Fungal communities represent the majority of root-specific transcripts in the transcriptomes of *Agave* plants grown in semiarid regions

Marina Pupke Marone [1], **Maria Fernanda Zaneli Campanari** [1], **Fabio Trigo Raya** [1], **Gonçalo Amarante Guimarães Pereira** [Corresp., 1, 2], **Marcelo Falsarella Carazzolle** [1, 2]

[1] Department of Genetics, Evolution, Microbiology, and Immunology, University of Campinas, Campinas, São Paulo, Brazil

[2] Center for Computing and Engineering Sciences, University of Campinas, Campinas, São Paulo, Brazil

Corresponding Author: Gonçalo Amarante Guimarães Pereira
Email address: goncalo@unicamp.br

Plants of the genus *Agave* present drought resistance mechanisms, commercial applications, and potential for bioenergy production. Currently, *Agave* species are used to produce alcoholic beverages and sisal fibers in semi-arid regions, mainly in Mexico and Brazil. Yet, because of their high productivities, low lignin content, and high shoot to root ratio, agaves are also interesting to be used as biomass feedstock to bioenergy production in marginal areas. Although several previous works have explored *Agave*'s morphological, physiological, and genetic adaptations to drought, very few have focused on plant-microbiome interactions that can be beneficial to the plant. Several studies have shown that plants host many microorganisms and researching them is important not only to understand the metabolism of the microorganisms interacting with the plant but also for biotechnological purposes. Here, we show the identification and functional characterization of fungal transcripts found in transcriptome datasets of three fiber-producing cultivars (*A. fourcroydes*, *Agave sisalana*, and hybrid 11648). We used leaf, stem, and root samples collected from the agave germplasm bank located in the state of Paraiba, in the Brazilian semiarid region, which has faced irregular precipitation periods. We used data from a *de novo* assembled transcriptome assembly (all tissues together). Regardless of the cultivar, we found around 10% of the transcripts belonging to fungi. Surprisingly, most root-specific transcripts were fungal (58%), not plant ones, and of these around 64% were identified as Ascomycota and 28% as Basidiomycota in the three communities. The functional characterization showed enriched terms related to heat shock proteins (HSPs) and transport across the membrane in Ascomycota and Basidiomycota, expressed in the communities of the three cultivars. Indeed, among the most expressed transcripts, many were annotated as HSPs, which are highly related to abiotic stress resistance. We compared our fungal transcripts datasets to other fungal genomes with many copies of

HSPs and found that most HSPs expressed by Ascomycota are small HSPs, which are not present in many copies in the other genomes and are highly related to dealing with temperature stresses. Also, some KEGG pathways suggest interaction with the roots, related to transport to outside the cell, such as *exosome* (present in the three Ascomycota communities) and *membrane trafficking*, which were further investigated. Also, we analyzed CAZymes, and among the secreted ones, we found chitinases, that can be related to pathogen control. We anticipate that our results can provide a starting point to the study of the potential uses of agaves' fungi. More advances in the field can pave the way for the exploitation of microbiomes as biotechnological tools to improve growth and production in dry environments, which is relevant especially in the climate change scenario.

Fungal communities represent the majority of root-specific transcripts in the transcriptomes of *Agave* plants grown in semiarid regions

Marina Pupke Marone*[1], Maria Fernanda Zaneli Campanari*[1], Fabio Trigo Raya[1], Gonçalo Amarante Guimarães Pereira[1,2], Marcelo Falsarella Carazzolle[1,2]

* Authors contributed equally to this manuscript.

[1]: Department of Genetics, Evolution, Microbiology, and Immunology, University of Campinas, Campinas, São Paulo, Brazil.
[2]: Center for Computing and Engineering Sciences, University of Campinas, Campinas, São Paulo, Brazil.

Corresponding Author:
Gonçalo Amarante Guimarães Pereira
Rua Carl Von Linaeus, Cidade Universitária, Campinas, São Paulo, 13083-864, Brazil
Email address: goncalo@unicamp.br

Abstract

Plants of the genus *Agave* present drought resistance mechanisms, commercial applications, and potential for bioenergy production. Currently, *Agave* species are used to produce alcoholic beverages and sisal fibers in semi-arid regions, mainly in Mexico and Brazil. Yet, because of their high productivities, low lignin content, and high shoot to root ratio, agaves are also interesting to be used as biomass feedstock to bioenergy production in marginal areas. Although several previous works have explored *Agave*'s morphological, physiological, and genetic adaptations to drought, very few have focused on plant-microbiome interactions that can be beneficial to the plant. Several studies have shown that plants host many microorganisms and researching them is important not only to understand the metabolism of the microorganisms interacting with the plant but also for biotechnological purposes. Here, we show the identification and functional characterization of fungal transcripts found in transcriptome datasets of three fiber-producing cultivars (*A. fourcroydes*, *Agave sisalana*, and hybrid 11648). We used leaf, stem, and root samples collected from the agave germplasm bank located in the state of Paraiba, in the Brazilian semiarid region, which has faced irregular precipitation periods. We used data from a *de novo* assembled transcriptome assembly (all tissues together). Regardless of the cultivar, we found around 10% of the transcripts belonging to fungi. Surprisingly, most root-specific transcripts were fungal (58%), not plant ones, and of these around 64% were identified as Ascomycota and 28% as Basidiomycota in the three communities. The functional characterization showed enriched terms related to heat shock proteins (HSPs) and transport across the membrane in Ascomycota and Basidiomycota, expressed in the communities of the three cultivars. Indeed, among the most expressed transcripts, many were annotated as HSPs, which are highly related to abiotic stress resistance. We compared our fungal transcripts datasets to other fungal genomes with many copies of HSPs and found that most HSPs expressed by Ascomycota are small HSPs, which are not present in many copies in the other genomes and are highly related to dealing with temperature stresses. Also, some KEGG pathways suggest interaction with the roots, related to transport to outside the cell, such as *exosome* (present in the three Ascomycota communities) and *membrane trafficking*, which were further investigated. Also, we analyzed CAZymes, and among the secreted ones, we found chitinases, that can be related to pathogen control. We anticipate that our results can provide a starting point to the study of the potential uses of agaves' fungi. More advances in the field can pave the way for the exploitation of microbiomes as biotechnological tools to improve growth and production in dry environments, which is relevant especially in the climate change scenario.

Introduction

Plants are host to many microorganisms and investigating them is important not only to understand the metabolism of the microorganisms interacting with the plant but also for biotechnological purposes. Of these sets of microorganisms, some will have an endophytic way of life, - colonizing the tissues and occupying the intra and intercellular spaces, in at least one period of their life cycle - or epiphytic - colonizing the vegetal surface, in the rhizoplane -, being able to assume a crucial role in development, growth, adaptability, and diversity of the plants

(Faust & Raes, 2012). ~~The relationship of~~ mutualism or pathogenicity ~~will~~ depend on biotic and abiotic factors, including species and genotypes of plants and microorganisms, environmental conditions, and the dynamic network of interactions within the plant biome (Hardoim *et al.*, 2015). ~~The~~ understanding of these microbial communities' structures and their interaction mechanisms with plants can ~~be useful for exploiting their capacity to contribute to various~~ phenotypes of interest, such as drought resistance and plant growth promoters, assisting in higher productivities (Wani *et al.*, 2015).

Agaves are semiarid plants with several drought resistance mechanisms, commercial uses, and the potential to be used as feedstock for bioenergy production in marginal areas (either ethanol or biogas). Their main drought resistance mechanism is the crassulacean acid metabolism (CAM), the most water-use efficient photosynthesis (Borland *et al.*, 2009), as well as retractable roots (Blunden *et al.*, 1973), waxy epidermis, and sunken stomata (Davis & Long, 2015). Currently, agaves are only used to produce sisal fibers - mainly in Brazil - and alcoholic beverages, such as tequila and mezcal - in Mexico. However, agaves are also interesting to be used as a biomass feedstock to bioenergy production in marginal areas because they present high productivities (Owen *et al.*, 2016), low lignin content (Yang *et al.*, 2015), and high shoot to root ratio (Borland *et al.*, 2009). *Agave*'s morphological and physiological adaptations to dry climates have been more studied than other aspects still under investigation, such as molecular mechanisms and genetics. In this context, more recently some works focused on the molecular aspects of agave's CAM metabolism and drought resistance mechanisms, mainly with the species *A. tequilana* Weber *var. azul* and *A. americana* (Gross *et al.*, 2013; Yang *et al.*, 2015; Abraham *et al.*, 2016). Considering fiber-producing agave cultivars, such as *A. sisalana* and hybrid 11648 ((*A. angustifolia* x *A. amaniensis*) x *A. amaniensis*), there are just a few recent studies available (Huang *et al.*, 2018, 2019; Sarwar *et al.*, 2019; Raya *et al.*, 2021), mostly approaching plant physiological mechanisms related to drought or cell wall biosynthesis. However, one aspect still to be vastly explored is the microbiome.

In this context, for agave plants, we believe tolerance to heat and drought stress is in part due to the microorganisms that inhabit them, and not only due to the physiological mechanisms of the plant itself. Thus, knowing these microorganisms and their possible impacts on plants becomes relevant. Some works have investigated the association between microbiota and agaves with a focus on understanding the populations' dynamics in different species and regions.

