Rodriguésia 74: e00492022. 2023
http://rodriguesia.jbrj.gov.br
DOI: http://dx.doi.org/10.1590/2175-7860202374020

Original Paper

The DNAJ gene family in yerba mate (*Ilex paraguariensis*): genome-wide identification, structural characterization, orthology based classification and expression analysis

Patricia Mabel Aguilera[1,2,4,8], *Humberto Julio Debat*[3], *María Lorena Castrillo*[2,5], *Gustavo Angel Bich*[2,6] & *Mauro Grabiele*[1,2,7,8]

Abstract

Dry leaves and twigs of yerba mate are widely infusion-consumed in southern Southamerica. Endemic and adapted to the Atlantic Forest, its extensive full-sun monoculture links to diverse biotic (pest, pathogens) and abiotic stresses (solar radiation, drought), impacting its productivity, ecology and socioeconomic niche. We focused in comprehensively characterize the DNAJ gene family in yerba mate to predict its possible roles on development and diverse stress responses to further assist crop manage. Our results suggest that yerba mate DNAJ proteins account 140 diverse members of six structural types displaying potential variable roles in protein homeostasis control. We were able to classify them into 51 distinct orthology groups, in agreement to *Arabidopsis*, and performed translational genomics of function, localization, expression and stress responsiveness data. Genome mapping and expression analysis indicated that yerba mate DNAJ genes differ in expression, nucleotide composition, length and exon-intron structure. Intronless or few introns genes -linked to rapid stress response- accounted 85 DNAJs. Promoters of DNAJ genes harbored a 73.2% of cis-acting regulatory elements involved in response to diverse stresses, hormones and light, simultaneously. We hypothesize that yerba mate DNAJs assist to plant survival during multiple stresses linked to current dominant agroecosystem but promote its growth under shade.

Key words: chaperones, crop tree, stress genes, translational genomics.

Resumen

Las hojas y ramitas secas de yerba mate son ampliamente consumidas como infusión en el sur de Sudamérica. Endémica y adaptada a la Mata Atlántica, el monocultivo extensivo de esta planta a pleno sol se vincula a diversos estreses bióticos (pestes, patógenos) y abióticos (radiación solar, sequía) que impactan en su productividad, ecología y nicho socioeconómico. El objetivo de este trabajo fue caracterizar exhaustivamente la familia de genes DNAJ en yerba mate a fin de predecir sus posibles roles en el desarrollo y en las respuestas a diversos estreses para así contribuir al manejo del cultivo. Nuestros resultados sugieren que las proteínas DNAJ de yerba mate contabilizan 140 miembros diversos de seis tipos estructurales, con diferentes roles potenciales en el control de la homeostasis proteica. Asimismo, fueron clasificadas en 51 grupos ortólogos distintos, de acuerdo con *Arabidopsis*, y se realizó la genómica traslativa de datos de función, localización, expresión y respuesta a estrés. El mapeo genómico y los análisis de expresión indicaron que los genes DNAJ de yerba mate difieren en expresión, composición nucleotídica, longitud y estructura exón-intrón. Se encontró que 85 genes DNAJ no presentan o poseen pocos intrones -ligados a una rápida respuesta a estrés-. Los promotores de genes DNAJ albergan un 73,2 % de elementos reguladores en cis involucrados en respuesta a diversos estreses, hormonas y luz, simultáneamente. Así, proponemos que las DNAJs de yerba mate asisten a la planta en su supervivencia durante múltiples estreses ligados al actual agroecosistema dominante, mientras que bajo sombra promueven su crecimiento.

Palabras clave: chaperonas, árbol cultivado, genes de estrés, genómica traslativa.

See supplementary material at <http://dx.doi.org/10.17632/294x7524bn.1>

[1] Universidad Nacional de Misiones, Instituto de Biología Subtropical (UNaM-CONICET), Posadas, Misiones, Argentina.

[2] Universidad Nacional de Misiones, Instituto de Biotecnología Misiones "Dra. María Ebe Reca", Sala de Biología Computacional, Posadas, Misiones, Argentina.

[3] Centro de Investigaciones Agropecuarias (INTA), Instituto de Patología Vegetal, Córdoba, Argentina. ORCID: <https://orcid.org/0000-0003-3056-3739>.

[4] ORCID: <https://orcid.org/0000-0002-5063-9350>. [5] ORCID: <https://orcid.org/0000-0002-4999-1672>.

[6] ORCID: <https://orcid.org/0000-0002-3658-5157>. [7] ORCID: <https://orcid.org/0000-0001-7386-4924>.

[8] Author for correspondence: maurograbiele@conicet.gov.ar; patriciamaguilera@gmail.com

Introduction

Dry leaves and twigs of the yerba mate or erva mate tree *Ilex paraguariensis* A. St.-Hil. (Aquifoliaceae) are normally consumed as an infusion called "mate" or "chimarrão" in Argentina, Brazil, Paraguay and Uruguay (Pereira Croge *et al*. 2021). The industry of yerba mate deeply influences the economy of the producer regions of Argentina, Brazil and Paraguay, which harbor about 160,000, 85,000 and 35,000 ha of this crop, respectively (Gortari *et al*. 2020), amounting a production of more than 900,000 tons in 2016 (Pereira Croge *et al*. 2021).

The yerba mate tree is native and adapted to the Atlantic Forest biome environment, however the full-sun monoculture practice of this crop is extensive (Montagnini *et al*. 2011). In this context, yerba mate is negatively affected by diverse insect and acari pests, in addition to fungi and viral pathogens (Sosa *et al*. 2011; Debat *et al*. 2014a; Rybak *et al*. 2014; Bejerman *et al*. 2017, 2020; Bergottini *et al*. 2017). The most important abiotic stress that affects yerba mate in this extensive agricultural system is the direct solar radiation that causes higher evapotranspiration demand and drought stress that ultimately reduces the plantation productivity (Gortari *et al*. 2019, 2020). To cope with those stressful situations and gain yield, diverse breeding programs were carried out in yerba mate which eventually obtained cultivars adapted to diverse environmental conditions (Belingheri & Prat Kricun 1997; Prat Kricun 2010; Stein *et al*. 2014). In addition, traditional mechanical harvesting of yerba mate leaves and twigs cause lesions that reduce its yield and mean life (Kurtz *et al*. 2014). Soil compaction may cause an anaerobic environment that affects the root-associated microbiome to this crop tree also (Bergottini *et al*. 2017). In this sense, model agronomical practices were implemented to maintain plants and soil health (Burtnik *et al*. 2006; Barbaro 2017; Zelada Cardozo & Gonzalez Villalba 2019). However, the multiple stress scenario that affect current yerba mate plantations is expected to be negatively enhanced by the global climate changes (Roeber *et al*. 2021) leading to major socioeconomic problems associated to an unsustainable system (Montagnini *et al*. 2011; MECON 2018). An alternative agroforestry approach, such as the return of yerba mate to its adapted forest conditions, is now considered (Dos Santos 2009; Montagnini *et al*. 2011; Marques *et al*. 2019; Salas *et al*. 2019; Pereira Croge *et al*. 2021), together with the recent study of associated microbes -some of them growth promoters- to the sustainable cultivation of this crop (Bergottini *et al*. 2017; Laczeski *et al*. 2020). As main outcomes of those studies in yerba mate, shade conditions decreased stress, influenced nutrient richness at the leaves and flavour, and increased diversity of the root-associated microbiome.

Plant stress is a state in which plants are exposed to unfavourable environmental or biological conditions that lead to increasing demands made upon it and consequently affecting growth, development and crop yields (Mosa *et al*. 2017). Plant response to stress involves the triggering of mechanisms to sense the stressful signal to enable an optimal growth response. As part of this response, specific transcription factors (TFs) bind to cis-acting regulatory elements (CAREs) at the upstream region of stress responsive genes (Verma *et al*. 2016). Considering the multiple biotic and abiotic stressful factors that could affect plant survival, development and growth, a growing amount of evidence is emerging to highlight the omnipresence of a crosstalk between the diverse response pathways from different stressor signals. In this sense, phytohormone-mediated regulation of stress response is well documented. In general, abscisic acid (ABA) mediates the response to drought, cold, heat and wounding stress, while salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) mediate the response to biotic stress caused by pathogen infection and pests (Bari & Jones 2009; Shi *et al*. 2010; Verma *et al*. 2016). The additional crosstalk between those hormones with auxins and gibberellins (GA) allows plants to cope with stressful situations and sustained growth (Verma *et al*. 2016). In addition, light influences gene expression and all aspects of plant growth and development, and could act as a stressor and take part in the crosstalk with abiotic and biotic pathway stress responses (Petrillo *et al*. 2014; Nawkar *et al*. 2017; Roeber *et al*. 2021). In this sense, environments with reduced light improve drought stress tolerance and the defense response to biotic invaders that affect plant growth (Roeber *et al*. 2021). As a significant outcome, light regulates the unfolded protein response to diverse stresses that affect protein folding, in which a set of molecular chaperones is expressed to alleviate stress (Nawkar *et al*. 2017).

As sessile organisms, plants depend on a complex machinery of specialized proteins to respond to stressful environmental conditions, but also to break through normal development and growth (Vierling 1991; Rajan & D'Silva 2009). Heat shock proteins (HSPs) take part of this machinery as molecular chaperones responsible for protein folding, assembly, translocation and degradation, under different stress conditions and normal cellular processes (Park & Seo 2015). The main plant chaperone families are named according to their approximate molecular weights, such as HSP70 (DNAK), J-protein/HSP40 (DNAJ), HSP60, HSP90, HSP100, and small HSP (sHSP) families (Fragkostefanakis *et al*. 2015). HSP70 chaperones have a major role in protein quality control and require nucleotide exchange factor proteins and the DNAJ proteins acting as co-chaperones for their functions.

The DNAJ proteins recognize unfolded substrates and deliver them to HSP70, stimulating its ATPase activity, which in turn induces conformational changes of HSP70 that stabilizes its interaction with the substrate (Pulido & Leister 2018). Moreover, there is increasing evidence of some HSP70-independent functions of the DNAJ proteins (Rajan & D'Silva 2009; Finka *et al*. 2011; Pulido & Leister 2018). Members of the DNAJ family possess one or more of the following domains: the DNAJ domain, responsible for binding of to the ATPase domain of HSP70, the Zinc finger domain, the C terminal domain, and the DNAJ-like domain (Pulido & Leister 2018). According to the arrangement of these domains, DNAJ proteins can be classified in the following structural types (A to F): A) contains the DNAJ, Zinc finger and C terminal domains; B) contains the DNAJ and C terminal domains; C) contains only the DNAJ domain; D) contains a DNAJ-like domain; E) contains the Zinc finger domain; and F) contains the C terminal domain (Rajan & D'Silva 2009; Finka *et al*. 2011; Pulido & Leister 2018). DNAJ types D, E and F are considered HSP70-independent chaperones and their evolution from canonical DNAJ types A, B and C was proposed (Pulido & Leister 2018).

Currently, the repertoire of DNAJ genes has been catalogued in diverse plants, showing high number and diversity of functions (Sarkar *et al.* 2013; Pulido & Leister 2018; Verma *et al.* 2019). Despite the importance of this stress related genes, the collection of DNAJs in yerba mate and *Ilex* as a whole remains uncharacterized to date.

To sum up, considering the constitutive stressful environmental conditions that yerba mate plantations are exposed to, an integral knowledge on the genetic basis of stress responses will be helpful for future considerations on the management of this crop. In this sense, the main goals of our study were to comprehensively characterize the DNAJ gene family in yerba mate and to predict possible roles of this family on development and diverse stress responses. For this purpose, we carried out a genomewide identification and structural characterization, proposed an orthology based classification, and performed expression analysis of the DNAJ gene family in this important crop tree.

Material and Methods

Identification of yerba mate DNAJs

Trinity-assembled transcriptome sequences from leaves at different stages of a mature tree of *Ilex paraguariensis* grown in monoculture under full sun radiation and other conventional cultivation conditions, deposited at the National Center for Biotechnology Information (NCBI, <https://ncbi.nlm.nih.gov>) under BioProject PRJNA251985, Accession GFHV00000000 (Aguilera *et al.* 2018), were used to build a local search database on the software Geneious 9.1.8 (Biomatters Ltd.). The plant of *I. paraguariensis* used here comprises the cultivar 16318 (CA 538 INTA) of the Instituto Nacional de Semillas (INASE, <https://gestion.inase.gob.ar/consultaGestion/>). It is an elite material with about 2000 planted hectares, showing higher yield of green leaves (17,950 kg/hectares) and lower susceptibility to drought.

Hidden Markov Model (HMM) profile of domains defining DNAJ proteins in *Arabidopsis thaliana* (Pulido & Leister 2018; Zhang *et al*. 2018), such as DNAJ (PF00226), DNAJ central or Zinc finger (PF00684), DNAJ C terminal (PF01556) and DNAJ-X (PF14308), were downloaded from the protein family database Pfam (Mistry *et al*. 2021, <http://pfam.xfam.org/family>) and used as query sequences (Suppl. Tab. 1, available at <http://dx.doi.org/10.17632/294x7524bn.1>) to search against putative yerba mate DNAJ protein genes in GFHV00000000 via BlastP with a cut off value of e-05. Additional BlastP reference searches (cut off e-05) with confirmed *A. thaliana* DNAJs (Pulido & Leister 2018; Zhang *et al*. 2018) were conducted in the same accession to identify potential yerba mate DNAJs with more dissimilar domains than supported by former analysis.

