strategy to cope with the water-limited conditions (considered tolerant) whereas *A. stenosperma* exhibits a less conservative transpiration behavior (considered sensitive), very close to that of the drought-sensitive peanut [24]. Changes in *Arachis* ploidy alter plant architecture, leaf morphology, and cell size, among other traits, and since these features are all involved in plant responses to drought imposition [24,47], here comparisons were made between species of the same ploidy.

In the present study, the removal of nutrient solution was used as an alternative model to simulate drought perception in *Arachis* and to facilitate rapid, uniform and easy root sampling, as previously validated in a number of different species [27,48–50]. The physiological analyses conducted at the early stages of the dehydration treatment indicated differences in leaf gas exchange between the two species, with *A. stenosperma* exhibiting a lower stomatal

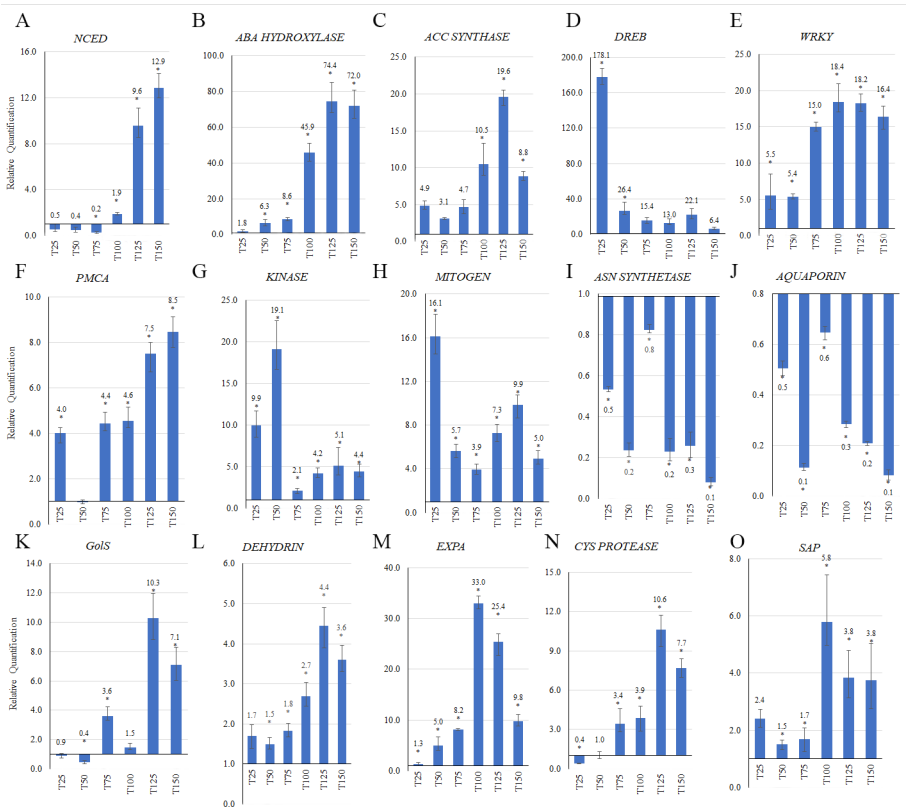



**Fig 8. Expression of DEGs as determined by qRT-PCR.** Relative quantification of mRNA levels of 15 candidate genes in *A*. *duranensis* roots during the dehydration treatment and collected after 25 (T25); 50 (T50); 75 (T75); 100 (T100); 125 (T125) and 150 (T150) min, relative to control (T0). Bars represent the standard deviation of three biological replicates. Significantly ($P < 0.05$) up- or downregulated genes are indicated by asterisks.

https://doi.org/10.1371/journal.pone.0198191.g008

conductance than *A*. *duranensis* that is normally associated with a lower net photosynthetic rate and transpiration. This lower rate of transpiration is likely to be responsible for the slightly greater leaf temperature of *A*. *stenosperma*. A previous study reported the opposite trend for the same species [51], though it is worth noting that in that study the differences in gas exchange parameters between two *A*. *duranensis* accessions were greater than those between *A*. *duranensis* and *A*. *stenosperma*. Moreover, whilst one of the *A*. *duranensis* accessions (K7988) in that study was the same as that used here, the *A*. *stenosperma* accessions were different (accessions SV2411 and V10309), possibly also contributing to the difference in gas exchange parameters between the two studies.

The similarity between *A*. *duranensis* and *A*. *stenosperma* genomes allowed the use of the *A*. *duranensis* genome [15] as a reference to the mapping of the raw reads of *A*. *stenosperma* transcriptome data. The low difference of percentage of reads mapping between the species confirmed that they are very close in terms of genome sequence and that it is possible to use *A*. *duranensis* as a reference genome to *A*. *stenosperma*. Twenty candidate genes identified in the transcriptome of *A*. *duranensis* and *A*. *stenosperma* as differentially expressed (DEGs) in response to the dehydration treatment were selected based on both the *in silico* analyses and their putative biological function. The expression profiles of these candidates obtained by qRT-PCR corroborated *in silico* analysis. Additionally, the comparative expression analyses between the drought-tolerant genotype of *A*. *duranensis* and the sensitive genotype of *A*. *stenosperma* showed that, in general, dehydration responsiveness, including the DEGs expression

magnitude, was greater in the tolerant genotype. In accordance with the physiological data, molecular evidence also indicated differences between the two species. *A*. *duranensis* exhibited, in general, a better adaptive response to water deficit at the molecular level than *A*. *stenosperma* in terms of activation of osmosensing cascades; control of hormone and osmolyte content; and protection of macromolecules. Since each candidate gene codes proteins that regulate distinct steps of biochemical pathways, putative roles in the mechanisms of receptor and effector responses and adaptation to dehydration stress in wild *Arachis* are discussed in more detail below.

## Regulator responses upon water deficit perception

Upon the perception of water deficit by plant cells, specific receptors trigger immediate molecular responses, in a process known as osmosensing. Activation of "osmosensor receptors" leads to the regulation (initiation or suppression) of a cascade of responses through signal-transducing pathways that in turn modulate the expression of genes involved in drought responses [7]. The signal sent by the osmosensor receptors is received by protein kinases (e.g., calcium-dependent kinases, CDPK, and mitogen activated kinases, MAPK) that catalyze protein phosphorylation and dephosphorylation [52]. In the present study, many receptor kinases were activated by the dehydration treatment, with differential modulation between *A*. *duranensis* and *A*. *stenosperma* (Fig 6A). Of these, two representatives of kinase-coding genes (*MITOGEN* and *KINASE*) were upregulated at all time points of the dehydration experiment. Indeed, several receptor kinases have been identified in plants subjected to water stress and shown to be involved in osmosensing process upon drought perception for the transduction of dehydration signals [7].

Besides kinases, the activation of "osmosensor receptors" upon water-deficit perception also modulates intracellular levels of calcium which in turn acts as a second messenger in a network of signalling events involved in the induction of stress responses [52]. The *PMCA* gene, coding for a calcium-ATPase present in the plasma membrane, is responsible for the active efflux of calcium from the cytosol in order to restore its low intracellular levels after the signalling events and maintain ionic homeostasis [53]. Since increases in cytosolic calcium are detrimental to plants, PMCA plays crucial regulatory roles in tolerating drought stress by activating the ABA-signalling pathway. Accordingly, we observed an induction of the *PMCA* gene under dehydration stress indicating the role of this calcium pump in the modulation of stress signalling and ionic homeostasis in *Arachis*.

Following the osmosensing process, several genes involved in ABA-dependent signaling responded to dehydration stress in this study (Fig 6), and two of the genes in this pathway were selected for point-by-point qRT-PCR analysis: *NCED* and *ABA HYDROXYLASE*. NCED catalyzes a main regulatory step in ABA synthesis by converting the epoxycarotenoid precursor to xanthoxin in the plastids that is in turn converted to ABA [54,55]. Once synthesized under water deficit conditions, ABA activates a number of ABRE-containing effector genes. Amongst several drought-related effects, ABA increases the hydraulic conductivity in root cells, directs the investment of photoassimilates towards root growth and induces stomatal closure [56]. Moreover, the overexpression of heterologous *NCED* genes in transgenic plants [56,57] demonstrated that drought tolerance could be improved through the modulation of ABA levels by this key regulatory gene. In accordance with its role in drought tolerance, in our study *NCED* gene expression levels in the tolerant genotype of *A*. *duranensis* were twice those in the sensitive genotype of *A*. *stenosperma* in response to dehydration, with a gradual increase in expression detected as the stress progressed.

The balance between biosynthesis, catabolism, conjugation and transport determines ABA content in plant organs rapidly decrease when the plants are recovering from abiotic stresses

[5] and this hormone is inactivated either by conjugation or oxidation, the latter of which is catalyzed by ABA 8-HYDROLASE. Here, the *A*. *duranensis ABA HYDROLASE* gene was immediately upregulated following the beginning of the dehydration treatment and showed the highest change in expression (>72-fold) of all of the candidate genes at the end of the treatment. Previous studies in citrus plants also indicated induction of this gene under drought and combined high temperature and drought treatments and, as we also observed in *Arachis*, this induction occurred simultaneously with the induction of NCED [58]. As a result of the need for a balance between ABA synthesis and catabolism, if the expression of *NCED* is higher in the tolerant *A*. *duranensis* higher expression of *ABA HYDROLASE* might also be anticipated, as was previously observed in sheepgrass [55].

Members of many transcription factor (TF) families, including MYC, MYB, bZIP, NAC, WRKY, and DREB, are regulatory proteins involved in signal transduction and regulation of gene expression in response to different abiotic stresses in both ABA-dependent and ABA-independent cascades, that are linked by crosstalk signaling pathways [5,52,59]. The genes belonging to the DREB (Dehydration-Responsive Element Binding) TF family play an important role in the ABA-independent abiotic stress-tolerance pathways that involve the regulation of genes possessing a conserved DRE or CRT motif in their promoters [52,60]. Here, a DREB-coding gene displayed the strongest upregulation (178-fold) among candidates in *A*. *duranensis* roots at early stages of dehydration (T25), as would be expected for a TF involved in the induction of drought-tolerance responses. Additionally, heterologous overexpression of *DREB* genes from diverse origins in transgenic plants is often described as conferring increased tolerance to drought, salt, freezing, and other abiotic stresses [60].