Inoculating *A. tequilana* with cultures of endophytic bacteria extracted from their leaf base lead to growth promotion in these plants (Martínez-Rodríguez *et al.*, 2014). Considering NGS technologies, a study investigated the interaction of microorganisms - both prokaryotic and eukaryotic - in distinct tissues of *A. tequilana*, *A. salmiana*, and *A. deserti* in different regions of Mexico and the USA (Coleman-Derr *et al.*, 2016). They have found a prokaryotic community common to the three species, even in different regions, which could assist in a drought resistance phenotype; the fungal community was otherwise modulated mainly by the characteristics of the regions. In Brazil, three papers explore the microbial community of *Agave sisalana,* two with a focus on nitrogen fixation (Santos *et al.*, 2019; Damasceno *et al.*, 2019) and the other on finding an antagonist to the etiological agent of the bole rot disease caused by *Aspergillus welwitschiae* (Candeias *et al.*, 2016), a cryptic species of *A. niger* (Duarte *et al.*, 2018). However, all Brazilian studies that conducted experiments with inoculant candidates used only classical microbiology methods to isolate endophytic fungi and microbial communities. Therefore, there are not many works describing agave microbiomes, especially considering fungi.

In our previous work (Raya *et al.*, 2021), we ~~have~~ assembled and analyzed the comprehensive transcriptomes (leaf, stem, and root tissues) of three fiber-producing agave cultivars collected at noon in a germplasm bank located in Monteiro, Paraíba, Brazil. Two of the cultivars are the most used in fiber production in Brazil - *A. sisalana* and hybrid 11648 (*A. amaniensis* x *A. angustifolia*) x *A. amaniensis*) - and one is the most used in Mexico - *A. fourcroydes*. ~~The~~ plants were in a region that faced irregular rainfall regime prior to sampling but looked apparently healthy. Interestingly, we have found that around 10% of the transcripts assembled were fungi transcripts, with the majority of them being root-specific. In this paper, we ~~describe our findings regarding these fungi communities, focusing on their identification~~ and functional characterization. Most fungal transcripts were classified as basidiomycete and ascomycete taxa and many of these transcripts have functions associated with transport and there are secreted carbohydrate-degrading enzymes (CAZymes), which may indicate they are interacting with their host plant. Interestingly, the most expressed transcripts are enriched with heat shock proteins that ~~can be~~ associated with resistance to heat and drought.

Methods

Transcriptome sequencing, assembly, and quantification

The samples of *A. fourcroydes, A. sisalana*, and hybrid 11648 were collected at the city of Monteiro, state of Paraíba, Brazil, at the agave germplasm bank owned by Embrapa (Brazilian Agricultural Research Corporation). The germplasm bank has non-calcic brown soil and prior to sampling the precipitation regime was highly irregular. We have harvested three different individuals for each cultivar at noon, collecting samples for root, stem, and leaf. Samples have not been disinfected, so both epiphytic and endophytic microorganisms might be present. The transcriptome for each agave cultivar was sequenced using Illumina/HiSeq 4000, the reads were assembled into transcripts using Trinity v. 2.5.1 (Haas *et al.*, 2013) and expression values were obtained with kallisto v. 0.44.0 (Bray *et al.*, 2016). Reads are available at SRA (accession number PRJNA746623). For each transcriptome, we have done the tissue specificity analysis using the tspex program (Camargo *et al.*, 2020) considering a threshold of SPM > 0.95 to identify root-specific transcripts. The complete pipeline for identifying plant transcripts can be seen in our previous work (Raya *et al.*, 2021). During the annotation of root-specific transcripts, we identified many fungal transcripts that led us to develop a new pipeline for their correct identification and further analysis. Fungal transcript and protein sequences for each cultivar are available in Supplemental Data S1.

In-house pipeline to separate plant from fungi

Kaiju software v. 1.6.3 (Menzel *et al.*, 2016) was used to infer the fungi taxonomic classification and to identify the plant and fungal sequences of our three assembled plant transcriptomes by comparison with the NCBI/NR database. However, manual annotation revealed that many plant transcripts were mistakenly classified as fungal sequences, so we decided to use another approach as well.

The second approach was based on the similarities and differences between the plant and fungi protein sequences available in the Uniref90 database. For this, we have done a BLASTx (E-value threshold of 1e-10) of the assembled transcriptome of *A. fourcroydes*, *A. sisalana*, and hybrid 11648 against Uniref90 and selected all the transcripts that presented at least 70% of the top 10 hits identified as fungi in Taxonomy DB.

Taxonomic classification of the fungi transcripts

Considering that different fungi groups can present distinct functional specializations, we decided to compare the two main phyla that we found in the previous Kaiju analysis: Ascomycota and Basidiomycota. We have developed a Perl pipeline to test different approaches to optimize this classification. Because there were many parameters to consider, being a multivariate set, we had to test the sensitivity of each parameter after fixating one of them. From 1,014 fungi genomes available on Ensembl Fungi (https://fungi.ensembl.org), we have randomly selected 2,000 CDS for Ascomycota and 700 for Basidiomycota to be used as a training and test dataset, in a proportion similar to what we had previously found in the datasets using Kaiju software. These 2,700 CDSs were blasted against all CDSs obtained from the 1,014 genomes (after the exclusion of these 2,700 CDS to avoid self-identification). The following parameters were tested to assess the ability to predict the two phyla of fungi: BLASTn (nucleotide similarity) or tBLASTx (amino acid similarity), E-value threshold of 1e-10 or 1e-20, the minimum number N of top BLAST hits (top N BLAST hits) necessary for the correct classification of Ascomycota and Basidiomycota. If a transcript could not be distinguished between Ascomycota and Basidiomycota based on our criteria, it would be classified as "Asco or Basidio". Moreover, if the transcript did not present similarity with Ascomycota nor Basidiomycota, but presented similarity with other fungi genomes, it was assigned to "Other Fungi". The metrics TPR (true positive rate, or sensitivity) and FPR (false positive rate, or specificity) were used (details in Methods S2).

Orthologous gene analysis

~~We have done an~~ orthologous analysis ~~using the~~ software OrthoFinder v. 2.5.2 (Emms & Kelly, 2019) configured with the parameter "-d" (nucleotide similarity) to compare the three communities' nucleotide sequences (*A. fourcroydes*, *A. sisalana* and hybrid 11648) ~~to understand how similar they are to each other~~. Then, we counted the number of orthologous nucleotide families ~~which are~~ exclusive (orphan genes) or shared between Ascomycota and Basidiomycota.

Annotation of fungi transcripts and enrichment tests

~~The fungi~~ transcripts from the three plants were annotated with RPS-Blast (E-value threshold of 0.01) using the CDD database (Lu *et al.*, 2020) as reference, with Pannzer2 (Törönen *et al.*, 2018) and to identify KEGG pathway (KO) groups (Kanehisa *et al.*, 2019) the

EggNOG-mapper program was used (Huerta-Cepas *et al.*, 2019). Using an R script, each fungal transcript was grouped considering the annotation of CDD and KO and divided between Ascomycota, Basidiomycota, Asco or Basidio, and Other Fungi groups. The significant KEGG terms and CDDs were detected by the hypergeometric test using all fungi transcripts as background, being accepted those with p-value < 0.05.

Transport proteins, CAZymes, and secreted proteins prediction

We have searched for membrane transport proteins using BLASTp with an E-value of 1e-5 against the curated Transporter Classification Database (Saier *et al.*, 2016), with an alignment coverage threshold of at least 70%. To identify carbohydrate-active enzymes (CAZymes), we have used the dbCAN software (Yin *et al.*, 2012). We have also used SignalP v. 5.0 (Almagro Armenteros *et al.*, 2019) to identify sequences of signal peptides in all fungal transcripts, proteins with scores > 0.5 were considered secreted.

Heat shock protein orthologous analysis

We have compared heat shock proteins (HSPs) between our fungi dataset (composed of Ascomycota and Basidiomycota sequences from *A. fourcroydes*, *A. sisalana*, and hybrid 11648) and all Ascomycota and Basidiomycota available genomes in Ensembl Fungi. To do so, we have searched for HSP domains using the manually curated database HSPIR (Ratheesh Kumar *et al.*, 2012), which has sequences from the six major groups of HSPs. To obtain all heat shock proteins in our fungi dataset and in the fungi genomes (about 1,000 genomes), we have done an HMM search using protein sequences with the program HMMER (Eddy, 2008), with a threshold of E-value < 0.001. We have selected the top 10 genomes for Ascomycota and the top 10 for Basidiomycota with more HSPs. Using only the HSPs sequences from the 26 datasets, we have done a protein orthologous analysis using OrthoFinder (Emms & Kelly, 2019)

Results

Root-specific fungi in Agave *plants*

Transcriptomic analysis of three fiber-producing agave cultivars (*A. fourcroydes*, *A. sisalana,* hybrid 11648) obtained from leaf, stem and root tissues revealed a large number of root-specific fungi transcripts (Fig. 1A). These plants were in a field without irrigation in a

semiarid region of Brazil that had irregular precipitation regimes prior to sampling, and we hypothesized that these fungi could be contributing to these plants' phenotype of high tolerance to heat and drought. Fungal transcripts constituted an average of 12.4% among the three agave transcriptomes, totalizing 2,966, 4,313, 3,433, fungi transcripts from *A. fourcroydes*, *A. sisalana*, and hybrid 11648, respectively (see Methods section for details). Interestingly, the amount of root-specific fungal transcripts is larger than root-specific plant transcripts in the three agave transcriptomes (average of 58% of root-specific transcripts are fungal). Although these transcripts represent a high percentage within the dataset, they do not present very high expression values, as these were calculated based on abundances of the plant transcripts (average of the three datasets top expressed plant transcript is 27,288 TPM); the highest fungal transcripts expression values were 197, 181 and 413 TPM for *A. fourcroydes*, *A. sisalana* and hybrid 11648, respectively (Table S3). The majority of fungi transcripts are root-specific, with only 6, 5, and 9 being non-root-specific (i.e., specificity measure, SPM < 0.95), for *A. fourcroydes*, *A. sisalana*, and hybrid 11648, respectively.