Structural characterization of yerba mate DNAJs

To further characterize accurate nucleotide (nt) composition and full-length transcript sequences of yerba mate, those detected were submitted to an iterative fine tuning back mapping step involving the 72M next generation sequencing (NGS) paired end reads (101 nt) of yerba mate of the mentioned BioProject PRJNA251985, SRA SRP043293 (Debat *et al*. 2014b) using the Geneious mapper (five to 10 rounds) at default values (Chapter 10, Geneious 9.0; <https://assets.geneious.com/documentation/geneious/GeneiousPrimeManual.pdf>) but a word length of 28 and 1% of maximum mismatch per read. Output curated refined and extended transcript sequences were supported by reads contiguity among overlapping reads, pairwise % identity, mean coverage of bases and Q20 values criteria. Coding sequences at transcripts were determined via the open reading frame (ORF) finder tool, subsequently translated and submitted to HMM searches at Pfam via the InterProScan tool to identify and annotate those DNAJ and related domains on proteins. Sequences lacking Pfam DNAJ annotations in the previous step were further compared to Superfamily database via HMM searches as described above, and their domains were properly annotated and checked via in-house BlastP HMM searches (cut off e-01). In addition, sequences lacking Superfamily annotations in the previous step were submitted to in-house BlastP HMM searches (cut off e-01) of DNAJ domains and properly annotated. Putative Yerba mate DNAJs were further classified in structural types (A to F) according to schemes of Pulido & Leister (2018). Subsequently, annotated yerba mate DNAJ proteins were submitted to best BlastP hits searches (cut off e-05) at NCBI Reference Proteins Database taxa sections Viridiplantae and *A. thaliana*, and *I. paraguariensis* BioProject PRJNA315513, Accession GEWR00000000 proteome (Fay *et al*. 2018), and posterior alignments with full-length protein of retrieved hits via Mafft v7.308 at default values. Validation of the full-length of coding sequences and the domain organization of the annotated yerba mate DNAJ proteins based in homology and orthology criteria, where applicable, considering best hits e-value, pairwise % identity and query coverage, added to global protein alignment features, pairwise % positive BLSM62 and domain organization correspondence. In addition, proteome NCBI-GEWR00000000 was also submitted to DNAJ searches and subsequent steps as described above to further characterize possible hits distinctively present among both yerba mate accessions. Protein sequences of identified yerba mate DNAJ genes were analyzed with the Expasy Compute pI/Mw tool (Gasteiger *et al*. 2005, <https://web.expasy.org/compute_pi/>) to obtain molecular weights and theoretical isoelectric points. Further, coding sequences of yerba mate DNAJ proteins were examined for polymorphisms, microsatellites -via Phobos 3.12- and splice variants in Geneious.

Orthology based classification of yerba mate DNAJs

Phylogenetic clustering analysis of yerba mate DNAJs were conducted at Ensembl Plants (Bolser *et al*. 2016) by translating fully curated orthology group information of *A. thaliana* (<https://plants.ensembl.org/Arabidopsis_thaliana/Info/Index>) onto the appropriate yerba mate ortholog. Phylogenetic trees of yerba mate and *A. thaliana* related protein sequences were constructed via multiple alignments -Mafft v7.308 at default values- submitted to Neighbour Joining (NJ) clustering using a p-distance substitution model and 1,000 bootstrap replicates. Additional translation of information between *A. thaliana* and yerba mate DNAJs such as gene functional description and gene ontology (GO), average tissue expression status of genes, and abiotic stress (ABS) response status of genes were conducted at TAIR (Huala *et al*. 2001, <https://www.arabidopsis.org/>), NCBI AceView (Thierry-Mieg & Thierry-Mieg 2006, <https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html>) and *Arabidopsis* eFP (Winter *et al*. 2007, <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>), respectively.

Expression analysis in yerba mate DNAJs

The resulting mappings of aligned reads generated to curate the consensus DNAJs identified in yerba mate were employed to calculate transcript expression levels, using coding sequences as reference in Geneious 9.1.8, measured as fragment per kilo base of exon model per million mapped reads (FPKM) values. A clustered heat map of those log2-transformed FPKM values was obtained using Heatmapper (<http://www.heatmapper.ca/expression/>).

Genomic mapping of yerba mate DNAJs

Finally, curated yerba mate DNAJ transcript sequences were mapped against the 32,521 genomic scaffolds of *I. paraguariensis* NCBI Bioproject PRJEB36685, Accession GCA_905181385.1 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB36685>) using the Geneious mapper at default values but a maximum mismatch per sequence of 3% and a maximum intron size of 30 thousand nt. Genomic validation of curated transcripts was based on global mapping alignment, pairwise % identity and single locus mapping criteria. The locus, direction, size and exon-intron structure of DNAJ genes were properly annotated at the corresponding genomic region. In addition, promoter sequences 1.5 kbp (kilo base pairs) in length upstream of translation start site of DNAJ genes were identified according to Kaur *et al*. (2017) criteria, annotated at the corresponding genomic scaffold, and submitted to an identification of CAREs at PlantCARE (Lescot *et al*. 2002, <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) following the criteria of Lis & Walther (2016) on the orientation of such elements. The analysis focused in CAREs that respond to stresses (ARE, AT-rich sequence, DREs, LTR, MYB, MYC, STRE, TC-rich repeats, W box, WRE3, WUN-motif, etc.), hormones (ABREs, AuxRR-core, CGTCA/TGACG-motif, ERE, GARE-motif, P-box, SARE, TATC-box, TCA-element, TGA-element, etc.) and light (AT1-motif, Box 4, G-Box, GA-motif, GATA-motif, Gap-box, GT1-motif, TCT-motif, etc.) (Pla *et al*. 1993; Seki *et al*. 1997; Sakuma *et al*. 2002; Narusaka *et al*. 2003; Kumar *et al*. 2009; Gao *et al*. 2013; Feng *et al*. 2016; Kaur *et al*. 2017; Su *et al*. 2018, 2021; Baruah *et al*. 2020; Huang *et al*. 2021; Islam *et al*. 2021). In addition, heat shock resposive elements (HSEs) were identified according to criteria of Kumar *et al*. (2009) and Kaur *et al*. (2017) and at PLACE (Higo *et al*. 1999, <https://www.dna.affrc.go.jp/PLACE/?action=newplace>).

Results

Identification of yerba mate DNAJs

BlastP searches in yerba mate NCBI-GFHV00000000 using HMM profile of domains defining HSP40 proteins such as DNAJ (PF00226), DNAJ central (PF00684), DNAJ C terminal (PF01556) and DNAJ-X (PF14308) resulted in ninety-one (91) distinct protein hits with E-values between e-135 to e-06 (Suppl. Tab. 1 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Eighty-nine (89) of these hits were further characterized as yerba mate DNAJs (Tab. 1) while two of them currently belong to A-type reminiscents of DNAJs from aphids and bacteria from aphids (Suppl. Tab. 2 and Suppl. File 1). An extra DNAJ was captured via internal BlastP similarity searches (e-05) with those yerba mate DNAJ hits (Tab. 1). Additional HMM BlastP searches at NCBI-GEWR00000000 proteome produced seventeen (17) additional hits not present among both yerba mate accessions with E-values between e-83 to e-06 (See Suppl. Tab. 1 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Only four of these hits were categorized as proper yerba mate DNAJs (Tab. 1) while three exhibited partial domains and ten originated most probably from psyllids and fungi (Suppl. Tab. 2 and Suppl. File 1 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Furthermore, conducted BlastP searches in yerba mate NCBI-GFHV00000000 using *A. thaliana* as reference to identify potential DNAJs with more dissimilar domains than the ones supported by former analysis resulted in forty-five (45) additional protein hits, added to one hit at NCBI-GEWR00000000 (Tab. 1).

Structural characterization of yerba mate DNAJs

Curated nucleotide composition and potential full-length transcripts of the 140 identified yerba mate DNAJ protein hits were obtained throughout an iterative back mapping of sequence reads strategy (Suppl. File 2). Subsequent downstream analysis were supported by the attained overlapping reads, a mean 99.7% pairwise identity, a mean 84.2 X coverage and a mean 99.0% Q20 value associated to the polished consensus sequences (Suppl. Tab. 3 and Suppl. File 2 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Transcripts length ranged from 401 to 8,707 nt (Suppl. Tabs 3 and 4 available at <http://dx.doi.org/10.17632/294x7524bn.1>), mean 1,667 nt, and their largest recognized ORFs varied from 348 to 7,764 nt in length (Tab. 1 and Suppl. Tab. 5 available at <http://dx.doi.org/10.17632/294x7524bn.1>), mean 1,221.4 nt. Overall, 140 translated ORF sequences (Suppl. Tab. 6 available at <http://dx.doi.org/10.17632/294x7524bn.1>) were then submitted to Pfam to identify and annotate DNAJ and related domains on proteins by which 90 yerba mate DNAJs were confirmed through their highly

Table 1 – Characterization of the Hsp40/DNAJ landscape in yerba mate (*Ilex paraguariensis*).

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
1C3	DNAJ_C	1	720	2	2022.58	239	28.87	9.39	XP_011078468	AT4G07990	HMM
1C24	DNAJ_C	1	1437	1	7997.86	478	52.67	9.11	XP_028099893	AT5G23240	HMM
1C33	DNAJ_C	1	864	1	3164.83	287	33.32	9.50	XP_028096129	AT2G18465	HMM
1C34	DNAJ_C	1	999	-	11378.59	332	37.30	7.95	XP_028122099	AT2G42750	HMM
1C35	DNAJ_C	1	759	4	7605.61	252	29.08	7.02	XP_024986631	AT4G09350	HMM
1C67	DNAJ_C	1	2325	6	1613.18	774	86.76	8.73	XP_010656024	AT5G53150	HMM
1C68	DNAJ_C	1	2289	6	1858.25	762	84.97	8.64	XP_028099578	AT5G53150	HMM
1C72	DNAJ_C	1	909	-	2443.41	302	34.62	9.34	XP_003634709	AT5G23240	HMM
1C74	DNAJ_C	1	2208	7	2742.54	735	81.15	8.72	XP_027081410	AT2G05230	HMM
1C75	DNAJ_C	1	2202	3	1960.22	733	80.95	8.53	XP_027081410	AT2G05230	HMM
1C79	DNAJ_C	1	3219	6	3391.34	1072	119.97	7.06	XP_028061106	AT5G27240	HMM
1C87	DNAJ_C	1	972	4	3481.45	323	35.97	8.68	XP_022849585	AT2G01710	HMM
1C88	DNAJ_C	1	960	2	2597.35	319	35.70	9.28	XP_022849585	AT2G01710	HMM
1D89	DNAJ_D	1	1470	6	2458.81	489	53.86	5.86	XP_028093057	AT5G64360	HMM
1D91	DNAJ_D	1	1692	11	2755.39	563	63.08	7.14	XP_027162935	AT5G64360	HMM
1C102	DNAJ_C	1	2439	13	1073.87	812	90.54	9.30	XP_010656024	AT5G53150	HMM
1D104	DNAJ_D	1	738	1	2952.78	245	27.33	5.29	XP_012443363	AT5G21430	Reference[P,Z]
2B7	DNAJ_B	2	1020	2	2249.41	339	37.12	9.26	XP_010264628	AT3G08910	HMM
2B9	DNAJ_B	2	1041	3	2737.43	346	38.08	9.25	XP_028121398	AT2G20560	HMM
2B10	DNAJ_B	2	1029	4	2301.00	342	37.48	9.29	XP_010264628	AT3G08910	HMM
2B22	DNAJ_B	2	993	7	1466.53	330	36.67	9.44	XP_028120226	AT3G47940	HMM
2B26	DNAJ_B	2	1032	3	4994.70	343	38.61	6.11	XP_027118347	AT3G62600	HMM

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
2C31	DNAJ_C	2	1050	4	2025.49	349	40.34	7.61	XP_028067901	AT5G49060	HMM
2C38	DNAJ_C	2	1053	5	2865.42	350	40.62	7.08	XP_028067901	AT5G49060	HMM
2B39	DNAJ_B	2	1041	2	5082.35	346	38.16	9.15	XP_011096356	AT2G20560	HMM
2B42	DNAJ_B	2	1035	3	7358.99	344	38.81	6.29	XP_027164880	AT3G62600	HMM
2C51	DNAJ_C	2	1065	3	3600.44	354	40.10	8.62	XP_021275443	AT3G57340	HMM
2C60	DNAJ_C	2	1719	4	4168.74	572	62.89	9.25	XP_031257163	AT3G08970	HMM
2B61	DNAJ_B	2	909	2	3642.06	302	33.71	8.98	XP_004236995	AT1G10350	HMM
2B81	DNAJ_B	2	915	3	2003.74	304	33.63	8.95	XP_019240827	AT1G10350	HMM
2F94	DNAJ_F	2	1179	2	497.62	392	44.47	9.44	XP_028105832	AT1G44160	HMM
2C105	DNAJ_C	2	1065	2	3570.93	354	40.25	8.13	XP_002285124	AT3G57340	HMM
3A1	DNAJ_A	3	1683	5	1294.80	560	60.93	9.32	XP_028083440	AT3G17830	HMM
3A11	DNAJ_A	3	1359	-	2204.81	452	50.15	9.04	XP_028072492	AT1G28210	HMM
3A14	DNAJ_A	3	1500	4	2381.70	499	54.04	9.22	XP_011081933	AT1G80030	HMM
3C17	DNAJ_C	3	507	2	25396.16	168	18.92	5.95	XP_016538735	AT1G56300	HMM
3A19	DNAJ_A	3	1254	5	27010.48	417	46.33	5.84	XP_022746015	AT5G22060	HMM
3C27	DNAJ_C	3	849	5	8082.72	282	33.20	9.66	XP_028056066	AT1G77930	HMM
3C28	DNAJ_C	3	417	-	5904.13	138	16.38	4.75	XP_009629854	AT5G16650	HMM
3C29	DNAJ_C	3	573	-	14590.57	190	22.46	5.82	XP_009630693	AT1G71000	HMM
3A44	DNAJ_A	3	1251	5	5962.75	416	46.46	6.52	XP_006445376	AT5G22060	HMM
3A46	DNAJ_A	3	1359	4	4710.27	452	49.76	9.07	XP_028125030	AT5G48030	HMM
3C48	DNAJ_C	3	714	3	19368.67	237	26.32	6.26	XP_028103021	AT3G14200	HMM
3A54	DNAJ_A	3	1323	4	8544.48	440	47.41	9.42	XP_028085432	AT2G22360	HMM
3A58	DNAJ_A	3	1254	4	196408.68	417	46.45	6.15	XP_007010484	AT5G22060	HMM
3A66	DNAJ_A	3	1335	2	7596.58	444	47.88	9.16	XP_011097809	AT2G22360	HMM