The hormone ethylene also plays a role in modulating plant responses to abiotic stresses and acts together with ABA-signaling cascades to determine how plant organs respond to drought [6]. Recent studies indicated that ethylene signaling might be implicated in both stomatal opening and closure under stress conditions [61], The upregulation of the gene coding for ACC SYNTHASE, a key step in the production of ethylene, was observed in tolerant *A*. *duranensis* plants at all time points along the dehydration treatment. Interestingly, *Arabidopsis* and rice mutants containing elevated ethylene levels showed increased tolerance to drought during the formation of panicles [61,62]. Together, these results suggested that ethylene production mediated by ACC SYNTHASE may be important in the determination of *A*. *duranensis* drought tolerance.

As in the case of ACC SYNTHASE, the activity of TFs belonging to the WRKY family is also modulated by both ABA- and ethylene-signaling pathways, being part of the crosstalk between these two hormones in the complex multicomponent networks of drought responses [5,63]. The WRKY-coding genes are also involved in the downstream transcriptional regulation of ABA-responsive element binding factors (ABFs/AREBs) through W-box sequences present in the promoters of the bZIP-type ABFs/AREBs TF family [63]. Here, *in silico* and *in vitro* expression analyses indicated that a *WRKY* gene was induced upon dehydration perception in both *Arachis* species, as would be expected for a regulator responsible for the induction of drought-responsive genes [5,63]. However, as for the other candidates involved in hormone signaling responses, *NCED*, *ABA HYDROXYLASE*, *PMCA*, *DREB* and *ACC SYNTHASE*, the genes in the tolerant *A*. *duranensis* had a higher magnitude of expression than their orthologs in the sensitive *A*. *stenosperma*.

## Effector responses to dehydration stress

The signaling events described above that occur upon the initial perception of water scarcity and activation of regulators ultimately lead to the expression of effector genes in *Arachis*

including those associated with osmoregulation processes, macromolecular protection, and senescence (Fig 6). Under water deficit conditions, the interaction of cationic and anionic amphiphilic substances with membranes results in changes in their physical state or in the protein-lipid interactions that relay osmosensing by cells [7]. Significantly, changes in the physical state of membranes may also regulate the activity of major proteins involved in the plasticity and osmotic balance of the plasma membrane-cell wall system. Among them, Expansins and Aquaporins play essential roles in the control of cell volume and turgor homeostasis under drought stress [23,64]. In the present study, many genes coding for cell wall modifying proteins were activated in response to dehydration and an *EXPANSIN A (EXPA)* was amongst the most highly upregulated of the genes analyzed by qRT-PCR (Fig 8). These results support our previous hypothesis that wild *Arachis* Expansins are common and crucial players in responses to multiple and simultaneous stresses, including drought, ultraviolet light exposure, and nematode infection [19,21,23,65]. In contrast, *AQUAPORIN* transcripts were downregulated in all expression analyses conducted on both species. Repression of *AQUAPORIN* expression in *A. duranensis* after drought perception, concomitant with the induction of key enzymes in hormone biosynthesis, as observed here for NCED, ABA HYDROXYLASE and ACC SYNTHASE, agreed with previous studies showing that the inhibition of AQUAPORIN activity is highly correlated with ethylene and ABA accumulation when roots are directly exposed to water stress [64,66]. In accordance, heterologous overexpression of certain AQUAPORINS in different transgenic plants led to an enhanced ability to withstand water stress conditions [67].

In addition to the control of turgor pressure via alterations to the membrane-cell wall system under dehydration conditions, plants will also accumulate sugars and other osmolytes in order to adjust their osmotic potential. This adjustment is accompanied by the influx of water into cells, or at least reduced efflux, resulting in maintenance of the turgor necessary for cell integrity and survival [6,7,55]. In the present study, expression of many genes involved in sugar metabolism was activated during the 150 min of dehydration stress in both *Arachis* species. *GOLS* was selected as a candidate gene representative of the carbohydrate metabolism category, and, and although positively regulated in both species, showed a higher magnitude in the tolerant *A. duranensis*. GOLS catalyzes the formation of galactinol from UDP-galactose and *myo*-inositol and has been extensively studied as part of plant adaptive responses to handle dehydration stress. Once synthesized, galactinol is in turn used as a galactosyl donor to form the raffinose series of oligosaccharides (raffinose, stachyose, and verbascose), which may act as osmoprotectants during drought stress, but also to protect cellular membranes during seed desiccation. Indeed, there are encouraging examples of drought tolerance improvements in transgenic plants overexpressing GOLS [68,69].

As part of the overall macromolecular protection mechanisms of plants to avoid water loss during drought stress, a gene coding for an SK2-type of DHN was also selected as a candidate. DHNs have multiple protective roles in plants subject to abiotic stress, acting as chaperones, antifreeze proteins, cryoprotectants, free radical-scavengers, and ion-binding proteins in order to retain water, maintain cell turgor, and ultimately function of cellular components [70,71]. Accordingly, DHNs accumulate under various abiotic stresses that cause cell dehydration, including drought and desiccation, in processes controlled by the ABA-dependent signaling pathway. Here, a SK2-type DHN was upregulated in all analyses and both species, but again with a higher magnitude in the tolerant *A. duranensis* genotype. This induction was maintained during the whole dehydration experiment and could indicate that DHNs have a different role in drought adaptation in these contrasting species, as previously suggested for sheepgrass [55]. Numerous reports reveal that the overexpression of DHN-coding genes might increase plant tolerance to abiotic stresses, particularly drought, cold, and salinity,

characterized by the reduction in the osmotic potential and lipid peroxidation, and accumulation of proline, sugar and potassium ions [70,71].

In addition to reduced plant growth, water deficit often leads to stomatal closure that, under severe stress conditions, may in turn lead to sugar starvation and senescence. ASPARAGINE (ASN) SYNTHETASE activity is dependent on the carbon/nitrogen status of the cell and is differently affected by various environmental stresses that may affect the requirement of asparagine for nitrogen mobilization and/or storage [72]. The regulation of plant ASN SYNTHETASE-coding genes by diverse stresses suggests convergent signal perception and transduction related to sugar-starvation conditions [73]. In our study, a decrease of *ASN SYNTHETASE* transcripts was observed in *A*. *duranensis* roots in response to dehydration stress. Since peanut is known to transport significant quantities of assimilated nitrogen in the form of asparagine [74], the decrease in *ASN SYNTHETASE* transcript levels may, therefore, reflect a prioritization under water-limited conditions of other physiological processes in wild *Arachis*, apart from nitrogen metabolism.

From the various senescence-related genes regulated during the 150 min of dehydration in both wild *Arachis* species (Fig 7) two candidates were further analyzed: those coding for a CYSTEINE (CYS) PROTEASE and a SENESCENCE-ASSOCIATED PROTEIN (SAP). CYS PROTEASEs are among the best-characterized plant proteases and known to be strongly expressed upon abiotic stress perception, being ultimately involved in programmed cell death [75]. Being a functional senescence-associated gene, CYS PROTEASE activity is normally further blocked by a protease inhibitor in drought-tolerant plants to prevent stress-related premature protein degradation. As expected, herein, a *CYS PROTEASE* gene was upregulated in both analyses and in both *Arachis* species. Accumulation of *CYS PROTEASE* transcripts was also observed in a drought-tolerant peanut cultivar under severe drought stress, whereas in the sensitive cultivar, accumulation occurred under only moderate stress [76]. Likewise, the *SAP* gene has a similar expression behavior, being induced in the drought-tolerant *A*. *duranensis* and strongly repressed in the drought-sensitive species during the imposition of dehydration. The majority of the characterized drought-induced *SAP* genes code for proteases, nucleases, and other enzymes involving nutrient recycling and stress responsive regulators, that are regulated via the ABA-dependent signaling pathway in association with other phytohormones [77]. However, most of the senescence studies refer to leaf senescence, and little is known about changes in senescent roots from a molecular, cellular and physiological perspective. Furthermore, efficient remobilization of amino acids and nutrients and the control of hydrolytic processes, by proteins such as ASN SYNTHETASE, CYS PROTEASE and SAP, could be crucial for the plant survival and reduction in water loss at whole-plant level, under water-limited conditions [6,75,77].

## Conclusions

Drought stress is a complex issue that involves an intricate chain orchestrated by physiological events affecting transcriptional and translational regulation. Different plant genotypes often harbor distinct mechanisms of response and adaptation to drought, which may include the expression of a particular set of genes that trigger changes in specific biochemical pathways. Following stress signaling, the regulators will signal towards diverse downstream pathways, which involve the action of effectors in osmoregulation, stress signaling, and macromolecule protection. In this comparative transcriptome study, many regulators and effectors were identified as being differentially regulated in two contrasting genotypes of wild *Arachis* species during dehydration. As these genes have different roles in the perception of water scarcity and subsequent establishment of stress tolerance, they might play a role in the tolerant genotype of

*A. duranensis* earlier and/or faster response to dehydration stress in comparison with the sensitive genotype of *A. stenosperma*. In this manner we were able to explore the nature of drought tolerance in wild species and discover potentially useful wild alleles related to adaptability to adverse environments. The recent advances in our understanding of *Arachis* genome structure, together with functional genomic integrative approaches, has enabled new insights into gene occurrence, evolution, function, and regulation, as well as their importance in the control of desirable agronomical traits. The findings in this study will increase our understanding of the transcriptional networks that control the responses to water deficit in wild *Arachis* and will broaden the use of their extensive genetic diversity for the improvement of cultivated peanut, and other legumes, by either introgression via interspecific hybridization or transgenic approaches.