When comparing the three communities in an orthologous analysis (orthologous nucleotide), it is possible to notice that they are very similar to each other, only presenting few exclusive gene families (Fig. 1B; Table S4). As expected, due to the greater amount of Ascomycota transcripts, we have a higher proportion of Ascomycota-exclusive families (square brackets) compared to Basidiomycota (curled brackets), however the opposite occurs in hybrid 11648-exclusive families (5 and 18 for Ascomycota and Basidiomycota, respectively). Furthermore, we have found that more than half of all transcripts are orphans (1,484, 2,702, and 1,619 for *A. fourcroydes*, *A. sisalana*, and hybrid 11648, respectively), i.e, don't form orthologous clusters with any other gene, probably indicating that these fungi are using different metabolisms in each plant.

Taxonomic analysis

Taxonomy inference of the whole transcriptomes (Table 1; Table S2) classified 11.91% ± 1.49 (values are the mean for the three plant communities) as fungi transcripts, of which 67.7% ± 5.5 were Ascomycota and 30.3% ± 5.3 Basidiomycota. The most represented Ascomycota genera were Talaromyces (9.3% ± 1.5) and Corynespora (3.8% ± 0.2). For Basidiomycota, there

was resolution only to classify until order, of which the main were Agaricales (6.2% ± 2.5) and Auriculariales (2.7% ± 1.4).

Functional annotation of the fungi transcripts

To understand the main functions of the fungi transcripts expressed in *A. fourcroydes*, *A. sisalana,* hybrid 11648 and classified by the Ascomycota and Basidiomycota phyla, we performed the functional annotation and enrichment analysis of these transcripts using the databases of the Conserved protein domain (CDD) and KEGG pathway (KO). Figure 2 shows the frequency of statistically significant (p-value < 0.05) enriched KEGG pathways (Fig. 2A and 2C) and conserved domains (CDD) (Fig. 2B and 2D). Generally, Ascomycota presented more categories than Basidiomycota, which could be due to the presence of more transcripts of the first one.

The most frequent enriched KEGG pathway in both Ascomycota and Basidiomycota was related to *chaperones and folding catalysts*, absent only in the *A. fourcroydes* Basidiomycota community. Some pathways could suggest a type of interaction with the roots, related to transport to outside the cell, such as *exosome* (present in the three Ascomycota communities, but for Basidiomycota, it is exclusive in *A. sisalana*), *membrane trafficking*, and *transporters*. Other routes could be associated with root development and elongation, such as *Citrate Cycle (TCA)* and *Glyoxylate and dicarboxylate metabolism*, observed exclusively in the samples of *A. fourcroydes* for Basidiomycota. Still regarding metabolism, *Glycosylphosphatidylinositol (GPI)* is present in all three datasets, within the phylum of Ascomycota and *GTP-binding protein* is exclusive of Basidiomycota in *A. sisalana*.

Regarding CDD, there were also domains related to chaperones and heat shock proteins. The most frequent domain present in all communities was *ACD sHSPs-like* (CDD:107221), a subunit of small heat shock proteins, that plays an important role in stress protection, and it is found in prokaryotes and eukaryotes alike (Ganea, 2001). Some domains are enriched in just one community, such as DnaJ (CDD:199909) in hybrid 11648 Ascomycota, which is also a domain present in HSPs. For Basidiomycota, *HSP90* (CDD:333906) is exclusive in *A. fourcroydes* and *molecular chaperone DnaK* (CDD:234715) in hybrid 11648. Regarding central metabolism, in the Ascomycota communities of *A. sisalana* and hybrid 11648 we observed *mannitol dehydrogenase (MDH)-like* (CDD:187610), which is responsible for catalyzing the conversion of

fructose to mannitol. We have also found the domain *Fungal hexose transporter*, which is specific to the hybrid 11648 Basidiomycota community.

Identification of transport proteins and carbohydrate-degrading enzymes

~~To deeper investigate transporters, we have blasted our~~ fungi transcripts against the curated TCDB. We have found 145, 201, and 157 proteins related to transport with alignment coverage > 70% in *A. fourcroydes*, *A. sisalana*, and hybrid 11648 respectively (Table S5). The profile of most frequent families is similar between the three communities (Table 2), although for *A. sisalana* "The Major Facilitator Superfamily (MFS)" only appears in Ascomycetes and for hybrid 11648, "The H+- or Na+-translocating F-type, V-type and A-type ATPase (F-ATPase) Superfamily" is exclusive of Basidiomycetes. When looking at the most expressed transcripts annotated as transporters, we can notice two families related to heat shock proteins ("The HSP90/CDC37 (HSP90/CDC37) Family" and "The Cation Channel-forming Heat Shock Protein-70 (Hsp70) Family"). In tumor cells, it has been described that members of the Hsp70 and Hsp90 families can be found in association with membranes, along with co-chaperones, regulating functions related to folding and trafficking (Gross *et al.*, 2003; Heider *et al.*, 2021). Other family with many highly expressed transcripts in all communities is "The Endoplasmic Reticular Retrotranslocon (ER-RT)", which is related to transport to and from the endosplamic reticulum, mostly for degradation of misfolded proteins (Römisch, 2005). Overall, both Ascomycota and Basidiomycota present highly expressed transporters (Table 2), except for *A. sisalana*, with the majority being from Ascomycota. Also, many families are related to regular transportation inside the cell, such as ABC transporters, ATPases, and transport to and from the mitochondria, and most of these have overall low expressions. More interestingly, we found transporters of ammonium (Q8NKD5|1.A.11.3.3) in *A. fourcroydes* and *A. sisalana*, phosphate (K4HTY2|2.A.1.9.11) in *A. fourcroydes,* and inorganic phosphate (Q7RVX9|2.A.1.9.2) in *A. sisalana* and hybrid 11648 with expressions varying from 0.84 to 5.56 TPM.

Carbohydrate Active enZYmes (CAZymes) were identified among all fungal transcripts and compared between Ascomycota and Basidiomycota (Fig. 3A). The majority were classified as glycoside hydrolases (GH) and the profile between Ascomycetes and Basidiomycetes in each plant is different, although the pattern is similar when comparing the same phyla. Considering percentages, Ascomycota presented more GHs than Basidiomycota, but these have more

enzymes with auxiliary activities (AA) and polysaccharide lyases (PL) are exclusive to them. The majority of these CAZymes can be related to the fungi's own carbohydrate metabolism.

To check for CAZymes that might be related to fungal-plant interactions, we have compared the secreted CAZymes of the three fungal communities to the ones secreted by the plant roots, considering root-specific transcripts (Fig. 3B). It is possible to notice again the majority of GHs. For Ascomycota, only *A. sisalana* presented other types of CAZymes other than GHs. The plant profile does not present a lot of variation except for *A. sisalana*, which presents more GHs. Focusing only on the fungal secreted CAZymes (Table 3), we can notice that Basidiomycetes have more secreted CAZymes than Ascomycetes, and GH128 is exclusive of Basidiomycetes. Interestingly, most secreted CAZymes can be related to either plant cell wall degradation (GH10, GH11, GH16, GH17, GH43, AA9, and CE1) or fungal cell wall degradation (GH18 and GH128). Moreover, hybrid 11648 presented GH10 and GH11 exclusively, which are related to the degradation of hemicellulose.

Heat shock protein orthologous analysis

Chaperones and heat shock proteins (HSPs) were one of the categories and pathways enriched and most frequent in the functional analysis, so we decided to carry out an analysis comparing our fungi communities with other fungal genomes. Moreover, there are not many studies focusing on a broad comparison between fungal HSPs. In our data, HSPs represent 6.21, 4.92, and 5.71 (5.61% average) of the total fungal transcripts for *A. fourcroydes*, *A. sisalana*, hybrid 11648, respectively. The total amount of transcripts annotated as HSPs by the HMM search was 433 Ascomycota and 161 Basidiomycota, considering the total for the three communities. To compare these numbers to other fungi, we did a protein orthologous analysis with HSPs prospected in about 1,000 fungi genomes (Ensembl Fungi) and selected the top 10 genomes for Ascomycota and the top 10 for Basidiomycota with more HSPs. The list of these fungi can be found on Figure 4.

On the orthologous analysis, there were 161 HSP orthologous families. Among these, only 22 contained at least one family from our agave datasets (Fig. 4), of which 4 are exclusive (no orthologous in the other fungi's genomes), presenting a minimum coverage of 70% compared to the expected length of the HSP type. These exclusive families were all annotated as ACD (alpha-crystallin domain), a domain present in small HSPs. There were 120 orphan genes

considering all datasets, of which 6 are from agave communities but were not considered in the analysis because their length was much shorter than expected for their families. Also, only 8 families are common to all the analyzed datasets. The family with more proteins (419), considering all datasets, was annotated as DnaJ (OG000000), a type of co-chaperone that acts helping the folding performed by Hsp70 (Genest *et al.*, 2019), although it is not abundant in our agave datasets, especially in Basidiomycota. The family with more proteins from our datasets (OG0000001), which are more abundant in Ascomycota (a mean of 50 transcripts in the three sets), was annotated as ACD.

Considering all fungi species analyzed, the types of HSP with more protein orthologous families were Hsp100, Hsp70, and DnaJ with 78, 75, and 60 families, respectively. The type of HSP with fewer families was Hsp90 (only 8 families). Curiously, some large Hsp100 families are exclusive (or almost) of some genomes and do not contain any representant in our fungi transcripts. For instance, *Amanita muscaria* has a family with 28 proteins (and 1 from *Piloderma croceum*, OG0000025), *Galerina. marginata* has one with 12 (and 1 from *A. muscaria*, OG0000064), another with 15 proteins (OG0000056) and *Exidia glandulosa* has one with 8 proteins (OG0000072). Another interesting family that did not contain any representant of our fungi transcripts was OG0000005, annotated as Hsp70; with only 1 protein in an Ascomycota genome (*Fusarium oxysporum*), there were 32 proteins in *Serendipita vermifera*, 24 in *A. muscaria*, 17 in *G. marginate*, and other Basidiomycota with less than 10.