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
3A69	DNAJ_A	3	1254	2	12765.85	417	46.51	6.11	XP_007010484	AT5G22060	HMM
3C71	DNAJ_C	3	1608	3	2749.48	535	60.76	7.61	XP_028074693	AT4G37480	HMM
3A84	DNAJ_A	3	1347	2	2185.56	448	49.43	8.77	XP_028125031	AT5G48030	HMM
3C97	DNAJ_C	3	576	2	6729.82	191	22.33	5.15	XP_009630693	AT1G71000	HMM
4C21	DNAJ_C	4	1194	4	6063.14	397	44.25	5.21	XP_002278176	AT1G21080	HMM
4C45	DNAJ_C	4	996	2	7804.92	331	37.26	8.59	XP_028062253	AT4G39150	HMM
4C57	DNAJ_C	4	999	2	2328.15	332	37.58	5.79	XP_028062253	AT4G39150	HMM
4C64	DNAJ_C	4	1143	3	1659.04	380	43.38	5.79	XP_010270365	AT1G77020	HMM
5C5	DNAJ_C	5	453	-	4134.97	150	16.93	9.04	XP_028106280	AT2G17880	HMM
5C23	DNAJ_C	5	471	1	10298.74	156	17.15	9.25	XP_002274505	AT2G17880	HMM
5C30	DNAJ_C	5	501	3	20576.97	166	18.23	10.07	XP_002274505	AT2G17880	HMM
6C6	DNAJ_C	6	1659	6	3997.43	552	63.31	8.63	XP_002529949	AT1G79030	HMM
6C8	DNAJ_C	6	4230	7	1325.07	1409	154.33	6.29	XP_028061517	AT5G12430	HMM
6C12	DNAJ_C	6	1446	2	3050.27	481	53.30	6.31	XP_028051920	AT5G03160	HMM
6C13	DNAJ_C	6	906	2	2231.79	301	35.41	8.85	XP_022723690	AT1G61770	HMM
6C40	DNAJ_C	6	1947	7	4245.56	648	74.08	7.63	XP_028100488	AT3G11450	HMM
6C50	DNAJ_C	6	7764	13	5960.27	2587	282.79	5.91	XP_002281542	AT2G26890	HMM
6C52	DNAJ_C	6	549	-	801.50	182	20.56	4.76	XP_018822468	AT4G10130	HMM
6C56	DNAJ_C	6	2238	6	2588.74	745	82.74	8.98	XP_028100689	AT5G49580	HMM
6C63	DNAJ_C	6	1623	2	2110.83	540	59.48	9.10	XP_028068340	AT2G35720	HMM
6C73	DNAJ_C	6	1839	16	544.93	612	68.85	5.09	XP_027124748	AT1G74250	HMM
6C76	DNAJ_C	6	1656	2	2574.89	551	63.58	8.83	XP_002529949	AT1G79030	HMM
6C78	DNAJ_C	6	4074	6	2252.72	1357	149.82	5.93	XP_010651821	AT5G12430	HMM
7C32	DNAJ_C	7	1233	5	4002.04	410	46.33	8.94	XP_017970522	AT1G24120	HMM

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
7C49	DNAJ_C	7	1245	4	4476.79	414	45.75	5.94	XP_011076517	AT1G68370	HMM
8C77	DNAJ_C	8	1533	4	6116.53	510	55.39	5.66	XP_028106966	AT1G21660	Reference[Z]
8C82	DNAJ_C	8	2154	11	5909.55	717	79.84	6.17	XP_028085607	AT1G75100	Reference[Z]
8C83	DNAJ_C	8	2241	6	7349.11	746	82.07	6.74	XP_028085607	AT1G75100	Reference[Z]
8C98	DNAJ_C	8	5082	14	1739.93	1693	192.49	5.18	XP_034919758	AT1G75310	Reference[Z]
8C99	DNAJ_C	8	1407	6	1064.80	468	52.92	4.90	XP_019078144	AT1G30280	Reference[Z]
8C101	DNAJ_C	8	2862	16	5212.72	953	105.15	5.90	XP_028104503	AT4G12770	Reference[Z]
9E96	DNAJ_E1	9	576	2	3619.55	191	20.60	9.37	XP_022866317	AT2G34860	HMM
9E120	DNAJ_E1	9	939	4	7553.48	312	34.25	6.45	XP_002280630	AT5G61670	Reference[P]
9E121	DNAJ_E1	9	954	-	4469.61	317	34.84	7.92	XP_028121331	AT5G06130	Reference[P]
9E123	DNAJ_E1	9	516	-	2456.74	171	19.26	9.12	XP_018845163	AT3G17668	Reference[P]
9E126	DNAJ_E1	9	432	1	4704.81	143	15.82	8.97	XP_028087295	AT2G24860	Reference[P]
9E127	DNAJ_E1	9	1278	2	2344.55	425	46.98	5.93	XP_028067512	AT2G38000	Reference[P]
9E136	DNAJ_E1	9	1245	1	3576.38	414	46.43	5.09	XP_028108577	AT4G13670	Reference[P]
10E122	DNAJ_E2	10	1203	6	3892.83	400	44.51	6.09	XP_028114241	AT3G57070	Reference[P]
10E131	DNAJ_E2	10	1110	4	15460.20	369	41.02	6.31	XP_028101590	AT1G64500	Reference[P]
10E137	DNAJ_E2	10	780	2	3774.29	259	28.98	8.91	XP_028089693	AT5G13810	Reference[P]
11E128	DNAJ_E1	11	396	-	16667.46	131	13.93	8.50	XP_034700044	AT3G47650	Reference[P]
11E129	DNAJ_E1	11	396	-	2328.15	131	13.71	8.72	XP_028071483	AT3G47650	Reference[P]
11E130	DNAJ_E1	11	429	3	3394.55	142	14.85	8.55	XP_018812873	AT3G47650	Reference[P]
12D108	DNAJ_D	12	348	-	1746.11	115	12.75	9.61	XP_031277242	AT3G59280	Reference[P]
12D109	DNAJ_D	12	351	-	1134.23	116	12.94	9.30	XP_031277242	AT3G59280	Reference[P]
12D113	DNAJ_D	12	351	-	3492.23	116	12.77	9.56	XP_027117214	AT3G59280	Reference[P]
13C59	DNAJ_C	13	633	2	50827.67	210	24.70	9.73	XP_028088068	AT4G13830	HMM

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
13C62	DNAJ_C	13	636	-	3953.47	211	24.30	9.80	XP_028055873	AT4G13830	HMM
13C86	DNAJ_C	13	489	2	42.85	162	19.01	9.30	XP_020222946	AT4G13830	HMM
14C43	DNAJ_C	14	2055	6	4608.72	684	76.14	5.63	XP_028082715	AT1G79940	HMM
14C55	DNAJ_C	14	2058	5	5482.70	685	76.41	5.50	XP_028082715	AT1G79940	HMM
15C4	DNAJ_C	15	495	4	1862.52	164	18.25	5.36	XP_019267692	AT1G80920	iBlastP
15C70	DNAJ_C	15	492	4	2044.23	163	18.23	5.78	XP_016477129	AT1G80920	HMM
16E133	DNAJ_E1	16	615	3	2197.55	204	22.31	5.58	XP_019166008	AT3G19220	Reference[P]
16E134	DNAJ_E1	16	579	4	1085.67	192	21.12	8.35	XP_019166008	AT3G19220	Reference[P]
17D114	DNAJ_D	17	648	1	2425.16	215	25.20	8.78	XP_028091083	AT5G11500	Reference[P]
17D115	DNAJ_D	17	648	2	5060.50	215	25.21	6.94	XP_011081836	AT5G11500	Reference[P]
18E139	DNAJ_E1	18	300	-	17670.68	99	10.12	9.69	XP_003631355	AT5G43260	Reference[P]
18E140	DNAJ_E1	18	300	3	3911.30	99	10.09	9.37	XP_003631355	AT5G43260	Reference[P]
19D80	DNAJ_D	19	1737	3	1791.35	578	64.61	8.89	XP_009616156	AT4G02100	Reference[Z]
20D110	DNAJ_D	20	915	2	3480.78	304	34.15	9.66	XP_002266453	AT1G08640	Reference[P]
21D90	DNAJ_D	21	2073	3	3042.43	690	78.02	6.32	XP_028119685	AT1G18700	HMM
22E119	DNAJ_E1	22	354	-	4261.70	117	12.67	9.83	XP_024026653	AT1G22630	Reference[P]
23C47	DNAJ_C	23	1785	13	3081.38	594	67.46	9.07	XP_022858076	AT1G65280	HMM
24C95	DNAJ_C	24	2160	4	3181.81	719	81.40	5.28	XP_022881860	AT1G69060	Reference[Z]
25E93	DNAJ_E1	25	468	2	5395.05	155	16.51	7.46	XP_012073817	AT1G75690	HMM
26D111	DNAJ_D	26	885	1	7623.72	294	31.17	9.63	XP_028085171	AT2G20920	Reference[P]
27E132	DNAJ_E1	27	348	2	1836.43	115	13.04	9.43	XP_015896200	AT2G24395	Reference[P]
28C100	DNAJ_C	28	339	1	5037.46	112	12.28	10.38	XP_019162613	AT2G35795	Reference[Z]
29C25	DNAJ_C	29	531	1	8207.73	176	19.83	9.59	XP_021685239	AT2G41000	HMM
30C53	DNAJ_C	30	825	1	6235.22	274	30.72	6.31	XP_006344630	AT3G05345	HMM

Name	Type[1]	Orthogroup[2]	ORF length (nt)	ORF SSRs (#)	Transcript expression (FPKM)	Protein length (AA)	Protein molecular weight (kDa)	Protein isoelectric point	Orthologous Viridiplantae locus[3]	Orthologous *A. thaliana* locus[3]	Search method
31C41	DNAJ_C	31	864	2	3334.59	287	32.98	8.16	XP_006490056	AT3G12170	HMM
32C2	DNAJ_C	32	486	2	905.39	161	17.63	8.57	XP_028116989	AT3G13310	HMM
32C15	DNAJ_C	32	474	1	3072.28	157	17.50	9.27	XP_028109809	AT3G13310	HMM
33D112	DNAJ_D	33	2502	4	2512.40	833	93.42	6.93	XP_028120791	AT3G19180	Reference[P]
34E135	DNAJ_E1	34	504	4	4926.54	167	17.84	9.26	XP_028104297	AT3G44020	Reference[P]
35E117	DNAJ_E1	35	492	-	3534.82	163	17.17	9.24	XP_028074428	AT3G45050	Reference[P]
36D106	DNAJ_D	36	837	-	2077.81	278	31.25	10.00	XP_024977429	AT3G51140	Reference[P]
37C16	DNAJ_C	37	771	2	3614.53	256	29.90	8.68	XP_009798146	AT2G42080	HMM
38C85	DNAJ_C	38	234	-	1119.30	77	8.91	8.85	XP_031280830	AT3G62190	HMM
39E124	DNAJ_E1	39	423	2	15900.79	140	14.45	9.72	XP_017242308	AT5G02160	Reference[P]
40C103	DNAJ_C	40	813	1	3505.12	270	31.68	8.34	XP_028069956	AT5G06410	Reference[Z]
41E138	DNAJ_E1	41	342	1	2818.29	113	12.07	8.41	XP_027112285	AT5G15802	Reference[P]
42E125	DNAJ_E1	42	480	-	5456.61	159	16.77	7.97	XP_028102281	AT5G17840	Reference[P]
43C37	DNAJ_C	43	936	4	3716.09	311	35.06	8.27	XP_019153639	AT5G18140	HMM
44C65	DNAJ_C	44	3060	5	1883.07	1019	115.86	5.74	XP_002276957	AT5G18750	HMM
45E116	DNAJ_E1	45	1173	3	2030.02	390	44.60	9.81	XP_002270296	AT5G20220	Reference[P]
46C20	DNAJ_C	46	765	1	4176.98	254	30.33	9.22	XP_022002466	AT5G22080	HMM
47D107	DNAJ_D	47	756	-	19290.41	251	28.17	9.85	XP_028085042	AT5G23040	Reference[P]
48C18	DNAJ_C	48	879	-	3635.26	292	32.78	9.23	XP_022898692	AT5G23590	HMM
49D92	DNAJ_D	49	2400	5	2427.10	799	87.92	4.73	XP_028099492	AT5G42480	HMM
50E118	DNAJ_E1	50	1275	4	5645.09	424	49.02	8.78	XP_019226561	AT5G53860	Reference[P]
51C36	DNAJ_C	51	897	2	3036.72	298	33.70	5.40	XP_028121608	AT5G59610	HMM