## Supporting information

**S1 Fig. *A. duranensis* 30-day-old plants at the beginning of the dehydration treatment.** A) aerial part and B) roots of *A. duranensis* plants at T0 when the nutrient solution was removed.
(TIF)

**S2 Fig. Distribution of exclusive and common Transcription Factors families in *A. duranensis* and *A. stenosperma* subjected to dehydration.**
(EPS)

**S3 Fig. Number of transcripts assigned to "stress" categories of MapMan in *A. duranensis* (blue) and *A. stenosperma* (orange).**
(EPS)

**S4 Fig. Number of transcripts assigned to metabolism categories of MapMan in *A. duranensis* (blue) and *A. stenosperma* (orange).**
(EPS)

**S1 Table. Candidate genes and primers used for qRT-PCR analysis.**
(DOCX)

**S2 Table. Physiological measurements.** Physiological measurements made on *A. duranensis* and *A. stenosperma* leaves during the dehydration treatment, using a portable Photosynthesis System (LI-COR, 150 model LI-6400).
(XLSX)

**S3 Table. Gene models identified in the transcriptome of *Arachis duranensis*.** Gene models (21,125) identified from reads of *A. duranensis* mapped to the reference *A. duranensis* genome, with their respective expression values (Log2FC and FDR) and functional annotation.
(XLSX)

**S4 Table. Gene models identified in the transcriptome of *Arachis stenosperma*.** Gene models (21,502) identified from reads of *A. stenosperma* mapped to the reference *A. duranensis* genome, with their respective expression values (Log2FC and FDR) and functional annotation.
(XLSX)

**S5 Table. Pairwise comparison between STR and CTR samples of *Arachis duranensis*.** Differentially Expressed Genes (1,235 DEGs) in pairwise comparison between STR (stressed) and CTR (control) samples of *A. duranensis*, with their respective expression values (Log2FC and FDR).
(XLSX)

**S6 Table. Pairwise comparison between STR and CTR samples of *Arachis stenosperma*.** Differentially Expressed Genes (799 DEGs) in pairwise comparison between STR (stressed) and CTR (control) samples of *A*. *stenosperma*, with their respective expression values (Log2FC and FDR).
(XLSX)

**S7 Table. Pairwise comparisons between samples of *Arachis duranensis* and *Arachis stenosperma*.** Differentially Expressed Genes (1,602 DEGs) in pairwise comparison between samples of *Arachis duranensis* and *A*. *stenosperma*, with their respective expression values (Log2FC and FDR).
(XLSX)

**S8 Table. Expression of *A*. *duranensis* DEGs homologs in peanut.** Expression profile of 20 *A*. *duranensis* candidate genes homologs in the *A*. *hypogaea* (cv. Tifrunner) expression atlas from the eFP Browser (http://bar.utoronto.ca/efp_arachis/cgi-bin/efpWeb.cgi).
(XLSX)

## Author Contributions

**Conceptualization:** Christina Cleo Vinson, Soraya Cristina Macedo Leal-Bertioli, Alexandre Lima Nepomuceno, Patricia Messenberg Guimaraes, Ana C. M. Brasileiro.

**Data curation:** Christina Cleo Vinson, Ana Paula Zotta Mota.

**Formal analysis:** Christina Cleo Vinson, Ana Paula Zotta Mota, Thais Nicolini Oliveira, Larissa Arrais Guimaraes, Thomas Christopher Rhys Williams, Alexandre Lima Nepomuceno, Mario Alfredo Passos Saraiva, Ana C. M. Brasileiro.

**Funding acquisition:** Soraya Cristina Macedo Leal-Bertioli, Ana Claudia Guerra Araujo, Patricia Messenberg Guimaraes, Ana C. M. Brasileiro.

**Investigation:** Thais Nicolini Oliveira, Larissa Arrais Guimaraes, Mario Alfredo Passos Saraiva.

**Methodology:** Ana Paula Zotta Mota.

**Supervision:** Patricia Messenberg Guimaraes, Ana C. M. Brasileiro.

**Validation:** Christina Cleo Vinson.

**Writing – original draft:** Christina Cleo Vinson, Ana Paula Zotta Mota, Thomas Christopher Rhys Williams, Patricia Messenberg Guimaraes, Ana C. M. Brasileiro.

**Writing – review & editing:** Christina Cleo Vinson, Ana Paula Zotta Mota, Larissa Arrais Guimaraes, Soraya Cristina Macedo Leal-Bertioli, Thomas Christopher Rhys Williams, Ana Claudia Guerra Araujo, Patricia Messenberg Guimaraes, Ana C. M. Brasileiro.

## References

1. Lawrence D, Vandecar K. Effects of tropical deforestation on climate and agriculture. Nat Clim Chang. 2015; 5: 27–36.
2. Shanker AK, Maheswari M, Yadav SK, Desai S, Bhanu D, Attal NB, et al. Drought stress responses in crops. Funct & Integr genomics. 2014; 14: 11–22.
3. Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA. Plant drought stress: effects, mechanisms and management. In: Lichtfouse E, Navarrete M, Debaeke P, Véronique S, Alberola C. eds. Sustainable Agriculture. Springer, Dordrecht; 2009. pp. 153–188.

4. Guerra D, Crosatti C, Khoshro HH, Mastrangelo AM, Mica E, Mazzucotelli E. Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. Front Plant Sci. 2015; 6: 57. https://doi.org/10.3389/fpls.2015.00057 PMID: 25717333
5. Golldack D, Li C, Mohan H, Probst N. Tolerance to drought and salt stress in plants: unraveling the signaling networks. Front Plant Sci.; 2014; 15.
6. Chaves MM, Maroco JP, Pereira JS. Understanding plant responses to drought—from genes to the whole plant. Funct Plant Biol. 2003; 30: 239–264.
7. Haswell ES, Verslues PE. The ongoing search for the molecular basis of plant osmosensing. J Gen Physiol. 2015; 145: 389–394. https://doi.org/10.1085/jgp.201411295 PMID: 25870206
8. Todaka D, Shinozaki K, Yamaguchi-Shinozaki K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. Front Plant Sci.; 2015; 6: 84. https://doi.org/10.3389/fpls.2015.00084 PMID: 25741357
9. Foyer CH, Lam H-M, Nguyen HT, Siddique KHM, Varshney RK, Colmer TD, et al. Neglecting legumes has compromised human health and sustainable food production. Nat Plants. 2016; 2: 16112. https://doi.org/10.1038/nplants.2016.112 PMID: 28221372
10. Bertioli DJ, Seijo G, Freitas FO, Valls JFM, Leal-Bertioli SCM, Moretzsohn MC. An overview of peanut and its wild relatives. Plant Genet Resour. 2011; 9: 134–149.
11. Janila P, Variath MT, Pandey MK, Desmae H, Motagi BN, Okori P, et al. Genomic tools in groundnut breeding program: status and perspectives. Front Plant Sci.; 2016; 7.
12. Stalker HT. Utilizing wild species for peanut improvement. Crop Sci.; 2017; 57: 1102–1120.
13. Ferguson ME, Jarvis A, Stalker HT, Williams DE, Guarino L, Valls JFM, et al. Biogeography of wild Arachis (Leguminosae):distribution and environmental characterisation. Biodivers Conserv. 2005; 14: 1777–1798.
14. Brasileiro ACM, Araujo ACG, Leal-Bertioli SC, Guimaraes PM. Genomics and genetic transformation in Arachis. Int J Plant Biol & Res. 2014; 2: 1017.
15. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, et al. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat Genet.; 2016; 48: 438–446. https://doi.org/10.1038/ng.3517 PMID: 26901068
16. Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V, et al. Draft genome of the peanut A-genome progenitor (Arachis duranensis) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci. 2016; 113: 6785–6790. https://doi.org/10.1073/pnas.1600899113 PMID: 27247390
17. Kumar K, Kirti P. Differential gene expression in Arachis diogoi upon interaction with peanut late leaf spot pathogen, Phaeoisariopsis personata and characterization of a pathogen induced cyclophilin. Plant Mol Biol. Springer Netherlands; 2011; 75: 497–513. https://doi.org/10.1007/s11103-011-9747-3 PMID: 21298396
18. Proite K, Leal-Bertioli SCM, Bertioli DJ, Moretzsohn MC, Silva FR, Martins NF, et al. ESTs from a wild Arachis species for gene discovery and marker development. BMC Plant Biol. 2007; 7: 7. https://doi.org/10.1186/1471-2229-7-7 PMID: 17302987
19. Brasileiro AM, Morgante C, Araujo AG, Leal-Bertioli SM, Silva A, Martins AQ, et al. Transcriptome profiling of wild Arachis from water-limited environments uncovers drought tolerance candidate genes. Plant Mol Biol Report; 2015; 1–17.
20. Guimaraes PM, Brasileiro ACM, Proite K, Araujo ACG, Leal-Bertioli SCM, Pic-Taylor A, et al. A study of gene expression in the nematode resistant wild peanut relative, Arachis stenosperma, in response to challenge with Meloidogyne arenaria. Trop Plant Biol.; 2010; 3: 183–192.
21. Guimaraes PM, Guimaraes LA, Morgante C V, Silva OB Jr, Araujo ACG, Martins ACQ, et al. Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS One. 2015; 10: e0140937. https://doi.org/10.1371/journal.pone.0140937 PMID: 26488731
22. Guimaraes PM, Brasileiro ACM, Morgante C V, Martins ACQ, Pappas G, Silva OB, et al. Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genomics. 2012; 13: 387. https://doi.org/10.1186/1471-2164-13-387 PMID: 22888963
23. Guimaraes LA, Mota APZ, Araujo ACG, de Alencar Figueiredo LF, Pereira BM, de Passos Saraiva MA, et al. Genome-wide analysis of expansin superfamily in wild Arachis discloses a stress-responsive expansin-like B gene. Plant Mol Biol. 2017; 1–2: 79–96.
24. Leal-Bertioli SCM, Bertioli DJ, Guimaraes PM, Pereira TD, Galhardo I, Silva JP, et al. The effect of tetraploidization of wild Arachis on leaf morphology and other drought-related traits. Environ Exp Bot. 2012; 84: 17–24.
25. Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimenes MA, et al. A microsatellite-based, gene-rich linkage map for the AA genome of Arachis (Fabaceae). Theor Appl Genet. 2005; 111: 1060–1071. https://doi.org/10.1007/s00122-005-0028-x PMID: 16088397

26. Roberts WR, Roalson EH. Comparative transcriptome analyses of flower development in four species of Achimenes (Gesneriaceae). BMC Genomics; 2017; 18: 240. https://doi.org/10.1186/s12864-017-3623-8 PMID: 28320315

27. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, et al. Differential gene expression and mitotic cell analysis of the drought tolerant soybean (Glycine max L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet Mol Biol. 2008; 31: 512–521.

28. Stolf-Moreira R, Lemos EGM, Carareto-Alves L, Marcondes J, Pereira SS, Rolla AAP, et al. Transcriptional profiles of roots of different soybean genotypes subjected to drought stress. Plant Mol Biol Report; 2011; 29: 19–34.