The number of transcripts in each family varied little between the six analyzed communities, showing the similarity between the expressed major HSP groups. However, regarding expression values, these numbers varied a lot inside the families, with only a few transcripts with expression above 10 TPM and the vast majority (83,9%) with values between 0.77 and 9.87 TPM. The families with more highly expressed transcripts are OG0000001, OG0000002, OG0000008, OG0000030, and OG0000114 (Table S6).

Discussion

In this study, we have investigated and characterized the molecular functions of transcripts of fungal communities found on the transcriptomes of three agave cultivars, namely *A. fourcroydes*, *A. sisalana*, and hybrid 11648. The percentage of fungal transcripts in our plant transcriptome dataset was a surprise. Even though we suspect it is possibly common to find

microorganisms in such experiments, we believe that they are presumably ignored or treated as contaminants in most cases. In fact, a similar study proposed the same type of investigation from stem samples in *Eucalyptus grandis*, finding around 21% of transcripts that did not map to the plant genome (Messal *et al.*, 2019). Still, it is remarkable that there are more fungal root-specific transcripts than plant ones in the analyzed agave cultivars. These fungi could be epiphytic or endophytic. In the latter case, they could also be localized in the velamen region of roots. Some studies have found fungi in the root tissues of CAM plants, especially in the velamen regions of orchid roots (Deepthi & Ray, 2018). This region contains a large community of fungal associates (due to the unique morphological characteristics of velamen roots) that are essential for the survival of orchids in specific natural habitats. Agave's roots also contain velamen regions, which in *A. sisalana* are composed of four layers of cells that have an irregular shape (Neto & Martins, 2012). According to these authors, absorbent hairs are found in groups on the roots of *A. sisalana*, where associations with fungal hyphae may occur.

Anyhow, fungal transcripts presented very low expression comparing to the plant ones, which was expected, since the primary aim was to sequence the plant. We understand that our datasets are not complete metatranscriptomes, but the number of assembled transcripts is sufficient to perform some exploratory analysis to shed light on fungal communities in dry environments. Studies have showed that microorganisms associated with agave plants can be beneficial in dealing with diseases or increasing productivity. Agave's endophytic bacteria have been used to control anthracnose in banana (Damasceno *et al.*, 2019) and the bole rot disease in *A. sisalana* (De Souza *et al.*, 2021). Moreover, association with endophytic fungi have increased growth of *Agave victoria-reginae* (Obledo *et al.*, 2003). However, to our knowledge, there are no studies describing such a huge number of fungi in plant root transcriptomes, nor many studies focusing solely on fungal microbiota in agaves' roots, neither studies on fungal metatranscriptomes in agave. Nevertheless, finding a large number of fungal transcripts in the roots might be more common in experiments carried out in the field, as some experiments showed that wild cultivars present a more rich microbiota than cultivated ones (Xu *et al.*, 2019; Tian *et al.*, 2020), although more research is needed as this can largely vary between environments and species.

The three fungal communities we analyzed are very similar to each other, probably because all the plants are on the same field and climate conditions. This is in agreement with

previous findings that fungal communities depend vastly on the soil (Lee & Hawkes, 2020) and, especially in agaves, geography matters more in fungi than in prokaryote communities (Coleman-Derr *et al.*, 2016). The classified fungi mainly belong to the phyla Ascomycota and Basidiomycota and are present in similar proportions and numbers across the plants, most transcripts belonging to Ascomycota compared to Basidiomycota. As our main interest is finding whether the fungi are improving or at least interacting with the plants, we focused on finding molecular functions that could be related to such interactions. To do so, we investigated enriched KEGG pathways and CDD domains. One of the most frequent enriched KEGG pathways for all Ascomycota and *A. sisalana* Basidiomycota communities was the exosome, which plays roles in cell communication and nutrients delivery. If we consider the plants, the exosome is involved with long-distance communication factors and bioactive compounds and acts on the immunity system when in a stressful situation (Akuma *et al.*, 2019). Similarly, the membrane trafficking pathway was also enriched in the same groups. Associated with signaling pathways, transport of small molecules, and metabolic processes (Geisler *et al.*, 2013), it could be related to plant-microbiome interactions. These interactions can be pathogenic or mutualistic, both extremely modulated by the plant itself or by interrelating microbiomes (Inada & Ueda, 2014). All these enriched pathways related to transport are more frequent in Ascomycota than Basidiomycota and looking to the numbers of transcripts annotated as transporters (Table S5), Ascomycetes indeed present more transporters.

The analysis of secreted CAZymes also provides a hint of possible interactions between the fungi and the host plants, as endophytes must break the plant cell wall (PCW) to colonize the host. The plant cell wall is mainly formed by cellulose, hemicellulose, lignin, and pectin, which form the first barrier against pathogens and other abiotic stresses (Benoit *et al.*, 2015). Both Ascomycetes and Basidiomycetes in all three communities presented CAZymes related to degradation of PCWs, such as GH16, GH17, GH43, and AA9 suggesting that they could be colonizing the plant roots. Most of them were annotated as glycoside hydrolases, which is the family mainly known to contain cellulolytic and hemicellulolytic enzymes. On the other hand, they also presented chitinases (GH18), related to the degradation of fungal cell walls, which could be a defense mechanism against pathogens (Aranda-Martinez *et al.*, 2016; Yang *et al.*, 2019) or even to remodel their cell wall (Gruber & Seidl-Seiboth, 2012). Although there is a lot

of variety in the distribution and quantity of CAZymes in fungal genomes, it has been demonstrated that symbiotic fungi can have fewer CAZymes (Zhao *et al.*, 2013).

Considering sugar metabolism, plant root exudates can contribute to the maintenance of the microbial community. The sugar transport system is important for all individuals, but especially in plants, it acts in sugar distribution, plant development, cell to cell communication, the constitution of signaling molecules, and environmental adaptation (Yamada *et al.*, 2011). These sugars are used in root exudates, essential for the rhizosphere microbiome (Kiers *et al.*, 2011). Thus, transport systems can help the roots to regulate carbohydrates in the rhizosphere for the maintenance and growth of new microorganisms. Knowing this, we have found enriched pathways and domains related to sugar metabolism in the fungi communities, that could be involved in this kind of interaction along with the transporters. Studies indicate that plants may detect, discriminate, and reward the best microorganisms by looking more closely at the fungi with more carbohydrates (Doidy *et al.*, 2012). The fungi, on the other hand, may respond to this cooperation by increasing nutrient transfer to the roots to gain more carbohydrates. Another important point is that the limitation of hexose may contribute to a decrease in the presence of mycorrhizae (Kiers *et al.*, 2011), which might explain why we have fewer Basidiomycetes' transcripts.

Bearing in mind that agaves are plants adapted to dry environments and that the ones analyzed in this study were going through a period of irregular precipitations, it was expected to find expressed transcripts related to abiotic resistance also among the fungal ones. We found enriched protein domains and KEGG pathways related to chaperones and heat shock proteins, which are also among the top expressed transcripts and among the most frequent transporter families. HSPs have many different functions, many related to protein folding in some way. Particularly in fungi, they act in stress resistance, sporulation, sexual/asexual development, virulence (Bui *et al.*, 2016; Chatterjee & Tatu, 2017), and some are important drug targets in fungal caused diseases (Lamoth *et al.*, 2016). One of their main functions, however, as the name suggests, is related to temperature stress. Our samples were collected at noon in a very dry region, and it has been described that in such situations, the soil surface can reach over 40ºC (Nobel, 2010). Therefore, it seems reasonable that we have found many HSPs as top expressed transcripts and within enriched categories. To further investigate the different HSPs found in our datasets and to compare them to other fungi, we did an HSP orthologous analysis.

The main type of HSP found in our agave fungal transcript datasets was small HSPs, annotated as ACD, which is a conserved domain through evolution, although the whole sequence of small HSPs varies (Kriehuber *et al.*, 2010). Small HSPs can act in response to temperature stress; some of them have been described as conferring tolerance to freezing (Pacheco *et al.*, 2009) and heat shock (Haslbeck *et al.*, 1999) in *Saccharomyces cerevisiae*. Also, small HSPs have been described as the first defense against many stresses (Haslbeck & Vierling, 2015). In filamentous fungi, the number of copies of small HSPs in their genomes does not vary much (3-5 copies each) and it has been shown that they diverge a lot across fungal species (Wu *et al.*, 2016). This could explain the difference between the number of proteins clustered in the other fungal genomes analyzed and our fungal transcripts (OG0000001), as maybe our samples presented closely related species grouped. However, we have more Ascomycota ACD-annotated proteins than Basidiomycota, whereas the Basidiomycota fungal genomes present more copies of ACDs, suggesting that even if the Basidiomycetes present in our community also have many copies in their genomes, they are not expressing them.

To our knowledge, there are no works comparing HSPs in such different fungal genomes. In this regard, one important protein family for the fungal genomes, but not so present in our fungi transcripts was DnaJ, present in a high number of copies in both Ascomycota and Basidiomycota genomes. In the enrichment analysis, hybrid 11648 Ascomycota exclusively had the DnaJ domain enriched. The DnaJ proteins are known molecules that confer stress protection, playing key roles in the cell death cycle and resistance to diseases (Liu & Whitham, 2013). Indeed, DnaJ was the third type of HSP more abundant in the whole dataset, with 60 families. The first one with more proteins through all the fungi genomes is Hsp100 with 78 families and Hsp70 is second with 75. In fact, Hsp70 and Hsp100 both form a complex that acts in the disaggregation of other proteins, which could explain the similar number of transcripts belonging to these families in our agave datasets (OG0000002 and OG0000003). All HSPs were more uniformly distributed through Ascomycota genomes, but Hsp100 presented many exclusive or almost exclusive families in many Basidiomycota genomes. Thus, Hsp100 are seemingly important chaperones for some Basidiomycetes in this environment.