[1] DNAJ structural types A to F follow schemes of Pulido & Leister (2018). [2] Orthology groups follow Ensembl_Plants database (<https://plants.ensembl.org/Arabidopsis_thaliana/Info/Index>; accessed at 10-2020). [3] Best BlastP hit at NCBI Reference Proteins database. ORF = open reading frame; nt: nucleotides; FPKM = fragment per kilobase of exon model per million mapped; AA = aminoacides; kDa = kilo Daltons; HMM = Hidden Markov Model. Reference[P] = Pulido & Leister (2018); Reference[Z] = Zhang *et al.* (2018); iBlastP = internal BlastP.

conserved major domains PF00226, PF00684, and PF01556 (Suppl. Tab. 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Those 90 DNAJs were further classified in structural types according to their domains: A (11; DNAJ-Zn finger-C terminal), B (9; DNAJ-Zn finger), C (67; DNAJ), D (2; DNAJ-like) and F (1; C terminal) (Tab. 1 and Suppl. Tab. 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). More dissimilar DNAJ and Zn finger domains were recognized at Superfamily by which 24 additional yerba mate DNAJ proteins were confirmed and classified into structural types C (9; DNAJ; e-04 to e-08), D (4; DNAJ-like; e-06 to e-08) and E (11; Zn finger; e-01 to e-09) (Tab. 1 and Suppl. Tab. 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Finally, the most dissimilar DNAJ and Zn finger domains were identified and annotated via a less stringent BlastP HMM analysis (cut off e-01) through which 26 novel yerba mate DNAJ proteins were recognized and classified into structural types D (10; DNAJ-like; e-01 to e-07) and E (16; Zn finger; e-02 to e-06) (Tab. 1 and Suppl. Tab. 7 available at <https://data.mendeley.com/??????????????????>). Hence, higher e-values of DNAJ and Zn finger domains were associated to mostly reference-found DNAJ-like (12/16) and type E DNAJ (25/27) proteins, respectively (Tab. 1). Moreover, alignments of the four major DNAJ domains among distinct yerba mate DNAJ proteins evidenced high sequence conservation (Suppl. Figs. 1-4 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Particularly, the central HPD motif is consistently present in the 96 DNAJ domain containing proteins, however a total of 16 DNAJ-like domain type D proteins lack HPD, which is also absent in the single F type DNAJ protein (Suppl. Fig. 1 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Regarding the Zn finger domain present in 40 DNAJ proteins of yerba mate, the typical motif CXXCXGXG is well-preserved along the 11 type A and all 27 type E proteins, however it is poorly conserved at both type C DNAJs (See Suppl. Figs. 2 and 3 available at <http://dx.doi.org/10.17632/294x7524bn.1>).

To validate the full-length nature of coding sequences and the domain organization of the 140 annotated DNAJ proteins of yerba mate, best BlastP hits were retrieved from NCBI Reference Proteins Database considering Viridiplantae and *A. thaliana* taxa, and *I. paraguariensis* NCBI-GEWR00000000 (Suppl. Tab. 8). Subsequently, protein alignments were performed for each annotated DNAJ protein of yerba mate and the full-length protein of retrieved hits, those of Viridiplantae and *A. thaliana* harboring database annotated features (Suppl. Tab. 9 and Suppl. File 3 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Identification of true homolog and ortholog proteins of yerba mate DNAJs was achieved based in hits E-value (median e-160), pairwise % identity (mean 79.0%) and query coverage (mean 89.7%), added to global protein alignment features, pairwise % positive BLSM62 (mean 79.2%) and equivalent domain organization (Tab. 1, Suppl. Tabs. 8 and 9, and Suppl. File 3 available at <http://dx.doi.org/10.17632/294x7524bn.1>).

Then, all 140 annotated and curated DNAJ proteins of yerba mate (Suppl. File 4 available at <http://dx.doi.org/10.17632/294x7524bn.1>), ranging from 77 to 2,587 AA in length (mean 406 AA), were submitted to Expasy where their molecular weights and theoretical isoelectric points were obtained (Tab. 1).

Further, coding sequences of yerba mate DNAJ curated transcripts were examined for polymorphisms, microsatellites (SSRs) and splice variants, and all identified features were annotated onto the corresponding sequences (Suppl. Files 5 and 6 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Overall, generated maps of sequence reads (Suppl. File 2 available at <http://dx.doi.org/10.17632/294x7524bn.1>) were scanned for polymorphisms at ORFs, which resulted in 218 sites identified at 84 yerba mate DNAJs (Suppl. Tab. 10 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Transition single nucleotide polymorphisms (SNPs) were found to be most prevalent (67.0%) followed by transversion SNPs (30.7%), while deletion and insertion types account merely to four and one sites, respectively (Suppl. Tab. 10 available at <http://dx.doi.org/10.17632/294x7524bn.1>). In addition, microsatellites varying from 2 to 10 nt were searched at ORF sequences of yerba mate transcripts, which resulted in 461 sites identified at 117 distinct DNAJs (Suppl. Tab. 11 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Tri-nucleotide SSRs constituted the principal class (39.5%) followed by hexanucleotide (22.8%), dinucleotide (12.6%), pentanucleotide (11.3%), tetranucleotide (9.5%), 7-nucleotide (2.8%), 9-nucleotide (0.9%) and 8-nucleotide (0.7%) SSRs, respectively, and their length ranged from 8 to 35 nt, with a mean of 11.4 nt (Suppl. Tab. 11 available at <http://dx.doi.org/10.17632/294x7524bn.1>).

Finally, splice variants at coding sequences were identified via transcripts alignments and/or mapping of sequence reads at four yerba mate DNAJs of types C (2) and E (2), source of different length protein sequences at the same locus (Suppl. File 6 available at <http://dx.doi.org/10.17632/294x7524bn.1>).

Orthology based classification of yerba mate DNAJs

A subsequent phylogenetic clustering of yerba mate DNAJs was conducted at Ensembl Plants by considering well-recognized orthogroups of *A. thaliana*, that is orthologs and paralogs contained at a particular orthology group, and translating this information onto the corresponding yerba mate ortholog protein (Tab. 1). Additional translation of information between *A. thaliana* and ortholog yerba mate DNAJs such as gene functional description, gene ontology (GO), average tissue expression and abiotic stress (ABS) response status of genes were conducted at TAIR, NCBI AceView and *Arabidopsis* eFP (Suppl. Tab. 12 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Overall *Arabidopsis* orthologs to yerba mate DNAJs are well-recognized as DNAJ members according to their TAIR functional description with the exception of 14 loci encoding structural types D and E DNAJs, characterized as such by other approaches (Suppl. Tab. 12 available at <http://dx.doi.org/10.17632/294x7524bn.1>). In addition, with the exception of 12 loci not evaluated to date, total *Arabidopsis* orthologs to yerba mate DNAJs showed a positive fold change value in gene expression during abiotic stress assays, 70 of them upregulated and constituting ABS responsive genes (Suppl. Tab. 12 available at <http://dx.doi.org/10.17632/294x7524bn.1>).

According to the presented phylogenetic approach, a total of 140 yerba mate DNAJs were classified into 51 distinct and coherent orthogroups (Tab. 1). Nomenclature of DNAJs employed here allows to recognize the orthogroup (1–51), the structural type (A–F) and the related number (1–140) of the protein, in that order. Orthogroups 1 to 18 and 32 hold one or more paralog members which comprehend 108 of total yerba mate DNAJs (Tab. 1). Rooted phylogenetic trees based on multiple alignments of full length DNAJ proteins of yerba mate and *A. thaliana* for each of those orthogroups were further built, which retrieved the same protein pair relationships -highly supported- than the ones presented at Table 1, and showed equally supported associations among yerba mate paralogs (Figs. 1-4). DNAJ domains of each one-member orthogroup harboring structural types C and D of yerba mate DNAJs were further aligned altogether with their corresponding *A. thaliana* orthologs and the resultant tree revealed well-supported protein pair relationships, the same than presented at Table 1 (Fig. 3h). The same strategy and results were obtained for one-member orthogroups of type E yerba mate DNAJs carrying Zinc finger domains (Fig. 4f).

Orthogroup 1 is the second largest with 17 DNAJ members of C (14) and D (3) structural types (Tab. 1 and Fig. 1a). Eight DNAJ paralogs of varying lengths (239 to 563 AA) harbor sole DNAJ domains at different protein positions (1C3, 1C33, 1C35, 1C87, 1C88, 1C89, 1D91 and 1D104). Particularly, DNAJs with marked similar domain organization were found to cluster together such as those harboring additional Fer4_13 domain (1C24, 1C72 and 1C34) and those with extra DUF3444 domain (1C67, 1C68, 1C74, 1C75 and 1C102), respectively, denoting a common ancestor for each subgroup of 1. Protein 1C79 is particular in having an arrange of two DUF3444 domains which accounts for being one of the largest DNAJs in yerba mate (1,072 AA). Orthogroup 2 is the third largest and contains 15 DNAJ members of types B (9), C (5) and F (1) (Tab. 1, Fig. 1b). Type C DNAJ proteins clustered together, including those with additional DUF1977 domain (2C31, 2C38, 2C51 and 2C105) and the largest 2C60. Protein 2F94 constitutes the unique type F DNAJ in yerba mate, which has a non-functional DNAJ domain lacking the HPD motif but a conserved C terminal domain. Type B DNAJs (2B7, 2B9, 2B10, 2B22 2B26, 2B39, 2B42, 2B61 and 2B81) limited to orthogroup 2 and exhibited similar lengths (mean 333 AA) and domain organization but arranged into three different subgroups. Orthogroup 3 of DNAJs is the largest in yerba mate with 18 members of types A (11) and C (7) (Tab. 1, Fig. 2a). Overall C-type members have varying protein lengths (138 to 535 AA) but grouped together (3C17, 3C27, 3C28, 3C29, 3C48, 3C71 and 3C97). Type A DNAJs are exclusively found in orthogroup 3 of yerba mate and all hold the Zinc finger domain embedded into the C terminal domain (3A1, 3A11, 3A14, 3A19, 3A44, 3A46, 3A54, 3A58, 3A66, 3A69 and 3A84). In addition, those A-type DNAJs exhibited similar lengths and domain organization but grouped into three distinct clusters of paralogs. Orthogroup 4

consists in four C-type DNAJ members (4C21, 4C45, 4C57 and 4C64) of similar lengths (331 to 397 AA) which particularly harbor an extra DNAJ-X domain (Tab. 1; Fig. 2b). Orthogroup 5 is formed by three structural type C DNAJs, small in size (151 to 166 AA) and with central DNAJ domains (5C5, 5C23 and 5C30), while orthogroup 7 comprises two C-type members (7C32 and 7C49) with about410 AA and DNAJ domains near the N-terminus of the protein (Tab. 1; Figs. 2c,e, respectively). Orthogroup 6 is the fourth largest in yerba mate with 12 type C protein members of extreme varying lengths (14-fold; 182 to 2587 AA), all but 6C13 and 6C40 harboring additional domains than DNAJ and displaying distinct domain organizations (Tab. 1; Fig. 2d). Protein 6C12 holds two TPR-like domains at central region while 6C63 ports a terminal DUF3395 domain and 6C50, by far the largest DNAJ in yerba mate with 2,587 AA, presents a central GYF_2 domain. Further, the most similar domain organizations of DNAJs also grouped together such as those that include an additional jiv90 domain (6C6, 6C56 and 6C76) and those with two extra and large TPR-like domains (6C68 and 6C78), respectively. Apart from the conserved DNAJ domains, proteins