29. Morgante C V, Brasileiro ACM, Roberts PA, Guimaraes LA, Araujo ACG, Fonseca LN, et al. A survey of genes involved in Arachis stenosperma resistance to Meloidogyne arenaria race 1. Funct Plant Biol. 2013; 40: 1298–1309.

30. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

31. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010; 26: 873–881. https://doi.org/10.1093/bioinformatics/btq057 PMID: 20147302

32. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013; 8: 1765–1786. https://doi.org/10.1038/nprot.2013.099 PMID: 23975260

33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

34. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16: 157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

35. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph. 2014; 20: 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248 PMID: 26356912

36. Prüfer K, Muetzel B, Do H-H, Weiss G, Khaitovich P, Rahm E, et al. FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinformatics. 2007; 8: 41. https://doi.org/10.1186/1471-2105-8-41 PMID: 17284313

37. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011; 6: e21800. https://doi.org/10.1371/journal.pone.0021800 PMID: 21789182

38. Zhao S, Fernald RD. Comprehensive algorithm for quantitative real-time polymerase chain reaction. J Comput Biol. 2005; 12: 1047–1064. https://doi.org/10.1089/cmb.2005.12.1047 PMID: 16241897

39. Morgante C V, Guimaraes PM, Martins A, Araujo ACG, Leal-Bertioli SCM, Bertioli DJ, et al. Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. BMC Res Notes. 2011; 4: 339. https://doi.org/10.1186/1756-0500-4-339 PMID: 21906295

40. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 2002; 30: 36.

41. Govind G, Vokkaliga ThammeGowda H, Jayaker Kalaiarasi P, Iyer D, Muthappa S, Nese S, et al. Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. Mol Genet Genomics. 2009; 281: 591–605. https://doi.org/10.1007/s00438-009-0432-z PMID: 19224247

42. Swain S, Myers ZA, Siriwardana CL, Holt BF. The multifaceted roles of NUCLEAR FACTOR-Y in Arabidopsis thaliana development and stress responses. BBA—Gene Regul Mech. 2017; 1860: 636–644.

43. Bemer M, van Dijk ADJ, Immink RGH, Angenent GC. Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. Trends Plant Sci. 2017; 22: 66–80. https://doi.org/10.1016/j.tplants.2016.10.007 PMID: 27814969

44. Clevenger JPC, Chu Y, Scheffler B, Ozias-Akins P. A developmental transcriptome map for allotetraploid Arachis hypogaea. Front Plant Sci. 2016; 7: 1446. https://doi.org/10.3389/fpls.2016.01446 PMID: 27746793

45. Tian X, Long Y, Wang J, Zhang J, Wang Y, Li W, et al. De novo transcriptome assembly of common wild rice (Oryza rufipogon Griff.) and discovery of drought-response genes in root tissue based on

transcriptomic data. PLoS One. 2015; 10: e0131455. https://doi.org/10.1371/journal.pone.0131455 PMID: 26134138

46. Bedada G, Westerbergh A, Müller T, Galkin E, Bdolach E, Moshelion M, et al. Transcriptome sequencing of two wild barley (Hordeum spontaneum L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. BMC Genomics. 2014; 15: 995. https://doi.org/10.1186/1471-2164-15-995 PMID: 25408241

47. Leal-Bertioli SCM, Santos SP, Dantas KM, Inglis PW, Nielen S, Araujo ACG, et al. Arachis batizocoi: a study of its relationship to cultivated peanut (A. hypogaea) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot. 2015; 115: 237–249. https://doi.org/10.1093/aob/mcu237 PMID: 25538110

48. Rabara RC, Tripathi P, Reese RN, Rushton DL, Alexander D, Timko MP, et al. Tobacco drought stress responses reveal new targets for Solanaceae crop improvement. BMC Genomics. 2015; 16: 484. https://doi.org/10.1186/s12864-015-1575-4 PMID: 26123791

49. Tripathi P, Rabara RC, Reese RN, Miller MA, Rohila JS, Subramanian S, et al. A toolbox of genes, proteins, metabolites and promoters for improving drought tolerance in soybean includes the metabolite coumestrol and stomatal development genes. BMC Genomics. 2016; 17: 102. https://doi.org/10.1186/s12864-016-2420-0 PMID: 26861168

50. Singh D, Singh CK, Taunk J, Tomar RSS, Chaturvedi AK, Gaikwad K, et al. Transcriptome analysis of lentil (Lens culinaris Medikus) in response to seedling drought stress. BMC Genomics. 2017; 18: 206. https://doi.org/10.1186/s12864-017-3596-7 PMID: 28241862

51. Azevedo Neto AD, Nogueira RJMC, Melo Filho PA, Santos RC. Physiological and biochemical responses of peanut genotypes to water deficit. J Plant Interact. 2010; 5: 1–10.

52. Knight H, Knight MR. Abiotic stress signalling pathways: specificity and cross-talk. Trends Plant Sci. 2001; 6: 262–267. PMID: 11378468

53. Huda KMK, Banu MSA, Tuteja R, Tuteja N. Global calcium transducer P-type Ca2+-ATPases open new avenues for agriculture by regulating stress signalling. J Exp Bot. 2013; 64: 3099–3109. https://doi.org/10.1093/jxb/ert182 PMID: 23918957

54. Qin X, Zeevaart JAD. The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean. Proc Natl Acad Sci. 1999; 96: 15354–15361. PMID: 10611388

55. Zhao P, Liu P, Yuan G, Jia J, Li X, Qi D, et al. New insights on drought stress response by global investigation of gene expression changes in Sheepgrass (Leymus chinensis). Front Plant Sci. 2016; 7: 954. https://doi.org/10.3389/fpls.2016.00954 PMID: 27446180

56. Wan X-R, Li L. Regulation of ABA level and water-stress tolerance of Arabidopsis by ectopic expression of a peanut 9-cis-epoxycarotenoid dioxygenase gene. Biochem Biophys Res Commun. 2006; 347: 1030–1038. https://doi.org/10.1016/j.bbrc.2006.07.026 PMID: 16870153

57. Qin X, Zeevaart JAD. Overexpression of a 9-cis-epoxycarotenoid dioxygenase gene in Nicotiana plumbaginifolia increases abscisic acid and phaseic acid levels and enhances drought tolerance. Plant Physiol. 2002; 128: 544–551. https://doi.org/10.1104/pp.010663 PMID: 11842158

58. Zandalinas SI, Rivero RM, Martínez V, Gómez-Cadenas A, Arbona V. Tolerance of citrus plants to the combination of high temperatures and drought is associated to the increase in transpiration modulated by a reduction in abscisic acid levels. BMC Plant Biol. 2016; 16: 105. https://doi.org/10.1186/s12870-016-0791-7 PMID: 27121193

59. Yoshida T, Mogami J, Yamaguchi-Shinozaki K. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Curr Opin Plant Biol. 2014; 21: 133–139. https://doi.org/10.1016/j.pbi.2014.07.009 PMID: 25104049

60. Lata C, Prasad M. Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot. 2011; 62: 4731–4748. https://doi.org/10.1093/jxb/err210 PMID: 21737415

61. Kazan K. Diverse roles of jasmonates and ethylene in abiotic stress tolerance. Trends Plant Sci. 2015; 20: 219–229. https://doi.org/10.1016/j.tplants.2015.02.001 PMID: 25731753

62. Du H, Wu N, Cui F, You L, Li X, Xiong L. A homolog of ETHYLENE OVERPRODUCER, OsETOL1, differentially modulates drought and submergence tolerance in rice. Plant J. 2014; 78: 834–849. https://doi.org/10.1111/tpj.12508 PMID: 24641694

63. Bakshi M, Oelmüller R. WRKY transcription factors: Jack of many trades in plants. Plant Signal Behav. 2014; 9: e27700. https://doi.org/10.4161/psb.27700 PMID: 24492469

64. Maurel C, Boursiac Y, Luu D-T, Santoni V, Shahzad Z, Verdoucq L. Aquaporins in plants. Physiol Rev. Am Physiological Soc; 2015; 95: 1321–1358.

65. Guimaraes LA, Pereira BM, Araujo ACG, Guimaraes PM, Brasileiro ACM. Ex vitro hairy root induction in detached peanut leaves for plant—nematode interaction studies. Plant Methods. 2017; 13: 25. https://doi.org/10.1186/s13007-017-0176-4 PMID: 28400855

66. Chen W, Yin X, Wang L, Tian J, Yang R, Liu D, et al. Involvement of rose aquaporin RhPIP1; 1 in ethylene-regulated petal expansion through interaction with RhPIP2; 1. Plant Mol Biol. 2013; 83: 219–233. https://doi.org/10.1007/s11103-013-0084-6 PMID: 23748738

67. Zargar SM, Nagar P, Deshmukh R, Nazir M, Wani AA, Masoodi KZ, et al. Aquaporins as potential drought tolerance inducing proteins: Towards instigating stress tolerance. J Proteomics. 2017; 169: 233–238. https://doi.org/10.1016/j.jprot.2017.04.010 PMID: 28412527

68. Selvaraj MG, Ishizaki T, Valencia M, Ogawa S, Dedicova B, Ogata T, et al. Overexpression of an Arabidopsis thaliana galactinol synthase gene improves drought tolerance in transgenic rice and increased grain yield in the field. Plant Biotechnol J. 2017; 15: 1465–1477 https://doi.org/10.1111/pbi.12731 PMID: 28378532

69. Himuro Y, Ishiyama K, Mori F, Gondo T, Takahashi F, Shinozaki K, et al. Arabidopsis galactinol synthase AtGolS2 improves drought tolerance in the monocot model Brachypodium distachyon. J Plant Physiol. 2014; 171: 1127–1131. https://doi.org/10.1016/j.jplph.2014.04.007 PMID: 24973584

70. Hanin M, Brini F, Ebel C, Toda Y, Takeda S, Masmoudi K. Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. Plant Signal Behav. 2011; 6: 1503–1509. https://doi.org/10.4161/psb.6.10.17088 PMID: 21897131

71. Banerjee A, Roychoudhury A. Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress. Plant Growth Regul. 2016; 79: 1–17.