Conclusions

In this study, we have identified and characterized the fungal transcripts found in the transcriptomes of three agave cultivars grown in the Brazilian semiarid without irrigation and under an irregular precipitation regime. We have found more fungal root-specific transcripts than plant ones. These fungi belong to two different phyla which are performing somewhat distinct functions, many related to interactions with the host plant and others related to drought resistance. Microbial communities can contribute to increase the host plant's resistance to many biotic and abiotic factors. Therefore, the current study underlies the importance of analyzing possible "contaminants" that may appear in transcriptome datasets, as valuable information might be present. In summary, our exploratory analysis of fungal communities indicates potential microorganisms that could be exploited to generate improved agronomical characteristics in agave or other cultures, trying to mitigate the damages of increasing temperatures on crops.

References

Abraham PE, Yin H, Borland AM, Weighill D, Lim SD, De Paoli HC, Engle N, Jones PC, Agh R, Weston DJ, *et al.* 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nature Plants* **2**: 16178.

Akuma P, Okagu OD, Udenigwe CC. 2019. Naturally occurring exosome vesicles as potential delivery vehicle for bioactive compounds. *Frontiers in Sustainable Food Systems* **3**: 23.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**: 420–423.

Aranda-Martinez A, Lenfant N, Escudero N, Zavala-Gonzalez EA, Henrissat B, Lopez-Llorca L V. 2016. CAZyme content of Pochonia chlamydosporia reflects that chitin and chitosan modification are involved in nematode parasitism. *Environmental Microbiology* **18**: 4200–4215.

Benoit I, Culleton H, Zhou M, DiFalco M, Aguilar-Osorio G, Battaglia E, Bouzid O, Brouwer CPJM, El-Bushari HBO, Coutinho PM, *et al.* 2015. Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. *Biotechnology for Biofuels* **8**: 107.

Blunden G, Yi Y, Jewers K. 1973. The comparative leaf anatomy of Agave, Beschorneria, Doryanthes and Furcraea species (Agavaceae: Agaveae). *Botanical Journal of the Linnean Society* **66**: 157–179.

Borland AM, Griffiths H, Hartwell J, Smith JAC. 2009a. Exploiting the potential of plants

with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany* **60**: 2879–2896.

Borland AM, Griffiths H, Hartwell J, Smith JAC. **2009b**. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany* **60**: 2879–2896.

Bray NL, Pimentel H, Melsted P, Pachter L. **2016**. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527.

Bui D-C, Lee Y, Lim JY, Fu M, Kim J-C, Choi GJ, Son H, Lee Y-W. **2016**. Heat shock protein 90 is required for sexual and asexual development, virulence, and heat shock response in Fusarium graminearum. *Scientific reports* **6**: 28154.

Camargo A, Vasconcelos A, Fiamenghi M, Pereira G, Carazzolle M. **2020**. tspex: a tissue-specificity calculator for gene expression data.

Candeias E, Santos M, Duarte E, Oliveira T, Bezerra J, Soares A. **2016**. FUNGOS ENDOFÍTICOS DE RAÍZES DE SISAL ANTAGONISTAS AO Aspergillus niger. *Agrotrópica* **28**: 29–36.

Chatterjee S, Tatu U. **2017**. Heat shock protein 90 localizes to the surface and augments virulence factors of Cryptococcus neoformans. *PLoS neglected tropical diseases* **11**: e0005836.

Coleman-Derr D, Desgarennes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, North G, Visel A, Partida-Martinez LP, Tringe SG. **2016**. Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytologist* **209**: 798–811.

Damasceno CL, Duarte EAA, dos Santos LBPR, de Oliveira TAS, de Jesus FN, de Oliveira LM, Góes-Neto A, Soares ACF. **2019**. Postharvest biocontrol of anthracnose in bananas by endophytic and soil rhizosphere bacteria associated with sisal (Agave sisalana) in Brazil. *Biological Control* **137**: 104016.

Davis SC, Long SP. **2015**. Sisal/Agave. In: Cruz VM V., Dierig DA, eds. Handbook of Plant Breeding. New York, NY: Springer New York, 335–349.

Deepthi AS, Ray JG. **2018**. Endophytic diversity of hanging velamen roots in the epiphytic orchid Acampe praemorsa. *Plant Ecology & Diversity* **11**: 649–661.

Doidy J, Grace E, Kühn C, Simon-Plas F, Casieri L, Wipf D. **2012**. Sugar transporters in plants and in their interactions with fungi. *Trends in plant science* **17**: 413–422.

Duarte EAA, Damasceno CL, de Oliveira TAS, Barbosa L de O, Martins FM, de Silva JRQ, de Lima TEF, da Silva RM, Kato RB, Bortolini DE, *et al.* 2018. Putting the mess in order: Aspergillus welwitschiae (and not A. niger) is the etiological agent of sisal bole rot disease in Brazil. *Frontiers in Microbiology* **9**.

Eddy SR. **2008**. A Probabilistic Model of Local Sequence Alignment That Simplifies Statistical Significance Estimation. *PLOS Computational Biology* **4**: e1000069.

Emms DM, Kelly S. **2019**. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Faust K, Raes J. **2012**. Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**: 538–550.

Ganea E. **2001**. Chaperone-like activity of alpha-crystallin and other small heat shock proteins. *Current Protein and Peptide Science* **2**: 205–225.

Geisler M, Murphy AS, Sze H. **2013**. Evolution of membrane signaling and trafficking in plants. *Frontiers in Plant Science*.

Genest O, Wickner S, Doyle SM. **2019**. Hsp90 and Hsp70 chaperones: Collaborators in protein remodeling. *Journal of Biological Chemistry* **294**: 2109–2120.

Gross C, Koelch W, DeMaio A, Arispe N, Multhoff G. **2003**. Cell surface-bound heat shock protein 70 (Hsp70) mediates perforin-independent apoptosis by specific binding and uptake of granzyme B. *The Journal of biological chemistry* **278**: 41173–41181.

Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A. **2013**. De novo transcriptome assembly of drought tolerant CAM plants, Agave deserti and Agave tequilana. *BMC Genomics* **14**: 1–14.

Gruber S, Seidl-Seiboth V. **2012**. Self versus non-self: fungal cell wall degradation in Trichoderma. *Microbiology (Reading, England)* **158**: 26–34.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, *et al.* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**: 1494–1512.

Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, Döring M, Sessitsch A. **2015**. The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and molecular biology reviews : MMBR* **79**: 293–320.

Haslbeck M, Vierling E. **2015**. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *Journal of Molecular Biology* **427**: 1537–1548.

Haslbeck M, Walke S, Stromer T, Ehrnsperger M, White HE, Chen S, Saibil HR, Buchner J. **1999**. Hsp26: a temperature-regulated chaperone. *The EMBO Journal* **18**: 6744–6751.

Heider M, Eichner R, Stroh J, Morath V, Kuisl A, Zecha J, Lawatscheck J, Baek K, Garz A-K, Rudelius M, *et al.* 2021. The IMiD target CRBN determines HSP90 activity toward transmembrane proteins essential in multiple myeloma. *Molecular cell* **81**: 1170-1186.e10.

Huang X, Wang B, Xi J, Zhang Y, He C, Zheng J, Gao J, Chen H, Zhang S, Wu W, *et al.* 2018. Transcriptome Comparison Reveals Distinct Selection Patterns in Domesticated and Wild Agave Species, the Important CAM Plants. *International Journal of Genomics* **2018**: 1–12.

Huang X, Xiao M, Xi J, He C, Zheng J, Chen H, Gao J, Zhang S, Wu W, Liang Y, *et al.* 2019. De Novo Transcriptome Assembly of Agave H11648 by Illumina Sequencing and Identification of Cellulose Synthase Genes in Agave Species. *Genes* **10**: 103.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, *et al.* 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**: D309–D314.

Inada N, Ueda T. **2014**. Membrane trafficking pathways and their roles in plant--microbe interactions. *Plant and Cell Physiology* **55**: 672–686.

Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. **2019**. New approach for understanding genome variations in KEGG. *Nucleic acids research* **47**: D590–D595.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, *et al.* 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *science* **333**: 880–882.

Kriehuber T, Rattei T, Weinmaier T, Bepperling A, Haslbeck M, Buchner J. **2010**. Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *The FASEB Journal* **24**: 3633–3642.

Lamoth F, Juvvadi PR, Steinbach WJ. **2016**. Heat shock protein 90 (Hsp90): A novel antifungal target against Aspergillus fumigatus. *Critical Reviews in Microbiology* **42**: 310–321.

Lee MR, Hawkes C V. **2020**. Plant and Soil Drivers of Whole-Plant Microbiomes: Variation in Switchgrass Fungi from Coastal to Mountain Sites. *Phytobiomes Journal* **5**: 69–79.

Liu J-Z, Whitham SA. **2013**. Overexpression of a soybean nuclear localized type--III DnaJ domain-containing HSP40 reveals its roles in cell death and disease resistance. *The plant journal* **74**: 110–121.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, *et al.* 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic acids research* **48**: D265–D268.

Martínez-Rodríguez J del C, Mora-Amutio MD la, Plascencia-Correa LA, Audelo-Regalado E, Guardado FR, Hernández-Sánchez E, Peña-Ramírez YJ, Escalante A, Beltrán-García MJ, Ogura T. **2014**. Cultivable endophytic bacteria from leaf bases of Agave tequilana and their role as plant growth promoters . *Brazilian Journal of Microbiology* **45**: 1333–1339.

Menzel P, Ng KL, Krogh A. **2016**. Fast and sensitive taxonomic classification for

metagenomics with Kaiju. *Nature Communications* **7**: 11257.

Messal M, Slippers B, Naidoo S, Bezuidt O, Kemler M. **2019**. Active fungal communities in asymptomatic eucalyptus grandis stems differ between a susceptible and resistant clone. *Microorganisms* **7**: 1–16.