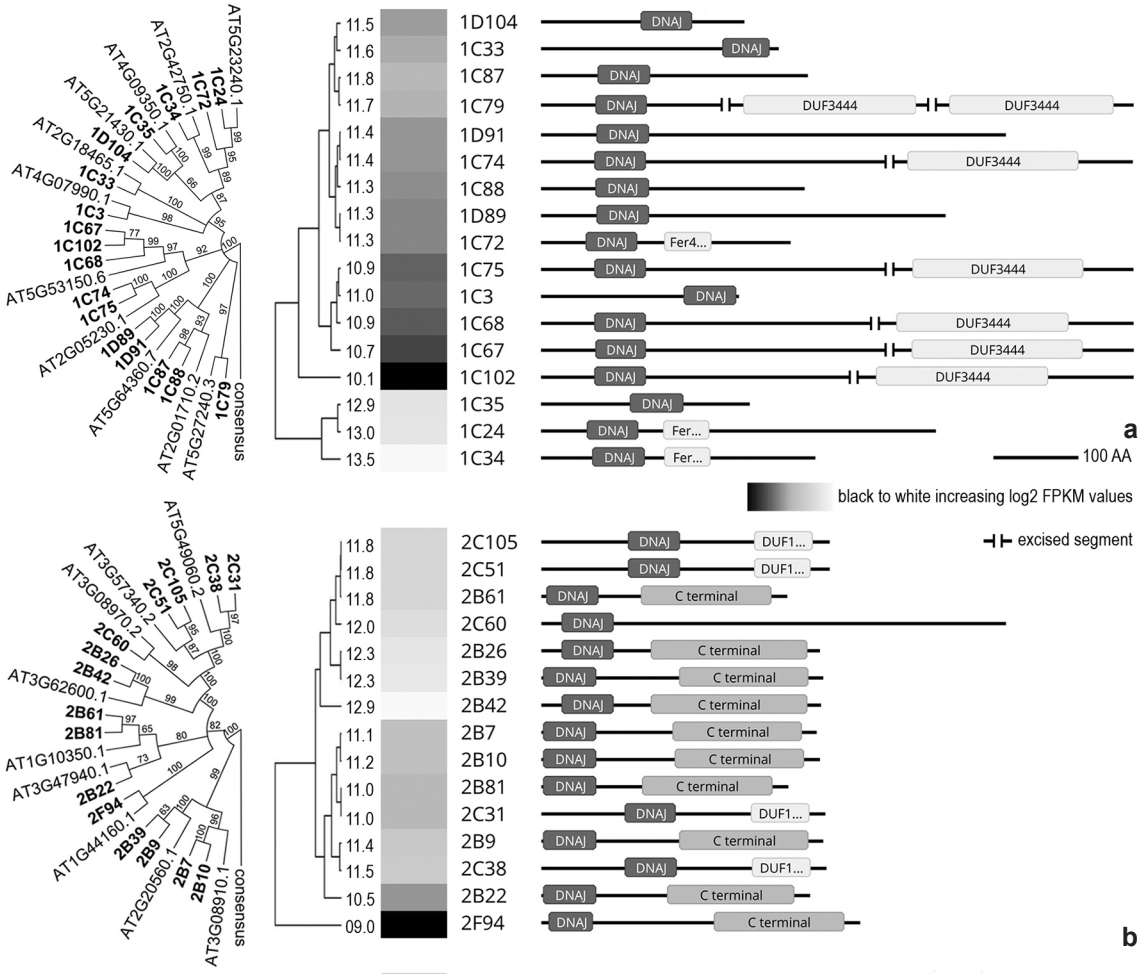

Figure 1 – a-b. Phylogenetic tree of yerba mate and *Arabidopsis thaliana* ortholog DNAJ proteins (left), clustered heat map of the expression profile of yerba mate DNAJs (middle) connected to the domain organization of those proteins (right). Full-length protein sequences were aligned via Mafft, submitted to NJ clustering using p-distance, 1,000 bootstrap replicates (support values denoted at the tree) and rooted with the group consensus sequence. log2-transformed FPKM values of the expression of yerba mate DNAJs are represented left to the heat map. AA, aminoacides – a. Orthogroup 1; b. Orthogroup 2.

6C52 and 6C73 also harbor vestigial Zinc finger domains. Orthogroup 8 with six C-type DNAJs is characterized by medium to large length proteins (468 to 1,693 AA) which hold a single DNAJ domain at the C-terminus of the protein (8C77, 8C82, 8C83, 8C98, 8C99 and 8C101) (Tab. 1; Fig. 3a). Similar domain organization DNAJs harboring additional domains to DNAJ were found

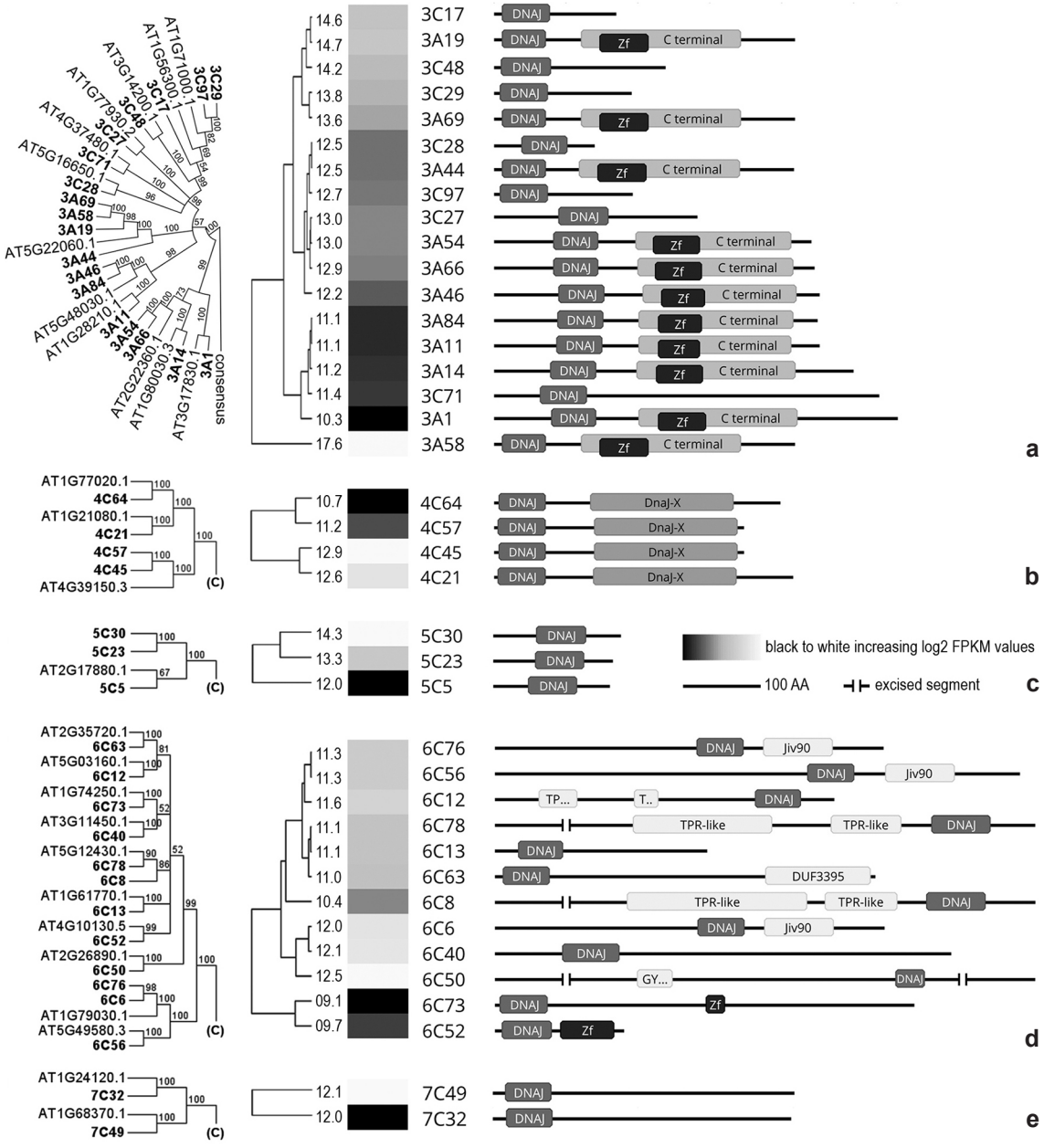

Figure 2 – a-e. Phylogenetic tree of yerba mate and *Arabidopsis thaliana* ortholog DNAJ proteins (left), clustered heat map of the expression profile of yerba mate DNAJs (middle) connected to the domain organization of those proteins (right). Full-length protein sequences were aligned via Mafft, submitted to NJ clustering using p-distance, 1,000 bootstrap replicates (support values denoted at the tree) and rooted with the group consensus sequence (c). log2-transformed FPKM values of the expression of yerba mate DNAJs are represented left to the heat map. AA, aminoacides – Orthogroups 3, 4, 5, 6 and 7, respectively.

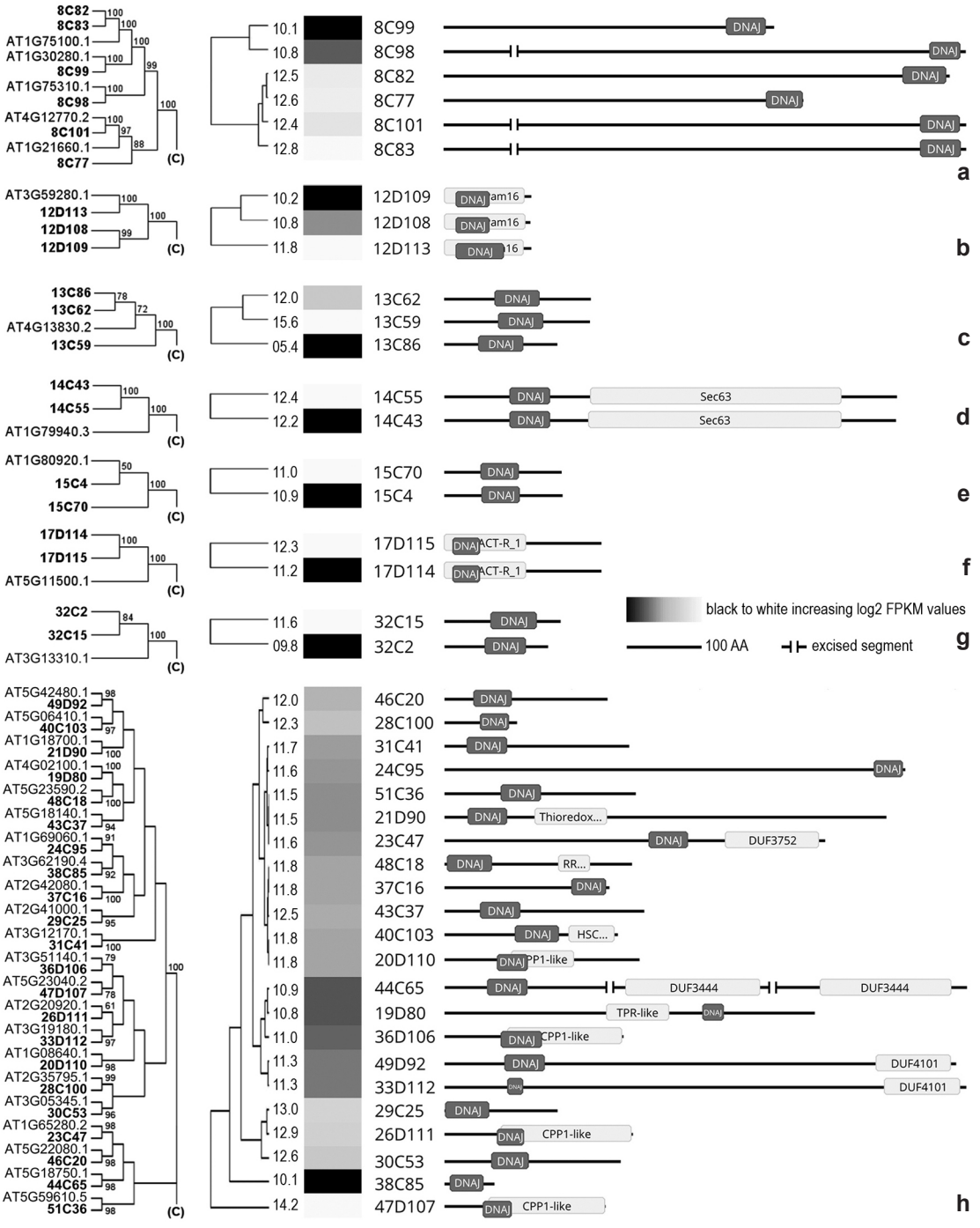

Figure 3 – a-h. Phylogenetic tree of yerba mate and *Arabidopsis thaliana* ortholog DNAJ proteins (left), clustered heat map of the expression profile of yerba mate DNAJs (middle) connected to the domain organization of those proteins (right). Full-length protein sequences were aligned via Mafft, submitted to NJ clustering using p-distance, 1,000 bootstrap replicates (support values denoted at the tree) and rooted with the group consensus sequence (c). log2-transformed FPKM values of the expression of yerba mate DNAJs are represented left to the heat map. AA, aminoacides – a-g. Orthogroups 8, 12, 13, 14, 15, 17 and 32, respectively; h. One-member orthogroups of structural types C and D.

at orthogroups 12 (D-type; 12D108, 12D109 and 12D113), 14 (C-type; 14C43 and 14C55) and 17 (D-type; 17D114 and 17D115), domains such as Pam16, Sec63 and NFACT-R_1, respectively (Tab. 1; Figs. 3b, d and f, respectively). In addition, structural type C DNAJ proteins of orthogroups 13