72. Ariz I, Asensio AC, Zamarreño AM, García-Mina JM, Aparicio-Tejo PM, Moran JF. Changes in the C/N balance caused by increasing external ammonium concentrations are driven by carbon and energy availabilities during ammonium nutrition in pea plants: the key roles of asparagine synthetase and anaplerotic enzymes. Physiol Plant. 2013; 148: 522–537. https://doi.org/10.1111/j.1399-3054.2012.01712.x PMID: 23061733

73. Gaufichon L, Reisdorf-Cren M, Rothstein SJ, Chardon F, Suzuki A. Biological functions of asparagine synthetase in plants. Plant Sci. 2010; 179: 141–153.

74. Peoples MB, Pate JS, Atkins CA, Bergersen FJ. Nitrogen nutrition and xylem sap composition of peanut (Arachis hypogaea L. cv Virginia Bunch). Plant Physiol. 1986; 82: 946–951. PMID: 16665171

75. Botha A, Kunert K, Cullis C. Cysteine proteases and wheat (Triticum aestivum L) under drought: A still greatly unexplored association. Plant Cell Environ. 2017; 40: 1679–1690 https://doi.org/10.1111/pce.12998 PMID: 28664627

76. Dramé KN, Clavel D, Repellin A, Passaquet C, Zuily-Fodil Y. Water deficit induces variation in expression of stress-responsive genes in two peanut (Arachis hypogaea L.) cultivars with different tolerance to drought. Plant Physiol Biochem. 2007; 45: 236–243. https://doi.org/10.1016/j.plaphy.2007.02.002 PMID: 17433701

77. Munné-Bosch S, Alegre L. Die and let live: leaf senescence contributes to plant survival under drought stress. Funct Plant Biol. 2004; 31: 203–216.

**REFERENCES**


Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., et al. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nat. Biotechnol.* 26, 909–915. doi:10.1038/nbt.1482.

Abedini, R., Ghanegolmohammadi, F., and Pishkamrad, R. (2017). Plant dehydrins : shedding light on structure and expression patterns of dehydrin gene family in barley. *J. Plant Res.* 130, 747–763. doi:10.1007/s10265-017-0941-5.

Adie, B. A. T., Perez-Perez, J., Perez-Perez, M. M., Godoy, M., Sanchez-Serrano, J.-J., Schmelz, E. A., et al. (2007). ABA Is an Essential Signal for Plant Resistance to Pathogens Affecting JA Biosynthesis and the Activation of Defenses in Arabidopsis. *Plant Cell Online* 19, 1665–1681. doi:10.1105/tpc.106.048041.

Agrios (2005). *Introduction to plant pathology*.

Albert, V. A., Barbazuk, W. B., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., et al. (2013). The Amborella Genome and the Evolution of Flowering Plants. *Science (80-. ).* 342. doi:10.1126/science.1241089.

Álvarez, C., Ángeles Bermúdez, M., Romero, L. C., Gotor, C., and García, I. (2012). Cysteine homeostasis plays an essential role in plant immunity. *New Phytol.* 193, 165–177. doi:10.1111/j.1469-8137.2011.03889.x.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.

Atkinson, N. J., Jain, R., and Urwin, P. E. (2015). Combined Stresses in Plants. 181–201. doi:10.1007/978-3-319-07899-1.

Atkinson, N. J., Lilley, C. J., and Urwin, P. E. (2013). Identification of Genes Involved in the Response of Arabidopsis to Simultaneous Biotic and Abiotic Stresses. *Plant Physiol.* 162, 2028–2041. doi:10.1104/pp.113.222372.

Atkinson, N. J., and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63, 3523–3543.

Bai, Y., Sunarti, S., Kissoudis, C., Visser, R. G. F., and Linden, C. G. Van Der (2018). The Role of Tomato WRKY Genes in Plant Responses to Combined Abiotic and Biotic Stresses. 9, 1–7. doi:10.3389/fpls.2018.00801.

Bao, F., Du, D., An, Y., Yang, W., Wang, J., Cheng, T., et al. (2017). Overexpression of Prunus

mume dehydrin genes in tobacco enhances tolerance to cold and drought. *Front. Plant Sci.* 8, 1–12. doi:10.3389/fpls.2017.00151.

Barbary, A., Djian-Caporalino, C., Marteu, N., Fazari, A., Caromel, B., Castagnone-Sereno, P., et al. (2016). Plant Genetic Background Increasing the Efficiency and Durability of Major Resistance Genes to Root-knot Nematodes Can Be Resolved into a Few Resistance QTLs. *Front. Plant Sci.* 7, 1–9. doi:10.3389/fpls.2016.00632.

Bartels, D., and Sunkar, R. (2005). Drought and Salt Tolerance in Plants. *CRC. Crit. Rev. Plant Sci.* 24, 23–58. doi:10.1080/07352680590910410.

Baxter, A., Mittler, R., and Suzuki, N. (2014). ROS as key players in plant stress signalling. *J. Exp. Bot.* 65, 1229–1240.

Beneventi, M., da Silva, O., de Sa, M. E., Firmino, A. A., de Amorim, R. M., Albuquerque, E. V., et al. (2013). Transcription profile of soybean-root-knot nematode interaction reveals a key role of phythormones in the resistance reaction. *BMC Genomics* 14, 322. Available at: http://www.biomedcentral.com/1471-2164/14/322.

Berens, M. L., Wolinska, K. W., Spaepen, S., Ziegler, J., Nobori, T., and Nair, A. (2018). Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. doi:10.1073/pnas.1817233116.

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of Arachis duranensis and Arachis ipaënsis, the diploid ancestors of cultivated peanut. *Nat Genet* 48, 438–446. doi:10.1038/ng.3517.

Bertioli, D. J., Leal-Bertioli, S. C. M., Lion, M. B., Santos, V. L., Pappas, G., Cannon, S. B., et al. (2003). A large scale analysis of resistance gene homologues in Arachis. *Mol. Genet. Genomics* 270, 34–45. doi:10.1007/s00438-003-0893-4.

Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: Finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi:10.1016/j.pbi.2010.04.010.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Bostock, R. M., Pye, M. F., and Roubtsova, T. V (2014). Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. *Annu. Rev. Phytopathol.* 52, 517–549.

Bouché, N., Yellin, A., Snedden, W. A., and Fromm, H. (2005). Plant-Specific Calmodulin-

Binding Proteins. *Annu. Rev. Plant Biol.* 56, 435–466. doi:10.1146/annurev.arplant.56.032604.144224.

Bowler, C., and Fluhr, R. (2000). The role of calcium and activated oxygens as signals for controlling cross-tolerance. *Trends Plant Sci.* 5, 241–246. doi:10.1016/S1360-1385(00)01628-9.

Brasileiro, A. M. C. M., Morgante, C. V., Araujo, A. G. C. G., Leal-Bertioli, S. C. M., Silva, A. K., Martins, A. C. Q., et al. (2015). Transcriptome profiling of wild Arachis from water-limited environments uncovers drought tolerance candidate genes. *Plant Mol. Biol. Report.*, 1–17. doi:10.1007/s11105-015-0882-x.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi:10.1038/nbt.3519.

Breitling, R., Armengaud, P., and Amtmann, A. (2005). Expression Responses Between Different Experimental Backgrounds. 7, 1–7. doi:10.1186/1471-2105-6-181.

Brossa, R., López-Carbonell, M., Jubany-Marí, T., and Alegre, L. (2011). Interplay Between Abscisic Acid and Jasmonic Acid and its Role in Water-oxidative Stress in Wild-type, ABA-deficient, JA-deficient, and Ascorbate-deficient Arabidopsis Plants. *J. Plant Growth Regul.* 30, 322–333. doi:10.1007/s00344-011-9194-z.

Cagnola, J. I., Dumont de Chassart, G. J., Ibarra, S. E., Chimenti, C., Ricardi, M. M., Delzer, B., et al. (2018). Reduced expression of selected FASCICLIN-LIKE ARABINOGALACTAN PROTEIN genes associates with the abortion of kernels in field crops of Zea mays (maize) and of Arabidopsis seeds. *Plant Cell Environ.* 41, 661–674. doi:10.1111/pce.13136.

Cao, F. Y., Yoshida, K., and Desveaux, D. (2011). The roles of ABA in plant – pathogen interactions. *J Plant Res* 124, 489–499. doi:10.1007/s10265-011-0409-y.

Carmo, L. S. T., Martins, A. C. Q., Martins, C. C. C., Passos, M. A. S., Silva, L. P., Araujo, A. C. G., et al. (2019). Comparative proteomics and gene expression analysis in Arachis duranensis reveal stress response proteins associated to drought tolerance. *J. Proteomics* 192, 299–310. doi:10.1016/j.jprot.2018.09.011.

Chan, Z. (2012). Expression profiling of ABA pathway transcripts indicates crosstalk between abiotic and biotic stress responses in Arabidopsis. *Genomics* 100, 110–115. doi:10.1016/j.ygeno.2012.06.004.

Chaves, M. M., Maroco, J. P., and Pereira, J. S. (2003). Understanding plant responses to

drought—from genes to the whole plant. *Funct. Plant Biol.* 30, 239–264.

Choudhury, F. K., Rivero, R. M., Blumwald, E., and Mittler, R. (2017). Reactive oxygen species , abiotic stress and stress combination. 856–867. doi:10.1111/tpj.13299.

Close, T. J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiol. Plant.* 97, 795–803. doi:10.1034/j.1399-3054.1996.970422.x.

Cosgrove, D. J. (2000). Loosening of plant cell walls by expansins. *Nature* 407, 321–326. Available at: http://dx.doi.org/10.1038/35030000.

Cosgrove, D. J. (2015). Plant expansins: diversity and interactions with plant cell walls. *Curr. Opin. Plant Biol.* 25, 162–172. doi:10.1016/j.pbi.2015.05.014.

Cottret, L., Rancurel, C., Briand, M., and Carrere, S. (2018). Family-Companion: analyse, visualise, browse, query and share your homology clusters. *bioRxiv*, 266742. doi:10.1101/266742.

Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the Plant Immune System from Dissection to Deployment. *Science (80-. ).* 341, 746–751. doi:10.1126/science.1236011.