Neto I, Martins F. **2012**. Anatomy of the vegetative organs of Agave sisalana perrine ex engelm (Agavaceae). *Revista Caatinga* **25**: 72–78.

Nobel PS. **2010**. *Desert Wisdom/Agaves and Cacti: CO2, Water, Climate Change*. New York, NY: NY: iUniverse.

Obledo EN, Barragán-Barragán LB, Gutiérrez-González P, Ramírez-Hernández BC, Ramírez JJ, Rodríguez-Garay B. **2003**. Increased photosyntethic efficiency generated by fungal symbiosis in Agave victoria-reginae. *Plant Cell, Tissue and Organ Culture* **74**: 237–241.

Owen NA, Fahy KF, Griffiths H. **2016**. Crassulacean acid metabolism (CAM) offers sustainable bioenergy production and resilience to climate change. *GCB Bioenergy* **8**: 737–749.

Pacheco A, Pereira C, Almeida MJ, Sousa MJ. **2009**. Small heat-shock protein Hsp12 contributes to yeast tolerance to freezing stress. *Microbiology* **155**: 2021–2028.

Ratheesh Kumar R, Nagarajan NS, Arunraj SP, Sinha D, Veedin Rajan VB, Esthaki VK, D'Silva P. **2012**. HSPIR: A manually annotated heat shock protein information resource. *Bioinformatics* **28**: 2853–2855.

Raya FT, Marone MP, Carvalho LM, Rabelo SC, de Paula MS, Campanari MFZ, Freschi L, Mayer JLS, Silva ORRF, Mieczkowski P, *et al.* 2021. Extreme physiology: Biomass and transcriptional profiling of three abandoned Agave cultivars. *Industrial Crops and Products* **172**: 114043.

Römisch K. **2005**. Endoplasmic reticulum-associated degradation. *Annual review of cell and developmental biology* **21**: 435–456.

Saier MHJ, Reddy VS, Tsu B V, Ahmed MS, Li C, Moreno-Hagelsieb G. **2016**. The Transporter Classification Database (TCDB): recent advances. *Nucleic acids research* **44**: D372-9.

Santos AF de J, Moreira ZPM, Souza JT de, Oliveira LM de, Barbosa HR, Silva E de S, Silva RM da, Soares ACF. **2019**. Culturable diazotrophic bacterial community associated with Agave sisalana P. plants from semi-arid regions in Brazil. *Revista Brasileira de Ciências Agrárias* **14**.

Sarwar MB, Ahmad Z, Rashid B, Hassan S, Gregersen PL, Leyva MD la O, Nagy I, Asp T, Husnain T. **2019**. De novo assembly of Agave sisalana transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Scientific Reports* **9**: 396.

De Souza JT, Silva ACM, de Jesus Santos AF, Santos PO, Alves PS, Cruz-Magalhães V,

Marbach PAS, Loguercio LL. **2021**. Endophytic bacteria isolated from both healthy and diseased Agave sisalana plants are able to control the bole rot disease. *Biological Control* **157**.

Tian L, Shi S, Sun Y, Tran L-SP, Tian C. **2020**. The compositions of rhizosphere microbiomes of wild and cultivated soybeans changed following the hybridization of their F1 and F2 generations. *European Journal of Soil Biology* **101**: 103249.

Törönen P, Medlar A, Holm L. **2018**. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research* **46**: W84–W88.

Wani ZA, Ashraf N, Mohiuddin T, Riyaz-Ul-Hassan S. **2015**. Plant-endophyte symbiosis, an ecological perspective. *Applied Microbiology and Biotechnology* **99**: 2955–2965.

Wu J, Wang M, Zhou L, Yu D. **2016**. Small heat shock proteins, phylogeny in filamentous fungi and expression analyses in Aspergillus nidulans. *Gene* **575**: 675–679.

Xu S, Tian L, Chang C, Li X, Tian C. **2019**. Cultivated rice rhizomicrobiome is more sensitive to environmental shifts than that of wild rice in natural environments. *Applied Soil Ecology* **140**: 68–77.

Yamada K, Kanai M, Osakabe Y, Ohiraki H, Shinozaki K, Yamaguchi-Shinozaki K. **2011**. Monosaccharide absorption activity of Arabidopsis roots depends on expression profiles of transporter genes under high salinity conditions. *Journal of Biological Chemistry* **286**: 43577–43586.

Yang Y, Liu X, Cai J, Chen Y, Li B, Guo Z, Huang G. **2019**. Genomic characteristics and comparative genomics analysis of the endophytic fungus Sarocladium brachiariae. *BMC Genomics* **20**: 782.

Yang L, Lu M, Carl S, Mayer JA, Cushman JC, Tian E, Lin H. **2015**. Biomass characterization of Agave and Opuntia as potential biofuel feedstocks. *Biomass and Bioenergy* **76**: 43–53.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. **2012**. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**: W445-51.

Zhao Z, Liu H, Wang C, Xu J-R. **2013**. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**: 274.

Figure 1

General numbers of fungal and plant transcripts in each plant and summary of the orthologous analysis of the three fungal communities.

A: The plant transcriptomes were assembled *de novo*, as described by Raya *et al*. (2021). Among the assembled transcripts we have found fungal transcripts, which are almost exclusively expressed in the plants' roots. Indeed, most root-specific transcripts were fungal ones. B: Number of orthologous gene families in each fungal community. The numbers in square brackets are only Ascomycota annotated families and in curled brackets are Basidiomycota annotated families. AF: *Agave fourcroydes*; AS: *Agave sisalana*; HY: hybrid 11648. Photos: Fabio T. Raya.

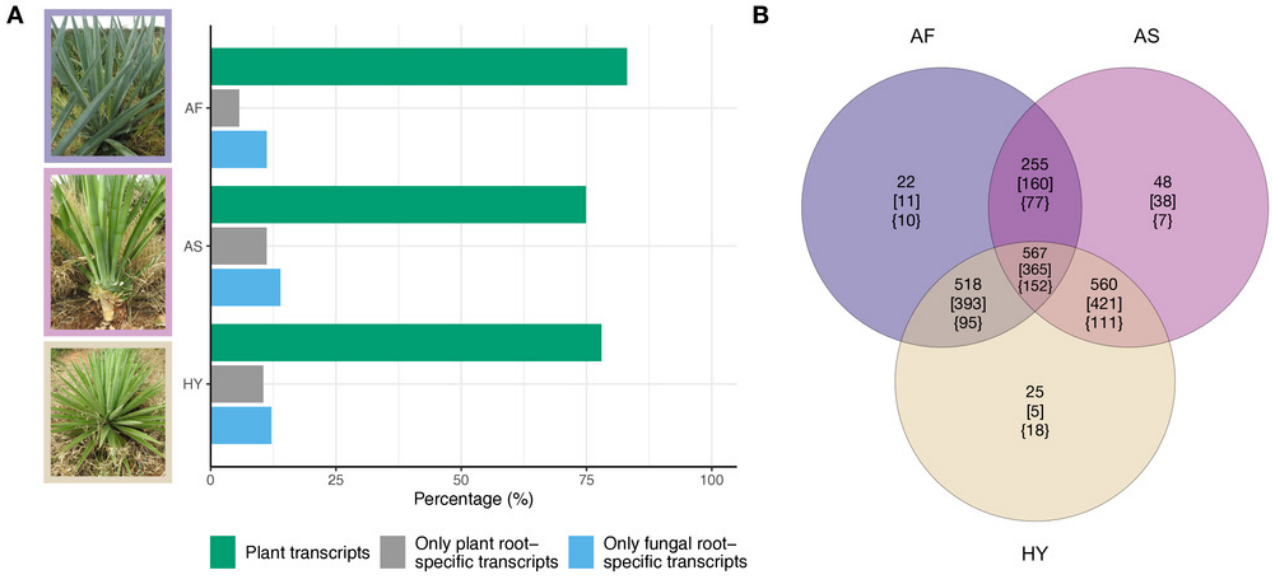

Figure 2

Functional characterization of KEGG pathways and protein domains

A, C: Frequency of enriched KEGG pathways for (A) Ascomycota and (C) Basidiomycota. B, D: Frequency of conserved protein domains (CDD) for (B) Ascomycota and (D) Basidiomycota. The hypergeometric test was used with p-value < 0.05 and only significantly enriched terms are shown. AF: *Agave fourcroydes*; AS: *Agave sisalana*; HY: hybrid 11648.