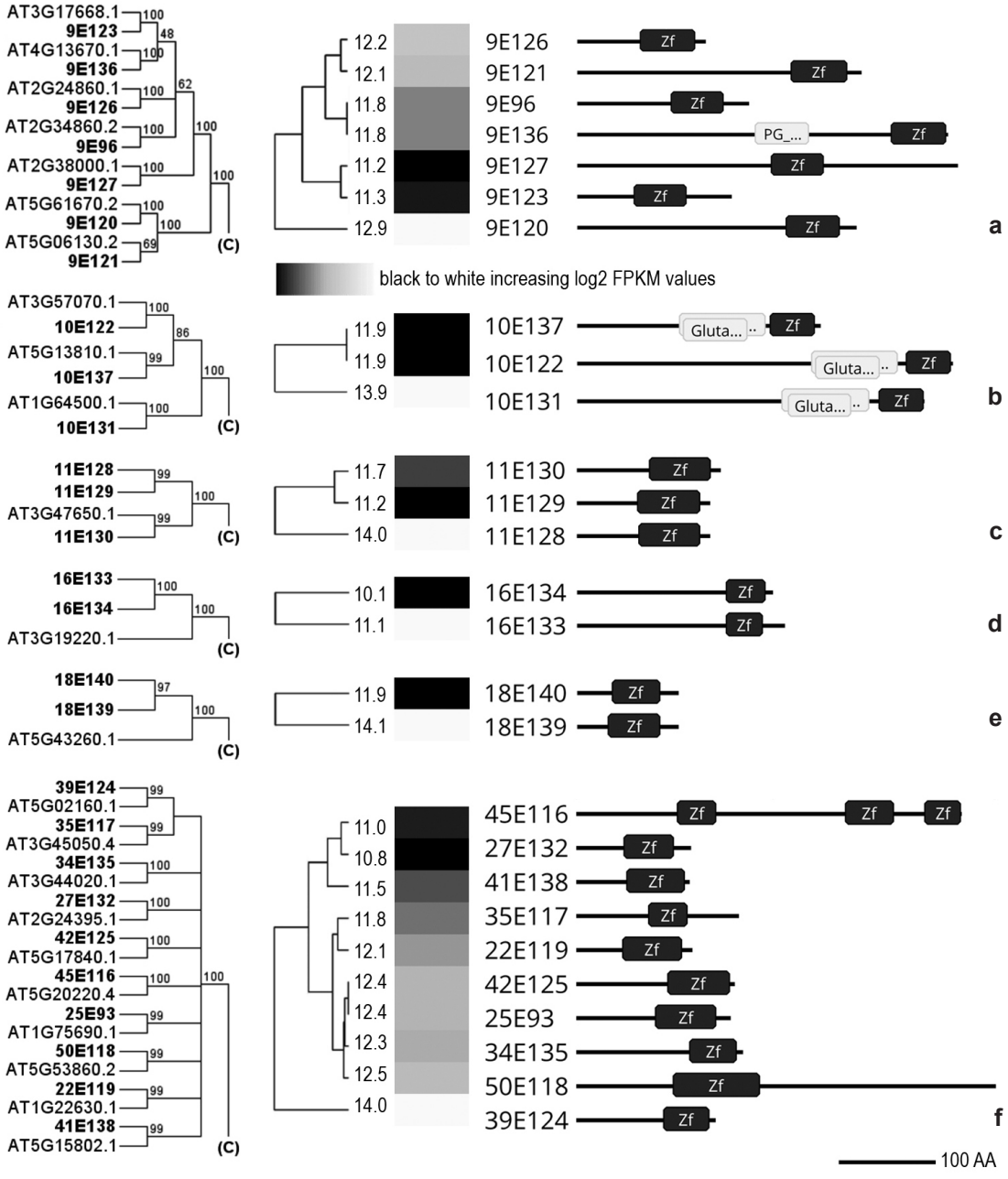

Figure 4 – a-f. Phylogenetic tree of yerba mate and *Arabidopsis thaliana* ortholog type E DNAJ proteins (left), clustered heat map of the expression profile of yerba mate DNAJs (middle) connected to the domain organization of those proteins (right). Full-length protein sequences were aligned via Mafft, submitted to NJ clustering using p-distance, 1,000 bootstrap replicates (support values denoted at the tree) and rooted with the group consensus sequence (c). log2-transformed FPKM values of the expression of yerba mate DNAJs are represented left to the heat map. AA, aminoacides – a-e. Orthogroups 9, 10, 11, 16 and 18, respectively; f. One-member orthogroups of structural type E.

(13C59, 13C62 and 13C86), 15 (15C4 and 15C70) and 32 (32C2 and 32C15) are defined by a small size (mean 175 AA) and a single central DNAJ domain (Tab. 1; Figs. 3c, e and g, respectively).

One-member orthogroups of structural types C (14) and D (8) of yerba mate DNAJs are diverse in length (77 to 1,019 AA) and domain organizations (Tab. 1; Fig. 3h). Proteins with a sole DNAJ domain (10) are exclusive of C-type (46C20, 28C100, 31C41, 24C95, 51C36, 37C16, 43C37, 29C25, 30C53 and 38C35). Other type C DNAJs have, besides DNAJ, additional domains such as DUF3752 (23C47), RRM_1 (48C18), HSCB_C (40C103) and DUF3444 (44C65). Moreover, all type D DNAJs harbor additional domains for instance Thioredoxin-like (21D90), CPP1-like (36D106, 47D107, 20D110 and 26D111), TPR-like (19D80) and DUF4101 (49D92 and 33D112).

Overall, 27 type E DNAJ proteins of yerba mate are contained within 15 distinct orthogroups (Tab. 1 and Fig. 4). Orthogroup 9 is the largest with seven DNAJs ranging from 143 to 425 AA and Zinc finger domains at different protein positions (Tab. 1; Fig. 4a). Apart from 9E136 wich holds a PGBD-like additional domain, the rest of the proteins port a unique Zinc finger domain (9E96, 9E120, 9E121, 9E123, 9E126 and 9E127). Orthogroup 10 members (10E122, 10E131 and 10E137) characterized by terminal Zinc finger domains added to a central Glutaredoxin domain, are classified as type E2 DNAJs (Tab. 1 and Fig. 4b). Orthogroups 11 (11E128, 11E129 and 11E130), 16 (16E133 and 16E134) and 18 (18E139 and 18E140) exhibited small size proteins ranging from 99 to 204 AA (Tab. 1; Figs. 4c-e). Finally, one-member orthogroups of structural type E (10) are diverse in length (113 to 424 AA) and hold a single Zinc finger domain, at central to terminal protein positions (25E93, 45E116, 35E117, 50E118, 22E119, 39E124, 42E125, 27E132, 34E135, 41E138) (Tab. 1 and Fig. 3f). Particularly 45E116 exhibited three Zinc finger domains which accounts to being the second largest (390 AA) after 50E118.

Expression analysis in yerba mate DNAJs

Generated mapping files of aligned reads for each of the 140 well-recognized yerba mate DNAJs (Suppl. File 2 available at <http://dx.doi.org/10.17632/294x7524bn.1>) were employed to calculate transcript expression levels -FPKM- using curated coding sequences as reference (Tab. 1; Suppl. Tab. 13 available at <http://dx.doi.org/10.17632/294x7524bn.1>). A clustered heat map of the expression profile of overall DNAJs in yerba mate leaves was attained through those log2-transformed FPKM values (Suppl. Tab. 13; Suppl. Fig. 5 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Relative abundance of DNAJ transcripts ranged from 8.96 to 17.58 (mean 11.85), excluding the biased lower value for 13C86 (Suppl. Tab. 3). Highest expression levels were associated to 15 yerba mate DNAJs of structural types A (3A58, 3A19 and 3A69), C (13C59, 3C17, 5C30, 3C48, 3C29, 1C34 and 5C23), D (47D107) and E (18E139, 11E128, 39E124 and 10E131), predominantly from orthogroup 3. When compared by structural types, mean expression profiles of A (12.76) and E (12.07) DNAJ proteins are the only that surpass the overall mean. Specific clustered heat maps of expression values for orthogroups 1 to 18 and 32 were further generated which denoted the relative abundance of paralogs within each group, respectively (Figs. 1-4). Same strategy was applied to those one-member orthogroups of structural types C and D, added to E of yerba mate DNAJs, to further compare their expression levels, respectively (Figs. 3h, 4f).

Genomic mapping of yerba mate DNAJs

A total of 140 curated yerba mate DNAJ transcript sequences were mapped against the 32,521 genomic scaffolds of *I. paraguariensis* NCBI-GCA_905181385.1. In this sense, all the DNAJ transcripts were found to have a genomic counterpart: 129 were unambiguously mapped to a single genomic region while three mapped to highly similar genomic regions in two or more scaffolds in addition to eight which different parts mapped to different scaffolds, mostly at the ends of those genomic sequences (Suppl. Tab. 14; Suppl. File 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Particularly, large scaffolds 4, 6, 38, 61, 188 and 275 carry two or more DNAJ genes (Suppl. Tab. 14; Suppl. File 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). DNAJ genes ranged from 0.4 (18E139) to 46.6 (1C34) kbp in length, with 94 genes under the mean of 8.4 kbp (Suppl. Tab. 14 and Suppl. File 7). In addition, the exon-intron structure of DNAJ genes of yerba mate was found to be deeply variable,

ranging from none (13 genes) to 21 introns (6C50), with 85 genes under the mean of 5 introns (Suppl. Tab. 14; Suppl. File 7 available at <http://dx.doi.org/10.17632/294x7524bn.1>). A minor positive linear association was found among gene size and intron number at yerba mate DNAJs according to the R square (R^2) coefficient value (0.39), however the association was strong (R^2=0.99) among gene size and intron fraction at those DNAJ genes. In addition, a negligible positive linear associations ($R^2 < 0.01$) were found among intron number and expression level, added to intron fraction compared to expression level at yerba mate DNAJs genes.

Finally, promoter sequences 1.5 kbp in length upstream of the putative translation start site of each yerba mate DNAJ gene were identified and annotated at the corresponding genomic scaffolds (Suppl. Files 7 and 8) and submitted to searches for regulatory elements. As a main result, all the 140 DNAJ gene loci were predicted to have CAREs at their promoter regions (Suppl. Tab. 15 and Suppl. File 9). According to their functions those 117 different CAREs could be further grouped into six major categories such as 1-stress response (24), 2-hormone response (20), 3-light response (31), 4-cellular development (13), 5-core cis-element (4) and 6-unknown function (25), with a mean of 26.6 different CAREs at each yerba mate DNAJ gene promoter (Suppl. Tab. 16 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Stress response accounts for a 34.8% of the total CAREs at promoter of yerba mate DNAJs followed by light response (20.7%), hormone response (17.7%), core cis (11.5%), cellular development (9.4%) and unknown function (5.9%) elements. Drought stress response CAREs such as MYB, MYC, MBS, DRE and their related sequences were present at 138 DNAJ genes. MBS is also involved in high salt and low temperature stresses, and always linked to other drought CAREs was found at 56 DNAJ genes. LTR and different HSE sequences, involved in low temperature and heat stress responsiveness were found at 44 and 17 yerba mate DNAJ genes, respectively. In total, 82 DNAJs were revealed as low temperature responsive genes via MBS and/or LTR CAREs. In addition, 96 DNAJs harboring ARE comprised anaerobic responsive genes. Defense stress responsive CAREs such as as-1, W box, WRE3, WUN-motif, TC-rich repeats, CCAAT-box, box S and AT-rich sequence, involved in the protection of plants from attacks by diverse pests and pathogens and/or wounding by herbivores and environmental mechanical stresses, were largely present at 133 yerba mate DNAJ genes. It is worth to mention here our findings of those sequences at yerba mate transcriptomes (GFHV00000000, GEWR00000000) that ultimately belong to aphids, bacteria from aphids, psyllids and fungi commonly linked to this crop (Suppl. Tab. 2 available at <http://dx.doi.org/10.17632/294x7524bn.1>). Light responsive CAREs were found at 139 DNAJ genes, and major elements included Box 4 (112), G-Box (102), GT1-motif (73), TCT-motif (58), GATA-motif (53) and TCCC-motif (35) among others. Of the 20 types of hormone responsive CAREs, those that involve ABA such as ABRE and related sequences (ABRE2, ABRE3a, ABRE4, AT~ABRE) were major, ocurring at 94 yerba mate DNAJ genes. Cis elements CGTCA-motif and TGACG-motif involved in the JA responsiveness were found at 84 genes, while the ET responsive element ERE harbor 80 genes, and SA responsiveness CAREs such as TCA-element and SARE occurred at 49 yerba mate DNAJ genes. GA responsive elements P-box, GARE-motif and TATC-box were found at 70 genes, while auxin responsive elements such as AuxRR-core, AuxRE, TGA-box and TGA-element, were present at 43 DNAJ genes.

Discussion

Genome wide identification and structural characterization of yerba mate DNAJs

The DNAJ gene family is characterized by highly conserved protein members among diverse organisms (Sarkar *et al*. 2013; Pulido & Leister 2018; Verma *et al*. 2019) such that true comparisons, inferences and translational genomics of annotations and functions are possible. In this context, DNAJs of *I. paraguariensis* could be identified according to conserved domains defining HSP40 proteins (DNAJ, DNAJ central, DNAJ C terminal, DNAJ-X) via a search strategy based on HMM profile of those Pfam domains (Mistry *et al*. 2021), or by linking shared features with confirmed DNAJ proteins of *Arabidopsis* (Pulido & Leister 2018; Zhang *et al*. 2018) directly through a reference-based search approach. Hence, in agreement to those identified and further characterized NGS-derived sequences, the DNAJ gene family in yerba mate comprises at least 140 distinct members. Similar genome wide analysis of the DNAJ family were accomplished in representative taxa of such diverse range as

monocots, rosids and asterids which found 104 to 115 genes in rice (Sarkar *et al*. 2013; Luo *et al*. 2019), 89 to 120 in *Arabidopsis* (Miernyk 2001; Rajan & D'Silva 2009; Finka *et al*. 2011; Zhang *et al*. 2018) and 76 in chili pepper (Fan *et al*. 2017), respectively, always from structural types A to D proteins.

Further, in yerba mate the central HPD motif at the DNAJ domain is consistently present at the 96 DNAJ proteins of A, B or C types, classified here according to the stringent criteria of Rajan & D'Silva (2009) on the occurrence of that motif. These 96 DNAJs that could interact with HSP70 chaperones mainly via HPD conserved motif are also classified as HSP70 dependent proteins, according to the criteria of Pulido & Leister (2018). On the contrary, those DNAJ proteins of yerba mate with DNAJ domains that lack the central HPD motif are recognized as HSP70 independent DNAJ-related ones, such as the 16 type D proteins, which still may support the folding of substrates according to those authors.