Das, M., Haberer, G., Panda, A., Laha, S. Das, Ghosh, T. C., and Schäffner, A. R. (2016). Expression pattern similarities support the prediction of orthologs retaining common functions after gene duplication events. *Plant Physiol.* 171, 2343–2357. doi:10.1104/pp.15.01207.

de Ollas, C., Arbona, V., and Gómez-Cadenas, A. (2015). Jasmonic acid interacts with abscisic acid to regulate plant responses to water stress conditions. *Plant Signal. Behav.* 10, 1–4. doi:10.1080/15592324.2015.1078953.

De Ollas, C., Arbona, V., Gómez-Cadenas, A., and Dodd, I. C. (2018). Attenuated accumulation of jasmonates modifies stomatal responses to water deficit. *J. Exp. Bot.* 69, 2103–2116. doi:10.1093/jxb/ery045.

De Waele, D., and Elsen, A. (2007). Challenges in Tropical Plant Nematology. doi:10.1146/annurev.phyto.45.062806.094438.

Decraemer, W., and Hunt, D. (2006). "Structure and Classification," in *Plant Nematology*, 3–32.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635.

Dubois, M., Skirycz, A., Claeys, H., Maleux, K., Dhondt, S., De Bodt, S., et al. (2013). ETHYLENE RESPONSE FACTOR6 Acts as a Central Regulator of Leaf Growth under Water-Limiting Conditions in Arabidopsis. *Plant Physiol.* 162, 319–332. doi:10.1104/pp.113.216341.

Dubois, M., Van den Broeck, L., and Inzé, D. (2018). The Pivotal Role of Ethylene in Plant Growth. *Trends Plant Sci.* 23, 311–323. doi:10.1016/j.tplants.2018.01.003.

Dutt, S., Pandey, D., and Kumar, A. (2011). Jasmonate signal induced expression of cystatin genes for providing resistance against karnal bunt in wheat. *Plant Signal. Behav.* 6, 821–830. doi:10.4161/psb.6.6.14743.

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 1–14. doi:10.1186/s13059-015-0721-2.

Fan, J., Hill, L., Crooks, C., Doerner, P., and Lamb, C. (2009). Abscisic Acid Has a Key Role in Modulating Diverse Plant-Pathogen Interactions. *Plant Physiol.* 150, 1750–1761. doi:10.1104/pp.109.137943.

Finkelstein, R. (2013). Abscisic Acid Synthesis and Response. *Arab. B.* 11, e0166. doi:10.1199/tab.0166.

Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N., and Travers, S. E. (2006). Climate Change Effects on Plant Disease: Genomes to Ecosystems. *Annu. Rev. Phytopathol.* 44, 489–509. doi:10.1146/annurev.phyto.44.070505.143420.

Glazebrook, J. (2005). Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi:10.1146/annurev.phyto.43.040204.135923.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–52. doi:10.1038/nbt.1883.

Graether, S. P., and Boddington, K. F. (2014). Disorder and function: a review of the dehydrin protein family. *Front. Plant Sci.* 5, 1–12. doi:10.3389/fpls.2014.00576.

Guimaraes, L. A., Mota, A. P. Z., Araujo, A. C. G., de Alencar Figueiredo, L. F., Pereira, B. M., de Passos Saraiva, M. A., et al. (2017). Genome-wide analysis of expansin superfamily in wild Arachis discloses a stress-responsive expansin-like B gene. *Plant Mol. Biol.* 94, 1–18.

doi:10.1007/s11103-017-0594-8.

Guimaraes, P. M., Guimaraes, L. A., Morgante, C. V., Silva, O. B., Araujo, A. C. G., Martins, A. C. Q., et al. (2015). Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. *PLoS One* 10, 1–22. doi:10.1371/journal.pone.0140937.

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signal. Behav.* 6, 1503–1509. doi:10.4161/psb.6.10.17088.

Hara, M., Fujinaga, M., and Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *J. Exp. Bot.* 56, 2695–2703. doi:10.1093/jxb/eri262.

Hill, W., Jin, X.-L., and Zhang, X.-H. (2016). Expression of an arctic chickweed dehydrin, CarDHN, enhances tolerance to abiotic stress in tobacco plants. *Plant Growth Regul.* 80, 323–334. doi:10.1007/s10725-016-0169-8.

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. 444, 323–329. doi:10.1038/nature05286.

Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., et al. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* 14, 946–961. doi:10.1111/mpp.12057.

Kazan, K., and Manners, J. M. (2013). MYC2: The master in action. *Mol. Plant* 6, 686–703. doi:10.1093/mp/sss128.

Kissoudis, C., Sunarti, S., Van De Wiel, C., Visser, R. G. F., Van Der Linden, C. G., and Bai, Y. (2016). Responses to combined abiotic and biotic stress in tomato are governed by stress intensity and resistance mechanism. *J. Exp. Bot.* 67, 5119–5132. doi:10.1093/jxb/erw285.

Kollist, H., Zandalinas, S. I., Sengupta, S., Nuhkat, M., Kangasjärvi, J., and Mittler, R. (2019). Rapid Responses to Abiotic Stress: Priming the Landscape for the Signal Transduction Network. *Trends Plant Sci.* 24, 25–37. doi:10.1016/j.tplants.2018.10.003.

Koonin, E. V (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–38. doi:10.1146/annurev.genet.39.073003.114725.

Kovacs, D., Kalmar, E., Torok, Z., and Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol.* 147, 381–390. doi:10.1104/pp.108.118208.

Ku, Y., Sintaha, M., Cheung, M., and Lam, H. (2018). *Plant Hormone Signaling Crosstalks*

*between Biotic and Abiotic Stress Responses*. doi:10.3390/ijms19103206.

Laetsch, D. R., and Blaxter, M. (2017). KinFin : Software for Taxon-Aware Analysis of Clustered Protein Sequences. 7, 3349–3357. doi:10.1534/g3.117.300233.

Lata, C., and Prasad, M. (2011). Role of DREBs in regulation of abiotic stress responses in plants. *J. Exp. Bot.* 62, 4731–4748.

Leal-Bertioli, S. C., José, A., Alves-Freitas, D. M., Moretzsohn, M. C., Guimarães, P. M., Nielen, S., et al. (2009). Identification of candidate genome regions controlling disease resistance in Arachis. *BMC Plant Biol.* 9, 112. doi:10.1186/1471-2229-9-112.

Leal-Bertioli, S. C. M., Bertioli, D. J., Guimaraes, P. M., Pereira, T. D., Galhardo, I., Silva, J. P., et al. (2012). The effect of tetraploidization of wild Arachis on leaf morphology and other drought-related traits. *Environ. Exp. Bot.* 84, 17–24. doi:10.1016/j.envexpbot.2012.04.005.

Leba, L. J., Cheval, C., Ortiz-Martín, I., Ranty, B., Beuzón, C. R., Galaud, J. P., et al. (2012). CML9, an Arabidopsis calmodulin-like protein, contributes to plant innate immunity through a flagellin-dependent signalling pathway. *Plant J.* 71, 976–989. doi:10.1111/j.1365-313X.2012.05045.x.

Lee, I. H., Shim, D., Jeong, J. C., Sung, Y. W., Nam, K. J., Yang, J. W., et al. (2018). Transcriptome analysis of root-knot nematode (Meloidogyne incognita)-resistant and susceptible sweetpotato cultivars. *Planta* 249, 431–444. doi:10.1007/s00425-018-3001-z.

Li, M., Wang, L., Chen, M., Chai, R., Wang, X., Chang, J., et al. (2015). Expression of TaWRKY44, a wheat WRKY gene, in transgenic tobacco confers multiple abiotic stress tolerances. *Front. Plant Sci.* 6, 1–14. doi:10.3389/fpls.2015.00615.

Love, M., Anders, S., and Huber, W. (2013). Differential analysis of RNA-Seq data at the gene level using the DESeq package. 1–32.

Ma, H., Liang, D., Shuai, P., Xia, X., and Yin, W. (2010). The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in Arabidopsis thaliana. 61, 4011–4019. doi:10.1093/jxb/erq217.

Magnan, F., Ranty, B., Charpenteau, M., Sotta, B., Galaud, J. P., and Aldon, D. (2008). Mutations in AtCML9, a calmodulin-like protein from Arabidopsis thaliana, alter plant responses to abiotic stress and abscisic acid. *Plant J.* 56, 575–589. doi:10.1111/j.1365-313X.2008.03622.x.

Marowa, P., Ding, A., and Kong, Y. (2016). Expansins: roles in plant growth and potential

applications in crop improvement. *Plant Cell Rep.* 35, 1–17. doi:10.1007/s00299-016-1948-4.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi:10.1046/j.1365-313X.1999.00606.x.

Mickelbart, M. V, Hasegawa, P. M., and Bailey-serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. 16. doi:10.1038/nrg3901.

Milisavljević, M. D., Radović, S. R., Konstantinović, M. M., Timotijević, G. S., and Maksimović, V. R. (2009). Ubiquitous aspartic proteinase as an actor in the stress response in buckwheat. *J. Plant Physiol.* 167, 61–68. doi:10.1016/j.jplph.2009.06.017.

Mistry, J., Bateman, A., and Finn, R. D. (2007). Predicting active site residue annotations in the Pfam database. 14, 1–14. doi:10.1186/1471-2105-8-298.

Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. doi:10.1016/j.tplants.2005.11.002.

Mittler, R., and Blumwald, E. (2010). Genetic Engineering for Modern Agriculture: Challenges and Perspectives. *Annu. Rev. Plant Biol.* 61, 443–462. doi:doi:10.1146/annurev-arplant-042809-112116.

Morgante, C. V., Brasileiro, A. C. M., Roberts, P. A., Guimaraes, L. A., Araujo, A. C. G., Fonseca, L. N., et al. (2013). A survey of genes involved in arachis stenosperma resistance to meloidogyne arenaria race 1. *Funct. Plant Biol.* 40, 1298–1309. doi:10.1071/FP13096.

Morgante, C. V, Guimaraes, P. M., Martins, A., Araujo, A. C. G., Leal-Bertioli, S. C. M., Bertioli, D. J., et al. (2011). Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes* 4, 339. doi:10.1186/1756-0500-4-339.

Mota, A. P. Z., Vidigal, B., Danchin, E. G. J., Togawa, R. C., Leal-Bertioli, S. C. M., Bertioli, D. J., et al. (2018). Comparative root transcriptome of wild Arachis reveals NBS-LRR genes related to nematode resistance. *BMC Plant Biol.* 18, 159. doi:10.1186/s12870-018-1373-7.