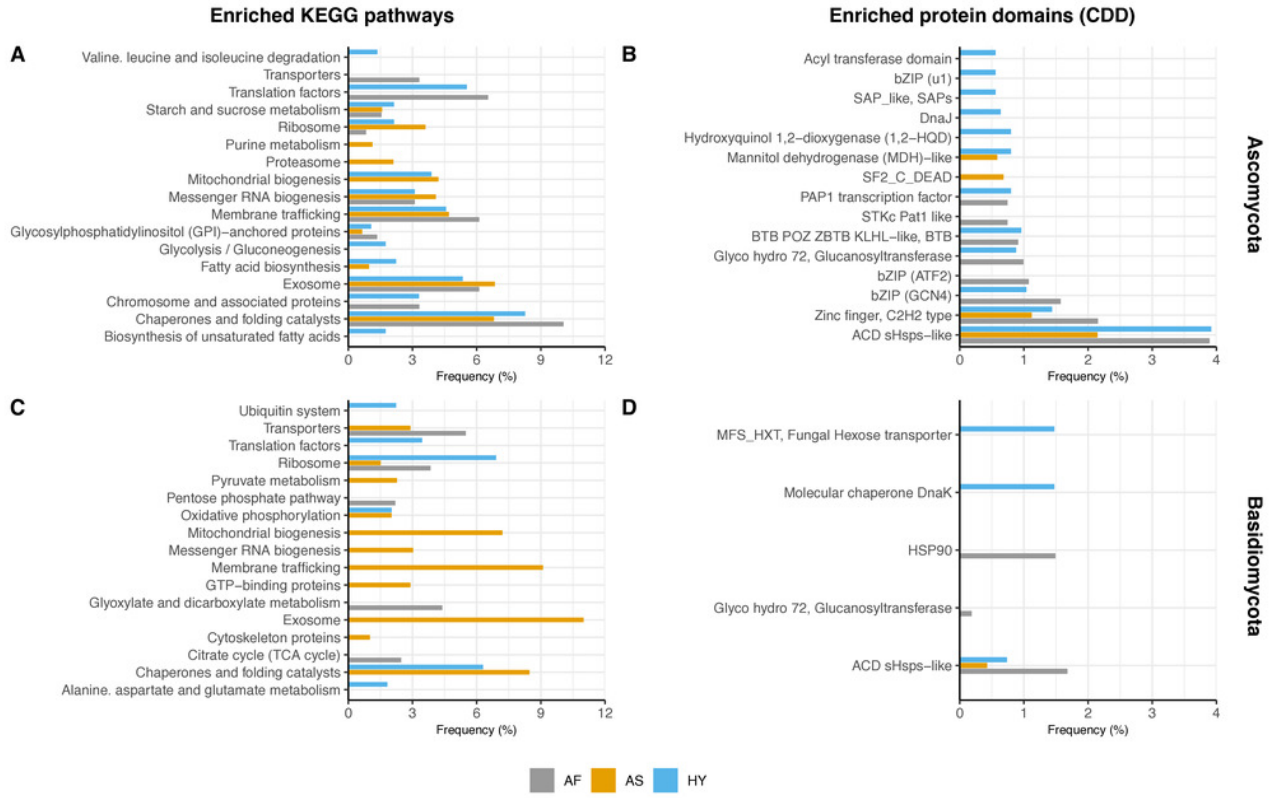

Figure 3

Number of proteins identified as different classes of Carbohydrate Active enZYmes (CAZymes)

A: All CAZymes found in each fungal community, showing the differences between the pattern of Ascomycota and Basidiomycota. B: Comparison of secreted CAZymes between the Ascomycota, Basidiomycota, and in the host plant. Secreted CAZymes have a signal peptide identified by SignalP. Asco: Ascomycota; Basidio: Basidiomycota; AA: auxiliary activities; CBM: carbohydrate-binding molecule; CE: carbohydrate esterases; GH: glycoside hydrolases; GT: glycosyltransferases; PL: polysaccharide lyases; AF: *Agave fourcroydes*; AS: *Agave sisalana*; HY: hybrid 11648.

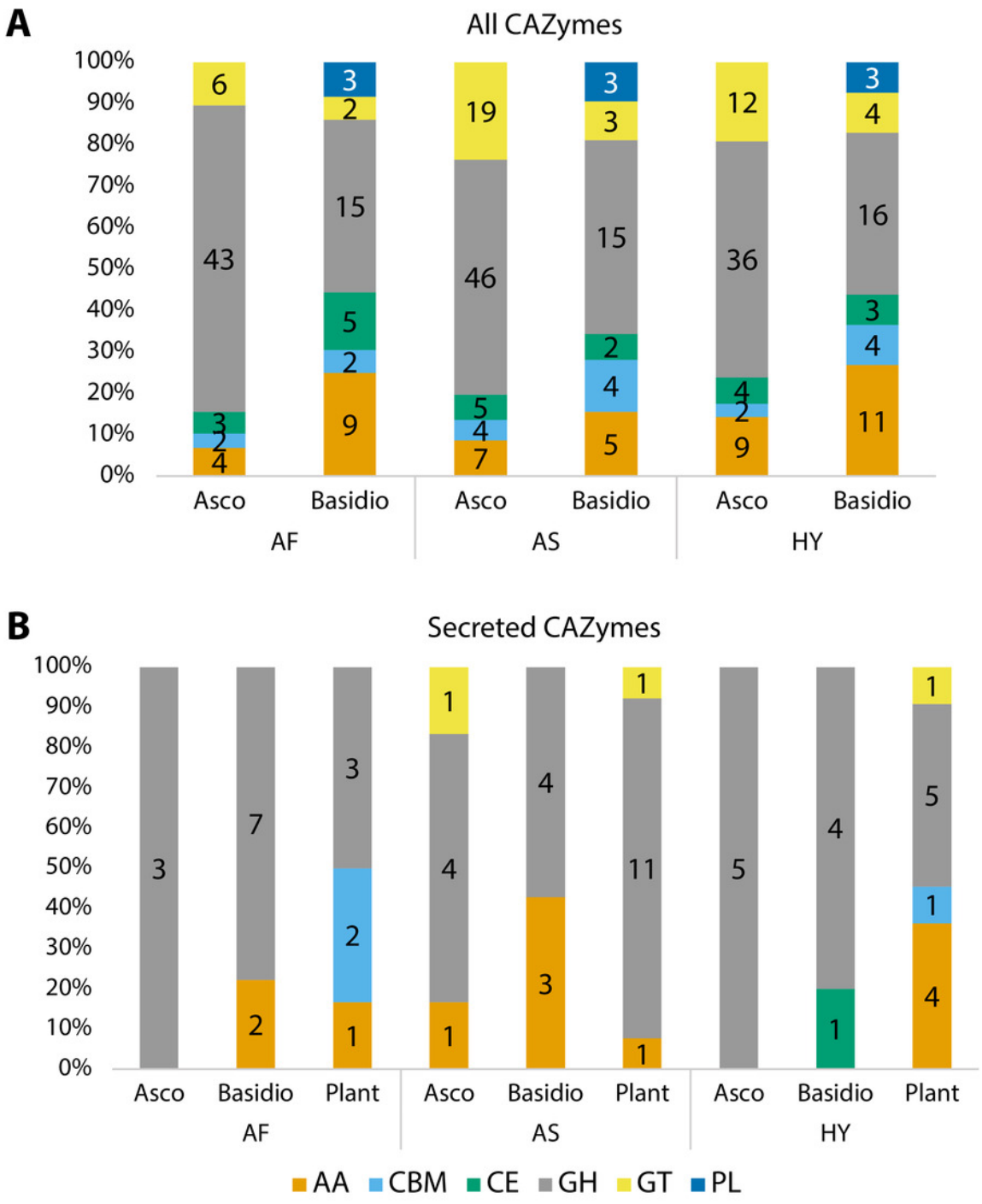
A
All CAZymes
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
6
43
3
2
4
3
2
15
5
2
9
19
46
5
4
7
3
3
15
2
4
5
12
36
4
2
9
3
4
16
3
4
11
Asco
Basidio
Asco
Basidio
Asco
Basidio
AF
AS
HY
B
Secreted CAZymes
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
3
2
7
1
2
3
1
4
1
3
4
1
11
1
5
1
4
1
5
1
4
Asco
Basidio
Plant
Asco
Basidio
Plant
Asco
Basidio
Plant
AF
AS
HY
AA
CBM
CE
GH
GT
PL

Figure 4

Protein orthologous analysis of the heat shock proteins (HSPs) in our Ascomycota and Basidiomycota datasets and the 10 Ascomycota and 10 Basidiomycota genomes with more HSPs.

Only the families presenting proteins from the agave datasets are represented. Annotation was according to the result of the HMM search against HSPIRDB. Scale is in number of proteins. AF: *Agave fourcroydes*; AS: *Agave sisalana*; HY: hybrid 11648.