In their work, Pulido & Leister (2018) also recognized novel HSP70 independent DNAJ-related proteins in *Arabidopsis*, including conventional structural type D proteins harboring DNAJ-like domain and others. Regarding that study, novel types E and F DNAJs harbor a single Zinc finger or C terminal domains, lacking the function of protein quality control of the DNAJ domain but being still important for substrate binding, which may have evolved from types A or B DNAJs, respectively. In view of this last update, *Arabidopsis* currently reaches 157 DNAJ genes and holds the most extensive classification of DNAJs in plants regarding types A to F (Pulido & Leister 2018; Zhang *et al*. 2018). In agreement to this comprehensive classification, that we followed throughout this work, we recognized and characterized additional HSP70 independent DNAJ-related proteins in yerba mate beyond type D. Those novel DNAJs of yerba mate of type E (27) harbor mainly a single Zinc finger domain while the single type F protein at orthogroup 2 carries a C terminal domain and a less conserved DNAJ-like domain lacking the HPD tripeptide. Hence and according to their structures, overall 140 yerba mate DNAJ proteins could be classified into types A to F (11 A, 9 B, 76 C, 16 D, 27 E, 1 F) comparable to *Arabidopsis* (Pulido & Leister 2018; Zhang *et al*. 2018), displaying potential variable roles in the control of protein homeostasis.

Orthology based classification and expression analysis of yerba mate DNAJs

Yerba mate DNAJ proteins showed medium to high global similarity and equivalent domain organization to corresponding *Arabidopsis* orthologs despite the 125 million years of the split of Asterids and Rosids (Guyot *et al*. 2012) to which these species belong to, respectively. Total *Arabidopsis* orthologs to yerba mate DNAJs are well-recognized DNAJ members according to their gene functional description complemented to GO data at TAIR (Huala *et al*. 2001) and by the detailed contributions of diverse authors (Miernyk 2001; Rajan & D'Silva 2009; Finka *et al*. 2011; Pulido & Leister 2018; Zhang *et al*. 2018). The subsequent phylogenetic clustering approach considering those total 140 yerba mate DNAJs and ortholog proteins of *Arabidopsis* properly grouped at Ensembl Plants (Bolser *et al*. 2016) showed 51 distinct and coherent DNAJ orthogroups in yerba mate which shed light on the evolution of structural types and domains organization added to paralog relationships in compound DNAJ clades. As a whole, this approach contributed to the understanding of the DNAJ proteins scenario in yerba mate, and is comparable in part to those of Pulido & Leister (2018) and Zhang *et al*. (2018) for *Arabidopsis* which still wait for a unified orthology based classification of DNAJs.

Additionally, the widespread correspondence between yerba mate and *Arabidopsis* ortholog DNAJ proteins allowed a precise interspecific translation of information between the crop tree and the model species, such as prediction of subcellular localization of DNAJ proteins and the ABS responsive status of DNAJ genes. At this regard, according to GO data at TAIR (Huala *et al*. 2001) and Zhang *et al*. (2018), *Arabidopsis* DNAJ proteins orthologs to yerba mate can function in different compartments, *i.e*. the cytosol, nucleus, organelles. Those DNAJs with similar localization may come from different clades and are involved in similar cellular events but with different roles (Zhang *et al*. 2018). Considering abiotic stress assays including cold, drought, genotoxic, heat, osmotic, oxidative, salt, UV-B and wounding treatments, those 128 *Arabidopsis* orthologs to yerba mate DNAJs with ABS data showed a positive fold change value in gene expression, and in particular 70 of those DNAJs were upregulated and further considered as ABS responsive genes (Kilian *et al*. 2007).

As a main outcome, expression profiles based on RNA-seq showed that overall 140 DNAJs were expressed in leaves of yerba mate. In addition, differences in gene expression were revealed, with 61 yerba mate DNAJs above the mean, and highest expression levels associated to 15 DNAJs of nine diverse clades (1, 3, 5, 10, 11, 13, 18, 39, 47) and structural types (A, C, D, E). *Arabidopsis* DNAJs orthologs to those of yerba mate with data on average tissue expression, such as AT2G42750:1C34, AT5G22060:3A19:3A58:3A69, AT2G17880:5C23:5C30, AT1G64500:10E131, AT3G47650:11E128, AT4G13830:13C59, AT5G43260:18E139 and AT5G02160:39E124, also showed high expression levels, while AT1G56300:3C17 and AT3G14200:3C48 are well expressed ABS genes, compared to moderately and low expressed DNAJs in this model species (Thierry-Mieg & Thierry-Mieg 2006; Kilian *et al*. 2007). Considering the 19 multiple DNAJ clades of yerba mate, mean gene expression levels above the overall mean were found at orthogroups 3 (types A or C), 5, 7 and 14 (type C), and 9, 10, 11 and 18 (type E) while most type B and the single F DNAJs exclusives of orthogroup 2 are the less expressed. In addition, within orthogroup differences at gene expression were also found at all compound DNAJ clades of yerba mate, even at related paralogs of the same structural type and similar domain organization such as those of the largest orthogroups 1, 2 and 3, which are expected to have redundant functions as occurs in *Arabidopsis* (Zhang *et al*. 2018). The availability of additional RNA-seq datasets of yerba mate derived from diverse environmental and stress conditions could provide further insights into the diversity and expression of DNAJs in this crop and complement the foundational baseline landscape of yerba mate DNAJs reported here.

Genomic mapping of yerba mate DNAJs

As a whole, 140 yerba mate DNAJ genes showed a great variability in nucleotide composition, length and their exon-intron structure comparable to other plant species such as *Capsicum annuum*, *A. thaliana* and *Oryza sativa* (Fan *et al*. 2017; Zhang *et al*. 2018; Luo *et al*. 2019). Close DNAJ paralogs of yerba mate that hold similar protein domain organization were found to share a comparable gene structure, such as 1C67:1C68:1C74:1C75:1C102, but the contrary also occurred, i.e. 2C31:2C38:2C51:2C105, as reported in *Arabidopsis* (Zhang *et al*. 2018). On the other hand, increased expression levels of yerba mate DNAJ genes were somewhat accompanied to increased intron number or fraction at those genes, in contrast to strong associations reported in *Arabidopsis* and rice considering a wide gene approach (Ren *et al*. 2006). Size of intron-containing yerba mate DNAJ genes strongly associated to the intron fraction of that gene. Further, the percentage of intronless DNAJ genes in yerba mate (13; 9.9%) was the lowest if compared to 32.9% in chili pepper (Fan *et al*. 2017), 22.2% in thale cress (Zhang *et al*. 2018) and 20.0% in rice (Luo *et al*. 2019). To stand out, yerba mate DNAJ genes with few introns (1–3) accounted for 45 members, and 85 genes were under the overall mean of five introns (0–5). This is important to the DNAJ genes of yerba mate since intronless or fewer intron genes were found to be rapidly regulated to respond timely to various stresses (Jeffares *et al*. 2008).

The CAREs are non-coding DNA sequences in gene promoters that control the transcription of their accompanying genes, and have been reported elsewhere, inclusive at DNAJ genes (Hernandez-Garcia & Finer 2014; Fan *et al*. 2017; Huang *et al*. 2021). Promoter regions of total 140 yerba mate DNAJs harbor diverse CAREs involved in the response to stresses (drought, high salt, low temperature, heat, anaerobic, defense), hormones (ABA, JA, ET, SA, GA, auxins) and light simultaneously, constituting the 73.2% of total cis elements in this family, while another 20.9% of those CAREs with known function are related to transcription efficiency or growth. These results are consistent with an scenario were crops require to adapt to such adverse situations and grow, by evolving crosstalking mechanisms among hormones, ligth and diverse stresses (Verma *et al*. 2016; Nawkar *et al*. 2017; Roeber *et al*. 2021). As a result of responding to diverse environmental and endogenous factors, yerba mate DNAJs may be involved in a variety of stressful and physiological processes and in coordinating plant growth under adverse conditions, as reported previously in other eukaryotes (Finka *et al*. 2011). As a whole, our study suggests that the 140 DNAJs characterized of yerba mate are potentially responsive genes to cope with the well-known major stresses that persistently affect this crop in the context of an universal full-

sun monoculture practice, such as those abiotic (Rakocevic *et al.* 2008; Gortari *et al*. 2019, 2020; Salas *et al.* 2019), biotic (Sosa *et al*. 2011; Rybak *et al*. 2014; Bejerman *et al*. 2017; Bergottini *et al*. 2017) and mechanical (Kurtz *et al*. 2014) shocks. We hypothesize that the large repertoire of DNAJ genes and these cis-acting regulatory elements are concomitant to the survival of the yerba mate plant during those multiple stresses. While in shade conditions DNAJs may be involved primarily in promoting growth in this crop tree, we hypothesize that DNAJs assist to yerba mate survival during multiple stresses associated to the current dominant agroecosystem. Recently, Salas *et al*. (2019) measured the behaviour of yerba mate plantlets under a light gradient varying from full-sun culture to conditions of Atlantic Forest. This experiment offered the basis to test our hypothesis, considering mature trees. In sum, our results introduce for the first time an exhaustive analysis of yerba mate DNAJs and provide functional and evolutionary insights into this important family of plant chaperones.

Acknowledgements

The authors would like to thank the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina); the Universidad Nacional de Misiones (UNaM, Argentina); and the Instituto Nacional de Tecnología Agropecuaria (INTA, Argentina). Also, we would like to thank the anonymous reviewers and editors, for their valuable comments and corrections to the manuscript. M. Lorena Castrillo and Gustavo A. Bich are CONICET assistant researchers. Humberto J. Debat is researcher at INTA. Patricia M. Aguilera and Mauro Grabiele are CONICET associate researchers.

References

Aguilera PM, Debat HJ & Grabiele M (2018) Dataset of the first transcriptome assembly of the tree crop "yerba mate" (*Ilex paraguariensis*) and systematic characterization of protein coding genes. Data in Brief 17: 1036-1040. DOI: 10.1016/j.dib.2018.02.015

Barbaro SE (2017) Estado de degradación de ultisoles bajo diferentes sistemas de control de malezas en cultivos de yerba mate en Misiones. Tesis de Maestría. Universidad Nacional de Buenos Aires, Buenos Aires. 83p.

Bari R & Jones JD (2009) Role of plant hormones in plant defence responses. Plant Molecular Biology 69: 473-488. DOI: 10.1007/s11103-008-9435-0

Baruah I, Baldodiya GM, Sahu J & Baruah G (2020) Dissecting the role of promoters of pathogen-sensitive genes in plant defense. Current Genomics 21: 491-503. DOI: 10.2174/1389202921999200727213500

Bejerman N, De Breuil S, Debat H, Miretti M, Badaracco A & Nome C (2017) Molecular characterization of yerba mate chlorosis-associated virus, a putative cytorhabdovirus infecting yerba mate (*Ilex paraguariensis*). Archives of Virology 162: 2481-2484. DOI: 10.1007/s00705-017-3363-8

Bejerman N, Acevedo RM, De Breuil S, Ruiz OA, Sansberro P, Dietzgen RG, Nome C & Debat H (2020) Molecular characterization of a novel cytorhabdovirus with a unique genomic organization infecting yerba mate (*Ilex paraguariensis*) in Argentina. Archives of Virology 165: 1475-1479. DOI: 10.1007/s00705-020-04609-3

Belingheri LD & Prat Kricun SD (1997) Programa de mejoramiento genético de la yerba mate en el INTA. *In*: Congresso Sul-Americano da Erva-Mate, II Reunião Técnica do Cone Sul Sobre a Cultura da Erva-Mate (org.) Anais. Curitiba. EMBRAPA-CNPF, Documentos, 33, Colombo. Pp. 267-278.

Bergottini VM, Hervé V, Sosa DA, Otegui MB, Zapata PD & Junier P (2017) Exploring the diversity of the root-associated microbiome of *Ilex paraguariensis* St. Hil. (Yerba Mate). Applied Soil Ecology 109: 23-31. DOI: 10.1016/j.apsoil.2016.09.013

Bolser D, Staines DM, Pritchard E & Kersey P (2016) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *In*: Edwards D (ed.) Plant bioinformatics. Methods in Molecular Biology. Vol. 1374. Humana Press, New York. Pp. 115-140. DOI: 10.1007/978-1-4939-3167-5_6

Burtnik OJ (2006) Yerba mate: manual de producción. 3rd ed. Ediciones INTA, Corrientes. 52p.

Debat HJ, Grabiele M, Aguilera PM, Bubillo R, Zapata PD, Marti DA & Ducasse DA (2014a) The complete genome of a putative endornavirus identified in yerba mate (*Ilex paraguariensis* St. Hil.). Virus Genes 49: 348-350. DOI: 10.1007/s11262-014-1096-2

Debat HJ, Grabiele M, Aguilera PM, Bubillo RE, Otegui MB, Ducasse DA, Zapata PD & Marti DA (2014b) Exploring the genes of yerba mate (*Ilex paraguariensis* A. St.-Hil.) by NGS and de novo transcriptome assembly. PLoS ONE 9: e109835. DOI: 10.1371/journal.pone.0109835

Dos Santos SCL (2009) Avaliação comparativa entre sistema agroflorestal com erva-mate e fragmento de floresta ombrófila mista quanto à fitossociologia, parametros quimicos e fisicos do solo e ciclagem de nutrientes. Tese de Mestrado. Universidade Federal do Paraná, Curitiba. 105p.