Ndimba, B. K., Chivasa, S., Simon, W. J., and Slabas, A. R. (2005). Identification of Arabidopsis salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 5, 4185–4196.

doi:10.1002/pmic.200401282.

Nieri, D., Di Donato, A., and Ercolano, M. R. (2017). Analysis of tomato meiotic recombination profile reveals preferential chromosome positions for NB-LRR genes. *Euphytica* 213, 206. doi:10.1007/s10681-017-1982-5.

Nobori, T., and Tsuda, K. (2019). The plant immune system in heterogeneous environments. *Curr. Opin. Plant Biol.* 50, 58–66. doi:10.1016/j.pbi.2019.02.003.

O'Brien, J. A., and Benková, E. (2013). Cytokinin cross-talking during biotic and abiotic stress responses. *Front. Plant Sci.* 4. doi:10.3389/fpls.2013.00451.

Olivas, N. H. D., Coolen, S., Huang, P., Severing, E., Verk, M. C. Van, Wittenberg, A. H. J., et al. (2016). Effect of prior drought and pathogen stress on Arabidopsis transcriptome changes to caterpillar herbivory.

Ollas, C., and Dodd, I. C. (2016). Physiological impacts of ABA – JA interactions under water-limitation. *Plant Mol. Biol.* 91, 641–650. doi:10.1007/s11103-016-0503-6.

Pandey, P., Ramegowda, V., and Senthil-kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations : physiological and molecular mechanisms. 6, 1–14. doi:10.3389/fpls.2015.00723.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi:10.1093/bioinformatics/btm071.

Peters, G. P., Marland, G., Quéré, C., Boden, T., Canadell, J. G., and Raupach, M. R. (2012). Rapid growth in CO 2 emissions after the 2008 – 2009 global financial crisis. *Nat. Clim. Chang.* 2, 2009–2011. doi:10.1038/nclimate1332.

Petitot, A. S., Kyndt, T., Haidar, R., Dereeper, A., Collin, M., De Almeida Engler, J., et al. (2017). Transcriptomic and histological responses of African rice (Oryza glaberrima) to Meloidogyne graminicola provide new insights into root-knot nematode resistance in monocots. *Ann. Bot.* 119, 885–899. doi:10.1093/aob/mcw256.

Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30, 36. doi:10.1093/nar/30.9.e36.

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2012). Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol.* 28, 489–521.

doi:10.1146/annurev-cellbio-092910-154055.

Pont, C., Murat, F., Confolent, C., Balzergue, S., and Salse, J. (2011). RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (Triticum aestivum L.). *Genome Biol.* 12, R119. doi:10.1186/gb-2011-12-12-r119.

Postnikova, O. A., Hult, M., Shao, J., Skantar, A., and Nemchinov, L. G. (2015). Transcriptome analysis of resistant and susceptible alfalfa cultivars infected with root-knot nematode Meloidogyne incognita. *PLoS One* 10, 1–17. doi:10.1371/journal.pone.0118269.

Pre, M., Atallah, M., Champion, A., De Vos, M., Pieterse, C. M. J., and Memelink, J. (2008). The AP2/ERF Domain Transcription Factor ORA59 Integrates Jasmonic Acid and Ethylene Signals in Plant Defense. *Plant Physiol.* 147, 1347–1357. doi:10.1104/pp.108.117523.

Proite, K., Carneiro, R., Falcao, R., Gomes, A., Leal-Bertioli, S., Guimaraes, P., et al. (2008). Post-infection development and histopathology of Meloidogyne arenaria race 1 on Arachis spp. *Plant Pathol.* 57, 974–980. doi:10.1111/j.1365-3059.2008.01861.x.

Prüfer, K., Muetzel, B., Do, H.-H., Weiss, G., Khaitovich, P., Rahm, E., et al. (2007). FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* 8, 41. doi:10.1186/1471-2105-8-41.

Ramegowda, V., and Senthil-kumar, M. (2015). The interactive effects of simultaneous biotic and abiotic stresses on plants : Mechanistic understanding from drought and pathogen combination. 176, 47–54.

Ramegowda, V., and Senthil-Kumar, M. (2015). The interactive effects of simultaneous biotic and abiotic stresses on plants: Mechanistic understanding from drought and pathogen combination. *J. Plant Physiol.* 176, 47–54. doi:10.1016/j.jplph.2014.11.008.

Rasmussen, S., Friis, P., Bressendorff, S., Bones, A. M., Nielsen, H. B., Barah, P., et al. (2013). Transcriptome Responses to Combinations of Stresses in Arabidopsis. *Plant Physiol.* 161, 1783–1794. doi:10.1104/pp.112.210773.

Remm, M., Storm, C. E. V, and Sonnhammer, E. L. L. (2001). Automatic Clustering of Orthologs and In-paralogs from Pairwise Species Comparisons. 1041–1052. doi:10.1006/jmbi.2001.5197.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., Mittler, R., et al. (2004). When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiol.* 134, 1683–1696. doi:10.1104/pp.103.033431.1.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Rosales, R., Romero, I., Escribano, M. I., Merodio, C., and Sanchez-Ballesta, M. T. (2014). The crucial role of Φ- And K-segments in the in vitro functionality of Vitis vinifera dehydrin DHN1a. *Phytochemistry* 108, 17–25. doi:10.1016/j.phytochem.2014.10.006.

Roux, F., Voisin, D., Badet, T., Balagué, C., Barlet, X., Huard-chauveau, C., et al. (2014). Resistance to phytopathogens e tutti quanti : placing plant quantitative disease resistance on the map. *Mol. Plant Pathol.* 15, 427–432. doi:10.1111/mpp.12138.

Sasidharan, R., Voesenek, L. A. C. J., and Pierik, R. (2011). Cell wall modifying proteins mediate plant acclimatization to biotic and abiotic stresses. *CRC. Crit. Rev. Plant Sci.* 30, 548–562.

Sendon, P. M., Seo, H. S., and Song, J. T. (2011). Salicylic Acid Signaling: Biosynthesis, Metabolism, and Crosstalk with Jasmonic Acid. *J. Korean Soc. Appl. Biol. Chem* 54, 501–506. Available at: http://www.ksabc.or.kr/admin/contribute/journal/kpaper/54-402054501-506.pdf.

Seo, P. J., Xiang, F., Qiao, M., Park, J., Lee, Y. N., Kim, S., et al. (2009). The MYB96 transcription factor mediates abscisic acid signaling during drought stress response. *Plant Physiol.* 151, 275–289. doi:10.1104/pp.109.144220.

Shaik, R., and Ramakrishna, W. (2014). Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant Physiol.* 164, 481–495. doi:10.1104/pp.113.225862.

Shen, Y., Tang, M. J., Hu, Y. L., and Lin, Z. P. (2004). Isolation and characterization of a dehydrin-like gene from drought-tolerant Boea crassifolia. *Plant Sci.* 166, 1167–1175. doi:10.1016/j.plantsci.2003.12.025.

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* 58, 221–227. doi:10.1093/jxb/erl164.

Siddique, S., Radakovic, Z. S., De La Torre, C. M., Chronis, D., Novák, O., Ramireddy, E., et al. (2015). A parasitic nematode releases cytokinin that controls cell division and orchestrates feeding site formation in host plants. *Proc. Natl. Acad. Sci.* 112, 12669–12674. doi:10.1073/pnas.1503657112.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015).

BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351.

Sinapidou, E., Williams, K., Nott, L., Bahkt, S., T??r, M., Crute, I., et al. (2004). Two TIR:NB:LRR genes are required to specify resistance to Peronospora parasitica isolate Cala2 in Arabidopsis. *Plant J.* 38, 898–909. doi:10.1111/j.1365-313X.2004.02099.x.

Snel, B., Bork, P., and Huynen, M. A. (2002). The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci.* 99, 5890–5895. doi:10.1073/pnas.092632599.

Song, H., Wang, P., Li, C., Han, S., Zhao, C., Xia, H., et al. (2017). Comparative analysis of NBS-LRR genes and their response to Aspergillus flavus in Arachis. *PLoS One* 12, e0171181. doi:10.1371/journal.pone.0171181.

Stöver, B. C., and Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11, 7. doi:10.1186/1471-2105-11-7.

Street, N. R., Sjödin, A., Bylesjö, M., Gustafsson, P., Trygg, J., and Jansson, S. (2008). A cross-species transcriptomics approach to identify genes involved in leaf development. *BMC Genomics* 9. doi:10.1186/1471-2164-9-589.

Suárez-Vidal, E., Sampedro, L., Voltas, J., Serrano, L., Notivol, E., and Zas, R. (2019). Drought stress modifies early effective resistance and induced chemical defences of Aleppo pine against a chewing insect herbivore. *Environ. Exp. Bot.* 162, 550–559. doi:10.1016/j.envexpbot.2019.04.002.

Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., and Mittler, R. (2014). Abiotic and biotic stress combinations. *New Phytol.* 203, 32–43.

Tang, D., Wang, G., and Zhou, J.-M. (2017). Receptor Kinases in Plant-Pathogen Interactions: More Than Pattern Recognition. *Plant Cell* 29, 618–637. doi:10.1105/tpc.16.00891.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi:10.1111/j.1365-313X.2004.02016.x.

Tian, T., You, Q., Zhang, L., Yi, X., Yan, H., Xu, W., et al. (2016). SorghumFDB: sorghum functional genomics database with multidimensional network analysis. *Database (Oxford).* 2016, 1–16. doi:10.1093/database/baw099.

Ton, J., Flors, V., and Mauch-mani, B. (2009). The multifaceted role of ABA in disease

resistance. doi:10.1016/j.tplants.2009.03.006.

Tripathi, D. K., Singh, S., Singh, V. P., and Prasad, S. M. (2016). Silicon Nanoparticles More Efficiently Alleviate Arsenate Toxicity than Silicon in Maize Cultiver and Hybrid Differing in Arsenate Tolerance. 4. doi:10.3389/fenvs.2016.00046.

Turco, E., Close, T. J., Fenton, R. D., and Ragazzi, A. (2004). Synthesis of dehydrin-like proteins in Quercus ilex L. and Quercus cerris L. seedlings subjected to water stress and infection with Phytophthora cinnamomi. *Physiol. Mol. Plant Pathol.* 65, 137–144. doi:10.1016/j.pmpp.2004.11.010.