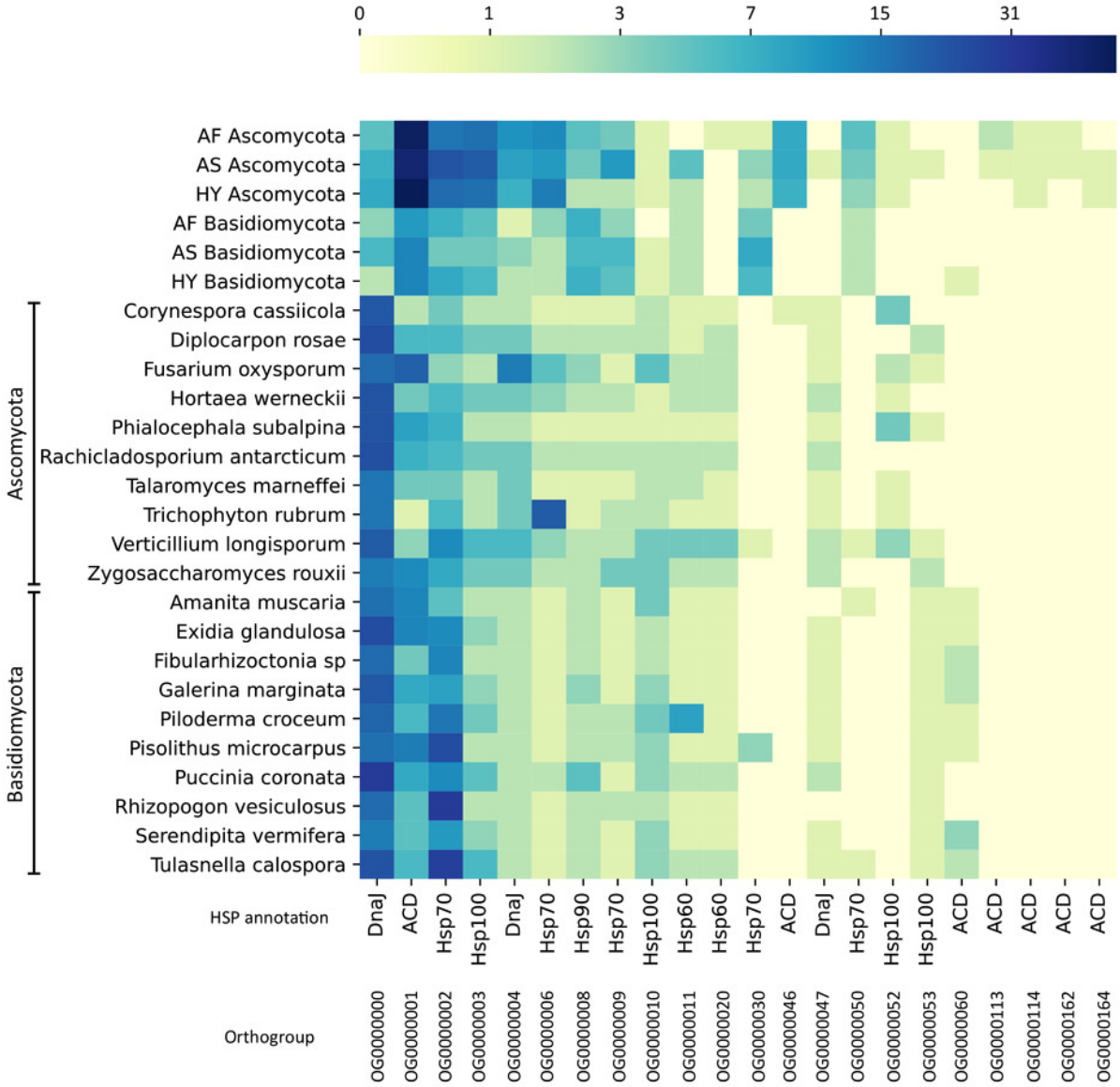
0
1
3
7
15
31
AF Ascomycota
AS Ascomycota
HY Ascomycota
AF Basidiomycota
AS Basidiomycota
HY Basidiomycota
Corynespora cassiicola
Diplocarpon rosae
Fusarium oxysporum
Hortaea werneckii
Phialocephala subalpina
Rachicladosporium antarcticum
Talaromyces marneffei
Trichophyton rubrum
Verticillium longisporum
Zygosaccharomyces rouxii
Amanita muscaria
Exidia glandulosa
Fibularhizoctonia sp
Galerina marginata
Piloderma croceum
Pisolithus microcarpus
Puccinia coronata
Rhizopogon vesiculosus
Serendipita vermifera
Tulasnella calospora
Ascomycota
Basidiomycota
HSP annotation
DnaJ
ACD
Hsp70
Hsp100
DnaJ
Hsp70
Hsp90
Hsp70
Hsp100
Hsp60
Hsp60
Hsp70
ACD
DnaJ
Hsp70
Hsp100
Hsp100
ACD
ACD
ACD
ACD
ACD
Orthogroup
OG0000000
OG0000001
OG0000002
OG0000003
OG0000004
OG0000006
OG0000008
OG0000009
OG0000010
OG0000011
OG0000020
OG0000030
OG0000046
OG0000047
OG0000050
OG0000052
OG0000053
OG0000060
OG0000113
OG0000114
OG0000162
OG0000164

Table 1(on next page)

Transcript numbers in each fungi group

Transcripts were identified with our in-house pipeline described in Methods S2. “Asco or Basidio” refers to transcripts that were either Ascomycota or Basidiomycota but could not be classified.

Cultivar	Ascomycota	Basidiomycota	Asco or Basidio	Other fungi	Total
Agave fourcroydes	1,927	797	19	253	2,996
Agave sisalana	3,012	1,036	23	242	4,313
Hybrid 11648	1,986	1,179	18	250	3,433

1

Table 2 (on next page)

Top expressed transcripts identified as transporters in fungal communities of the three agave cultivars.

Protein sequences were blasted against the Transporter Classification Database (TCDB) with E-value < 0.0001 and filtered for alignment coverage >= 70%. Expression values are in TPM.

Cultivar	ID	TCDB family	Fungal classification	Root mean expression (TPM)
A. fourcroydes	AF_DN51128_c6_g2	The HSP90/CDC37 (HSP90/CDC37)	Ascomycota	121.53
	AF_DN37348_c0_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Basidiomycota	97.41
	AF_DN51128_c6_g1	The HSP90/CDC37 (HSP90/CDC37)	Basidiomycota	86.05
	AF_DN46089_c1_g1	The Mitochondrial Carrier (MC)	Basidiomycota	69.20
	AF_DN39498_c0_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Asco or Basidio	63.63
	AF_DN52052_c2_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	48.26
	AF_DN103552_c0_g1	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	48.24
	AF_DN42546_c0_g1	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Basidiomycota	36.56
	AF_DN44188_c2_g2	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	28.82
	AF_DN50952_c2_g1	The Nuclear mRNA Exporter (mRNA-E)	Ascomycota	27.95
A. sisalana	AS_DN53864_c3_g1	The HSP90/CDC37 (HSP90/CDC37)	Ascomycota	99.52
	AS_DN59592_c8_g2	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	99.50
	AS_DN51359_c2_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	71.33
	AS_DN53419_c0_g2	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	67.31
	AS_DN53419_c0_g3	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	56.72
	AS_DN50667_c0_g1	The Mitochondrial Carrier (MC)	Ascomycota	47.87
	AS_DN54411_c1_g1	The Mitochondrial Carrier (MC)	Ascomycota	28.98
	AS_DN30106_c0_g1	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	24.04
	AS_DN48395_c0_g1	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Basidiomycota	23.72
	AS_DN56192_c3_g1	The Nuclear mRNA Exporter (mRNA-E)	Ascomycota	21.97
Hybrid 11648	HY_DN39331_c2_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	413.13
	HY_DN32985_c0_g1	The Mitochondrial Carrier (MC)	Basidiomycota	46.37
	HY_DN36716_c2_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	37.86
	HY_DN36716_c3_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	32.70
	HY_DN28452_c0_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Basidiomycota	26.55
	HY_DN38958_c4_g2	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Basidiomycota	23.08
	HY_DN39005_c7_g1	The HSP90/CDC37 (HSP90/CDC37)	Basidiomycota	22.64
	HY_DN10827_c0_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	21.63
	HY_DN70040_c0_g1	The Cation Channel-forming Heat Shock Protein-70 (Hsp70)	Ascomycota	20.58
	HY_DN39331_c3_g1	The Endoplasmic Reticular Retrotranslocon (ER-RT)	Ascomycota	19.66

1

Table 3(on next page)

All secreted CAZymes in the three fungal communities.

Plant cell wall (PCW) putative substrate was obtained from Kameshwar *et al.* (2019).

Expression values are in TPM.

Cultivar	ID	CAZy ID	CAZy classification	PCW putative substrate	Fungal classification	Root mean expression (TPM)
A. fourcroydes	AF_DN43332_c0_g1	GH79	β-glucuronidase	-	Asco or Basidio	6.92
	AF_DN16270_c0_g1	GH43	β-xylosidase	Hemicellulose	Basidiomycota	1.96
	AF_DN37792_c0_g1	GH18	chitinase	-	Basidiomycota	9.53
	AF_DN37367_c0_g1	GH18	chitinase	-	Basidiomycota	8.13
	AF_DN64727_c0_g1	GH16	β-glucanase	Hemicellulose	Ascomycota	2.52
	AF_DN54652_c0_g1	GH16	β-glucanase	Hemicellulose	Ascomycota	3.46
	AF_DN5518_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	7.91
	AF_DN44069_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	6.04
	AF_DN35139_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	7.18
	AF_DN84320_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	2.9
	AF_DN32526_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Basidiomycota	2.66
	AF_DN64239_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Basidiomycota	3.25
A. sisalana	AS_DN131388_c0_g1	GH76	cell wall α-1,6-mannotransglycosylase / α-1,6-mannanase	-	Basidiomycota	3.12
	AS_DN52648_c0_g1	GH76	cell wall α-1,6-mannotransglycosylase / α-1,6-mannanase	-	Ascomycota	4.39
	AS_DN25360_c0_g1	GH18	chitinase	-	Basidiomycota	9.34
	AS_DN46396_c2_g1	GH17	β-1,3-glucanase	Cellulose	Ascomycota	6.51
	AS_DN61972_c0_g1	GH17	β-1,3-glucanase	Cellulose	Ascomycota	1.65
	AS_DN41449_c0_g1	GH16	β-glucanase	Hemicellulose	Ascomycota	3.66
	AS_DN87265_c0_g1	GH16	β-glucanase	Hemicellulose	Basidiomycota	3.47
	AS_DN100681_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	12.16
	AS_DN42067_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Basidiomycota	2.31
	AS_DN114687_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Basidiomycota	3.5
	AS_DN40194_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Basidiomycota	5.42
	AS_DN16467_c0_g1	AA9	Lytic cellulose monooxygenase	Cellulose/Hemicellulose	Ascomycota	2.15
Hybrid 11648	HY_DN20444_c0_g1	GH79	β-glucuronidase	-	Asco or Basidio	3.19
	HY_DN23374_c0_g1	GH76	cell wall α-1,6-mannotransglycosylase / α-1,6-mannanase	-	Ascomycota	2.4
	HY_DN94917_c0_g1	GH18	chitinase	-	Basidiomycota	5.42
	HY_DN80813_c0_g1	GH17	β-1,3-glucanase	Cellulose	Ascomycota	1.27
	HY_DN39599_c0_g1	GH16	β-glucanase	Hemicellulose	Ascomycota	2.28
	HY_DN40185_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	4.89
	HY_DN27892_c0_g1	GH128	endo-β-1,3-glucanase	-	Basidiomycota	6.75
	HY_DN26068_c0_g1	GH11	Xylanase	Hemicellulose	Basidiomycota	6.82
	HY_DN56898_c0_g1	GH10	Xylanase	Hemicellulose	Ascomycota	5.87
	HY_DN50387_c0_g1	CE1	acetyl xylan esterase	Lignin	Basidiomycota	3.04

1