Fay JV, Watkins CJ, Shrestha RK, Litwiñiuk SL, Talavera Stefani LN, Rojas CA, Argüelles CF,

Ferreras JA, Caccamo M & Miretti MM (2018) Yerba mate (*Ilex paraguariensis* A. St.-Hil.) de novo transcriptome assembly based on tissue specific genomic expression profiles. BMC Genomics 19: 891. DOI: 10.1186/s12864-018-5240-6

Fan F, Yang X, Cheng Y, Kang Y & Chai X (2017) The DnaJ gene family in pepper (*Capsicum annuum* L.): comprehensive identification, characterization and expression profiles. Frontiers in Plant Science 8: 689. DOI: 10.3389/fpls.2017.00689

Feng K, Yu J, Cheng Y, Ruan M, Wang R, Ye Q, Zhou G, Li Z, Yao Z, Yang Y, Zheng Q & Wan H (2016) The SOD gene family in tomato: identification, phylogenetic relationships, and expression patterns. Frontiers in Plant Science 7: 1279. DOI: 10.3389/fpls.2016.01279

Finka A, Mattoo RUH & Goloubinoff P (2011) Meta-analysis of heat- and chemically upregulated chaperone genes in plant and human cells. Cell Stress and Chaperones 16: 15-31. DOI: 10.1007/s12192-010-0216-8

Fragkostefanakis S, Simm S, Paul P, Bublak D, Scharf K-D & Schleiff E (2015) Chaperone network composition in *Solanum lycopersicum* explored by transcriptome profiling and microarray meta-analysis. Plant, Cell and Environment 38: 693-709. DOI: 10.1111/pce.12426

Gao Z, Zhao R & Ruan J (2013) A genome-wide cis-regulatory element discovery method based on promoter sequences and gene co-expression networks. BMC Genomics 14(Suppl 1): S4. DOI: 10.1186/1471-2164-14-S1-S4

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. *In*: Walker JM (ed.) The proteomics protocols handbook. Humana Press, Totowa. Pp. 571-607.

Gortari F, Bulfe NM, Laczeski M, Onetto A, Cortese J, Castrillo ML, Bich GA, Zapata PD, Villalba LL & Niella F (2019) Tolerancia al estrés de plantines de yerba mate inoculados con bacterias promotoras de crecimiento vegetal nativas de Misiones. Yvyraretá 27: 6-10. Available at <http://www.yvyrareta.com.ar/index.php/ediciones-anteriores/diciembre-2019-n27>. Access on 01 February 2022.

Gortari F, Londero WO, Rocha P & Niella F (2020) Growth and physiological responses of yerba mate seedlings and mini-cuttings under drought stress. Cerne 26: 341-348.

Guyot R, Lefebvre-Pautigny F, Tranchant-Dubreuil C, Rigoreau M, Hamon P, Leroy T, Hamon S, Poncet V, Crouzillat D & De Kochko A (2012) Ancestral synteny shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum* Sp.) and rosid (*Vitis vinifera*) clades. BMC Genomics 13: 103. DOI: 10.1186/1471-2164-13-103

Hernandez-Garcia CM & Finer JJ (2014) Identification and validation of promoters and cis-acting regulatory elements. Plant Science 217-218: 109-119. DOI: 10.1016/j.plantsci.2013.12.007

Higo K, Ugawa Y, Iwamoto M & Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research 27: 297-300. DOI: 10.1093/nar/27.1.297

Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C & Rhee SY (2001) The Arabidopsis information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research 29: 102-105. DOI: 10.1093/nar/29.1.102

Huang B, Huang Z, Ma R, Chen J, Zhang Z & Yrjälä K (2021) Genome-wide identification and analysis of the heat shock transcription factor family in moso bamboo (*Phyllostachys edulis*). Scientific Reports 11: 16492. DOI: 10.1038/s41598-021-95899-3

Islam MdA, Rahman MdM, Rahman MdM, Jin X, Sun L, Zhao K, Wang S, Sikdar A, Noor H, Jeon J-S, Zhang W & Sun D (2021) In silico and transcription analysis of trehalose-6-phosphate phosphatase gene family of wheat: trehalose synthesis genes contribute to salinity, drought stress and leaf senescence. Genes 12: 1652. DOI: 10.3390/genes12111652

Jeffares DC, Penkett CJ & Bähler J (2008) Rapidly regulated genes are intron poor. Trends in Genetics 24: 375-378. DOI: 10.1016/j.tig.2008.05.006

Kaur A, Pati PK, Pati AM & Nagpal AK (2017) In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. PLoS ONE 12: e0184523. DOI: 10.1371/journal.pone.0184523

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J & Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. The Plant Journal 50: 347-363. DOI: 10.1111/j.1365-313X.2007.03052.x

Kumar GM, Mamidala P & Podile AR (2009) Regulation of polygalacturonase-inhibitory proteins in plants is highly dependent on stress and light responsive elements. Plant Omics 2: 238-249.

Kurtz V, Chifarelli D, Munaretto N & Korth S (2014) Poda de rebaje en plantaciones de yerba mate (*Ilex paraguariensis*): mortandad y productividad. *In*: Rago AM (ed.) Ciencia y tecnología de los cultivos industriales. Yerba mate, té mandica y stevia. Año 4, n°6. Ediciones INTA, Córdoba. Pp. 11-17.

Laczeski ME, Onetto AL, Cortese IJ, Mallozzi GY, Castrillo ML, Bich GA, Gortari F, Zapata PD & Otegui MB (2020) Isolation and selection of endophytic spore-forming bacteria with plant growth promoting properties isolated from *Ilex*

paraguariensis St. Hil. (yerba mate). Anais da Academia Brasileira de Ciências 92: e20181381. DOI: 10.1590/0001-3765202020181381

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, Rouzé P & Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30: 325-327. DOI: 10.1093/nar/30.1.325

Lis M & Walther D (2016) The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? BMC Genomics 17: 185. DOI: 10.1186/s12864-016-2549-x

Luo Y, Fang B, Wang W, Yang Y, Rao L & Zhang C (2019) Genome wide analysis of the rice J-protein family: identification, genomic organization, and expression profiles under multiple stresses. 3 Biotech 9: 358. DOI: 10.1007/s13205-019-1880-8

Marques ADC, Dos Reis MS & Denardin VF (2019) Yerba mate landscapes: forest use and socio-environmental conservation. Ambiente and Sociedade 22: e2822. DOI: 10.1590/1809-4422asoc201702822vu2019L3AO

MECON (2018) Informes de cadenas de valor. Yerba mate. Ministerio de Hacienda, Presidencia de la Nación, Argentina. Año 3, n°17. 28p. Available at <https://www.argentina.gob.ar/sites/default/files/sspmicro_cadenas_de_valor_yerba_mate.pdf>. Access on 12 January 2021.

Miernyk JA (2001) The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. Cell Stress and Chaperones 6: 209-218. DOI: 10.1379/1466-1268(2001)006<0209:tjdpoa>2.0.co;2

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD & Bateman A (2021) Pfam: the protein families database in 2021: nucleic acids research 49(D1): D412-D419. DOI: 10.1093/nar/gkaa913

Montagnini F, Eibl BI & Barth SR (2011) Organic yerba mate: an environmentally, socially and financially suitable agroforestry system. Bois et Forets des Tropiques 308: 59-74. DOI: 10.19182/bft2011.308.a20475

Mosa KA, Ismail A & Helmy M (2017) Plant stress tolerance. Springer, Cham. 83p. DOI: 10.1007/978-3-319-59379-1

Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K & Yamaguchi-Shinozaki K (2003) Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydratation and high-salinity stresses. The Plant Journal 34: 137-148. DOI: 10.1046/j.1365-313X.2003.01708.x

Nawkar GM, Kang CH, Maibam P, Park JH, Jung YJ, Chae HB, Chi YH, Jung IJ, Kim WY, Yun D-J & Lee SY (2017) HY5, a positive regulator of light signaling, negatively controls the unfolded protein response in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 114: 2084-2089. DOI: 10.1073/pnas.1609844114

Park C-J & Seo Y-S (2015) Heat shock proteins: a review of the molecular chaperones for plant immunity. The Plant Pathol Journal 31: 323-333. DOI: 10.5423/PPJ.RW.08.2015.0150

Petrillo E, Godoy Herz MA, Barta A, Kalyna M & Kornblihtt AR (2014) Let there be light: regulation of gene expression in plants. RNA Biology 11: 1215-1220. DOI: 10.4161/15476286.2014.972852

Pereira Croge C, Cuquel FL & Pintro PTM (2021) Yerba mate: cultivation systems, processing and chemical composition. A review. Scientia Agricola 78: e20190259. DOI: 10.1590/1678-992X-2019-0259

Pla M, Vilardell J, Guiltinan MJ, Marcotte WR, Niogret MF, Quatrano RS & Pagès M (1993) The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. Plant Molecular Biology 21: 259-266. DOI: 10.1007/BF00019942

Pulido P & Leister D (2018) Novel DNAJ-related proteins in *Arabidopsis thaliana*. New Phytologist 217: 480-490. DOI: 10.1111/nph.14827

Prat Kricun SD (2010) Evaluación y construcción estratégica del germoplasma de los cultivos de yerba mate y té. *In*: Mariotti JA (ed.) Cultivos industriales avances 2006/2009, Informe del programa nacional de cultivos industriales del INTA. Ediciones INTA, Buenos Aires. Pp. 69-71. ISBN 978-987-679-020-8

Rajan VBV & D'Silva P (2009) *Arabidopsis thaliana* J-class heat shock proteins: cellular stress sensors. Functional and Integrative Genomics 9: 433-446. DOI: 10.1007/s10142-009-0132-0

Rakocevic M, Medrado MJS, Lucambio F & Valduga AT (2008) Intensity of bitterness of processed yerba mate leaves originated in two contrasted light environments. Brazilian Archives of Biology and Technology 51: 569-579.

Ren X-Y, Vorst O, Fiers MWEJ, Stiekema WJ & Nap J-P (2006) In plants, highly expressed genes are the least compact. Trends in Genetics 22: 528-532. DOI: 10.1016/j.tig.2006.08.008

Roeber VM, Bajaj I, Rohde M, Schmülling T & Cortleven A (2021) Light acts as a stressor and influences abiotic and biotic stress responses in plants. Plant, Cell and Environment 44: 645-664. DOI: 10.1111/pce.13948

Rybak M, Rybak R, Cabrera MG & Alvarez RE (2014) Enfermedades de yerba mate y té en Misiones y norte de Corrientes. 2nd ed. Miscelánea N° 66. Ediciones INTA, Cerro Azul. 32p.

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K & Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain

of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochemical and Biophysical Research Communications 290: 998-1009. DOI: 10.1006/bbrc.2001.6299

Salas D, Valdez E & Fernández Y (2019) Aporte al conocimiento del comportamiento de la yerba mate (*Ilex paraguariensis* St. Hil.) bajo condiciones lumínicas naturales en la Reserva Natural del Bosque Mbaracayú. Revista de la Sociedad Científica del Paraguay 24: 74-98. DOI: 10.32480/rscp.2019-24-1.74-98

Sarkar NK, Thapar U, Kundnani P, Panwar P & Grover A (2013) Functional relevance of J-protein family of rice (*Oryza sativa*). Cell Stress and Chaperones 18: 321-331. DOI: 10.1007/s12192-012-0384-9

Seki H, Ichinose Y, Ito M, Shiraishi T & Yamada T (1997) Combined effects of multiple cis-acting elements in elicitor-mediated activation of PSCHS1 gene. Plant and Cell Physiology 38: 96-100. DOI: 10.1093/oxfordjournals.pcp.a029092

Shi Z, Maximova SN, Liu Y, Verica J & Guiltinan MJ (2010) Functional analysis of the *Theobroma cacao* NPR1 gene in *Arabidopsis*. BMC Plant Biology 10: 248. DOI: 1471-2229/10/248

Sosa DA, Barbaro S, Alvarenga FA, De Coll O, Ohashi DV, Rybak M & Agostini JP (2011) Yerba mate. Manual de campo. Ediciones INTA, Cerro Azul. 51p.

Stein J, Luna C, Espasandin F, Sartor M, Espinoza F, Ortiz JPA, Sansberro P & Pessino SC (2014) Construcción de un mapa genético preliminar de yerba mate (*Ilex paraguariensis*). Ciencias Agronómicas 23: 7-13.

Su T, Han M, Min J, Chen P, Mao Y, Huang Q, Tong Q, Liu Q & Fang Y (2018) Genome-wide survey of invertase encoding genes and functional characterization of an extracellular fungal pathogen-responsive invertase in *Glycine max*. International Journal of Molecular Sciences 19: 2395. DOI: 10.3390/ijms19082395

Su D, Xiang W, Wen L, Lu W, Shi Y, Liu Y & Li Z (2021) Genome-wide identification, characterization and expression analysis of BES1 gene family in tomato. BMC Plant Biology 21: 161. DOI: 10.1186/s12870-021-02933-7

Thierry-Mieg D & Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biology 7(Suppl 1): S12. DOI: 10.1186/gb-2006-7-s1-s12

Verma V, Ravindran P & Kumar PP (2016) Plant hormone-mediated regulation of stress responses. BMC Plant Biology 16: 86. DOI: 10.1186/s12870-016-0771-y

Verma AK, Tamadaddi C, Tak Y, Lal SS, Cole SJ, Hines JK & Sahi C (2019) The expanding world of plant J-domain proteins. Critical Reviews in Plant Sciences 38: 382-400. DOI: 10.1080/07352689.2019.1693716

Vierling E (1991) The role of heat shock proteins in plants. Annual Review of Plant Physiology and Plant Molecular Biology 42: 579-620. DOI: 10.1146/annurev.pp.42.060191.003051

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV & Provart NJ (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS ONE 2: e718. DOI: 10.1371/journal.pone.0000718

Zhang B, Qiu H-L, Qu D-H, Ruan Y & Chen D-H (2018) Phylogeny-dominant classification of J-proteins in *Arabidopsis thaliana* and *Brassica oleracea*. Genome 61: 405-415. DOI: 10.1139/gen-2017-0206

Zelada Cardozo NJ & Gonzalez Villalba JD (2019) Guía técnica cultivo de yerba mate. Facultad de Ciencias Agrarias, Universidad Nacional de Asunción, San Lorenzo. 60p.

Area Editor: Dra. Cassia Sakuragui
Received on April 10, 2022. Accepted on October 29, 2022.