Valéria, E., Albuquerque, S., Maria, R., and Gomes, D. (2010). Resistance to Meloidogyne incognita expresses a hypersensitive-like response in Coffea arabica. 365–373. doi:10.1007/s10658-010-9603-3.

Van Verk, M. C., Hickman, R., Pieterse, M. J., and Van Wees, S. C. M. Van (2013). RNA-Seq : revelation of the messengers. *Trends Plant Sci.* 18, 175–179. doi:10.1016/j.tplants.2013.02.001.

Vinson, C. C., Mota, A. P. Z., Oliveira, T. N., Guimaraes, L. A., Leal-Bertioli, S. C. M., Williams, T. C. R., et al. (2018). Early responses to dehydration in contrasting wild Arachis species. *PLoS One* 13, 1–28. doi:10.1371/journal.pone.0198191.

Visser, R. G. F., Sunarti, S., van der Linden, C. G., Bai, Y., and Kissoudis, C. (2018). The Role of Tomato WRKY Genes in Plant Responses to Combined Abiotic and Biotic Stresses. *Front. Plant Sci.* 9, 1–7. doi:10.3389/fpls.2018.00801.

Wang, H., Penmetsa, R. V., Yuan, M., Gong, L., Zhao, Y., Guo, B., et al. (2012). Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (Arachis hypogaea L.). *BMC Plant Biol.* 12, 10. doi:10.1186/1471-2229-12-10.

Wasternack, C. (2014). Action of jasmonates in plant stress responses and development - Applied aspects. *Biotechnol. Adv.* 32, 31–39. doi:10.1016/j.biotechadv.2013.09.009.

Weretilnyk, E. A. (2002). Maintaining Methylation Activities during Salt Stress. The Involvement of Adenosine Kinase. *Plant Physiol.* 125, 856–865. doi:10.1104/pp.125.2.856.

Woldesemayat, A. A., Modise, D. M., Gemeildien, J., Ndimba, B. K., and Christoffels, A. (2018). Cross-species multiple environmental stress responses: An integrated approach to identify candidate genes for multiple stress tolerance in sorghum (Sorghum bicolor (L.)

Moench) and related model species. *PLoS One* 13, 1–30. doi:10.1371/journal.pone.0192678.

Xie, Z., Nolan, T. M., Jiang, H., and Yin, Y. (2019). AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Front. Plant Sci.* 10, 1–17. doi:10.3389/fpls.2019.00228.

Xu, K., Chen, S., Li, T., Ma, X., Liang, X., Ding, X., et al. (2015). OsGRAS23, a rice GRAS transcription factor gene, is involved in drought stress response through regulating expression of stress-responsive genes. *BMC Plant Biol.* 15, 141. doi:10.1186/s12870-015-0532-3.

Xu, Y., Liu, F., Zhu, S., and Li, X. (2018). The Maize NBS-LRR Gene ZmNBS25 Enhances Disease Resistance in Rice and Arabidopsis. *Front. Plant Sci.* 9, 1–13. doi:10.3389/fpls.2018.01033.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57, 781–803. Available at: http://dx.doi.org/10.1146/annurev.arplant.57.032905.105444.

Yang, Y., He, M., Zhu, Z., Li, S., Xu, Y., Zhang, C., et al. (2012). Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic stress. *BMC Plant Biol.* 12, 140. doi:10.1186/1471-2229-12-140.

Yoshida, Y., Koutsovoulos, G., Laetsch, D. R., Stevens, L., Kumar, S., Horikawa, D. D., et al. (2017). *Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus*. doi:10.1371/journal.pbio.2002266.

Zandalinas, S. I., Mittler, R., Balfagón, D., Arbona, V., and Gómez-cadenas, A. (2018). Plant adaptations to the combination of drought and high. 2–12. doi:10.1111/ppl.12540.

Zang, L., Zheng, T., Chu, Y., Ding, C., Zhang, W., Huang, Q., et al. (2015). Genome-Wide Analysis of the Fasciclin-Like Arabinogalactan Protein Gene Family Reveals Differential Expression Patterns, Localization, and Salt Stress Response in Populus. *Front. Plant Sci.* 6, 1–16. doi:10.3389/fpls.2015.01140.

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., et al. (2013). Three keys to the radiation of angiosperms into freezing environments. *Nature* 506, 89–92. doi:10.1038/nature12872.

Zhang, C., Chen, H., Cai, T., Deng, Y., Zhuang, R., Zhang, N., et al. (2016). Overexpression of a

novel peanut NBS-LRR gene AhRRS5 enhances disease resistance to Ralstonia solanacearum in tobacco. *Plant Biotechnol. J.*, 1–17. doi:10.1111/pbi.12589.

Zhao, K., Zhang, X., Cheng, Z., Yao, W., Li, R., and Jiang, T. (2019). Plant Physiology and Biochemistry Comprehensive analysis of the three-amino-acid-loop-extension gene family and its tissue-di ff erential expression in response to salt stress in poplar. *Plant Physiol. Biochem.* 136, 1–12. doi:10.1016/j.plaphy.2019.01.003.

Zhao, S., Fernald, R. D., and Zhao, Sheng; Fernald, D. R. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.* 12, 1047--1064. doi:10.1089/cmb.2005.12.1047.Comprehensive.

Zhu, X., Xiao, K., Cui, H., and Hu, J. (2017). Overexpression of the Prunus sogdiana NBS-LRR subgroup gene PsoRPM2 promotes resistance to the root-knot nematode Meloidogyne incognita in tobacco. *Front. Microbiol.* 8, 1–13. doi:10.3389/fmicb.2017.02113.

**CURRICULUM VITAE resumido**

**MOTA, APZ**

## 1. DADOS PESSOAIS

**Nome:** Ana Paula Zotta Mota

**Local e data de nascimento:** Campo Grande, Mato Grosso do Sul, 30/11/1988

**Email:** anazottamota@gmail.com

## 2. FORMAÇÃO

2015 – 2019 PhD. Programa de Biologia Celular e molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

Título: Study of biological systems involved in the response to biotic and abiotic stress in *Arachis* spp

Supervisores: Maria Fátima Grossi de Sá
Doutorado sanduiche: Etienne Danchin

2010 – 2012 MSc. Departamento de Botânica, Universidade de Brasília, Brasília, Brasil

Título: Análise da sequencia cloroplastidial *Vigna unguiculata* (L.) Walp

Supervisor: Francisco José Lima Aragão

2006 – 2009 BSc. Ciências biológicas, Faculdades Integradas do Brasil, Curitiba, Brasil

Título: Avaliação molecular de clones de *Eucalyptus camaldulensis* inoculados com *Agrobacterium tumefaciens.*

Supervisora: Marguerite GG Quoirin

## 3. ESTÁGIOS

2014 – 2015 Projeto: Diversidade de virus em plantas daninhas e cultivadas

Financiamento: CNPq

Instituição: EMBRAPA Recursos Genéticos

Supervisores: Roberto Coiti Togawa / Simone Ribeiro

2013 – 2014 Projeto: Desenvolvimento de banco de dados utilizando Drupal

Financiamento: FAP-DF

Instituição: EMBRAPA Genetic Resources

Supervisores: Roberto Coiti Togawa / Ana Brasileiro / Patricia Messemberg

## 4. ARTIGOS COMPLETOS PUBLICADOS

**Genome-wide analysis of expansin superfamily in wild Arachis discloses a stress-responsive expansin-like B gene** (2017). Plant Molecular Biology
Larissa Arrais Guimaraes, Ana Paula Zotta Mota, Ana Claudia Guerra Araujo, Lucio Flavio de Alencar Figueiredo, Bruna Medeiros Pereira, Mario Alfredo de Passos Saraiva, Raquel Bispo Silva, Etienne G. J. Danchin, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

**Early responses to dehydration in contrasting wild *Arachis* species** (2018) PlosOne
Christina Cleo Vinson, Ana Paula Zotta Mota, Thais Nicolini Oliveira, Larissa Arrais Guimaraes, Soraya Cristina Macedo Leal-Bertioli, Thomas Christopher Rhys Williams, Alexandre Lima Nepomuceno, Mario Alfredo Passos Saraiva, Ana Claudia Guerra Araujo, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

**Comparative root transcriptome of wild *Arachis* reveals NBS-LRR genes related to nematode resistance** (2018) BMC Plant Biology
Ana Paula Zotta Mota, Bruna Vidigal, Etienne GJ Danchin, Roberto Coiti Togawa, Soraya CM Leal-Bertioli, David John Bertioli, Ana Claudia Guerra Araujo, Ana Cristina Miranda Brasileiro, Patricia Messenberg Guimaraes

**Contrasting effects of wild *Arachis* dehydrin under abiotic and biotic stresses** (2019) Frontiers in Plant Science
Ana Paula Zotta Mota, Thais Nicolini Oliveira, Christina Cleo Vinson, Thomas Christopher Rhys Williams, Marcos Mota do Carmos Costa, Ana Claudia Guerra Araujo, Etienne GJ Danchin, Maria Fátima Grossi-de-Sá, Patricia Messenberg Guimaraes, Ana Cristina Miranda Brasileiro

## *5.* RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

*Apresentação oral*

**Plant and Animal Genomes** – Root genomics session – San Diego, USA
Identification of evolutionarily conserved plant genes responsive to root-knot nematodes infection

**9eme Journées doctorants SPE** – Montpellier, France
Survey of orthologous genes expression during the nematode incompatible interactions with four plant species

I**II Encontro Científico da RECOBIO/EBB** – Brasília, Brasil
Survey of orthologous genes expression during the nematode incompatible interactions with four plant species

**Workshop Plant-Stress Biotic and Abiotic** – Brasília, Brasil
Identification of evolutionarily conserved genes responsive to nematode infection in four plants

*Apresentação de poster*

**21º Evolutionary Biology Meeting** - Marseille, France
Identification of evolutionary conserved genes responsive to nematode infection in four plant species.

**XXII Encontro do Talento Estudantil – EMBRAPA** – Brasilia, Brasil
Identification of evolutionary conserved genes responsive to nematode infection in four plant species – Prêmio de 3º melhor trabalho na àrea de plantas