Received: 25 April 2017 | Accepted: 3 July 2017

DOI: 10.1111/efp.12372

ORIGINAL ARTICLE

WILEY Forest Pathology

Adaptation of the pine fungal pathogen *Grosmannia clavigera* to monoterpenes: Biochemical mechanisms revealed by RNA-seq analysis

E. Kligun[1] | B. Ostretsov[1] | A. Titievsky[1] | M. Farkov[1,2] | S. M. Alamouti[3] | L. Brodsky[1,2]

[1]Tauber Bioinformatics Research Center, University of Haifa, Haifa, Israel

[2]The Department of Evolutionary and Environmental Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel

[3]Department of Wood Science, University of British Columbia, Vancouver, BC, Canada

Correspondence
Leonid Brodsky
Tauber Bioinformatics Research Center, University of Haifa, Haifa, Israel.
Email: lbrodsky@research.haifa.ac.il

Funding information
Laszlo N. Tauber Family Foundation

Editor: C. G. Fossdal

Summary

The blue stain fungus *Grosmannia clavigera (G. clavigera)* is a pathogen of pines that can tolerate monoterpenes in oleoresin. *Grosmannia clavigera* has developed several mechanisms to cope with the host's monoterpene defence: a monoterpene efflux system mediated by an ABC transporter and enzymes that utilize or modify monoterpenes. In this study, we integrated two expression data sets based on a match between their multidimensional distributions. The data sets are raw RNA-seq and already processed transcriptome expression data with partially matching biological conditions. As a result, several new regulatory mechanisms, including the upregulation of the stress response and the developmental process at early-time-point terpene treatment, were involved in the adaptation of *G. clavigera* to monoterpenes and were supported by both data sets. In addition, several genes related to terpenoid modification were upregulated within 36 h (in rich media) and 7 days (in poor media) of terpene treatment. These results led us to the assumption that *G. clavigera* used terpenes as an energy source within 7 days when no other energy source was available, while within 36 h of terpene treatment on rich media these genes took part in terpene detoxification. As the *G. clavigera* genome is not well annotated, we performed detection of new putative transcripts as genome fragments enriched by read mapping. Several newly detected transcripts annotated as putative retrotransposons were upregulated mostly under early-time-point terpene treatment, while others annotated as putative ribonuclease and protease were downregulated mostly during early-time-point terpene treatment. Based on these results, it can be hypothesized that these new transcripts could play a role in previously unidentified *G. clavigera* terpene response regulatory mechanisms. Overall, this study identified gene expression regulations supported by two data sets as well as new putative transcripts related to new processes that might be important for *G. clavigera* tolerance to terpenes.

1 | INTRODUCTION

In recent years, phytophagous insects with fungal pathogens attacking coniferous trees have drawn a lot of attention, both from industrial companies and the scientific community, due to the enormous amount of damage they cause. In addition to local environmental destruction and economic harm to forest products industries, the damage caused by these fungal pathogens has important implications with regard to

climate change through the obvious impact of tree mortality on forest carbon dynamics (Kurz et al., 2008). Pines have several defence mechanisms against insects, pests and fungi. One of these mechanisms is terpene-based, including the secretion of a complex mixture of monoterpenes, sesquiterpenes and diterpenoid acids as oleoresin (Bohlmann, 2012; Franceschi, Krokene, Christiansen, & Krekling, 2005; Phillips & Croteau, 1999). Some bacteria mitigate the toxic effects of monoterpenes by modifying their membrane lipids, transforming monoterpenes and/or using active transport by efflux pumps. It has been shown that in concentrations below the toxic level, microorganisms can use monoterpenes as their sole carbon and energy source (Marmulla & Harder, 2014). Similar utilization of monoterpenes is expected in fungi, *Grosmannia clavigera* in particular.

Grosmannia clavigera is a fungal pathogen of pine trees that forms an association with the *Dendroctonus ponderosae* [mountain pine beetle] and its sister species, the *D. jeffreyi* [Jeffrey pine beetle]. *Grosmannia clavigera* is considered as symbiotically acting with these beetles to subvert the conifers' natural defences, and it is a central component of the mountain pine beetle as well as the Jeffrey pine beetle epidemic (Alamouti et al., 2011). However, this hypothesis was challenged by Six and Wingfield (2011), who claimed that phytopathogenicity might be important to the fungi rather than to the bark beetle because fungi need to survive in a defensive host, up to a year, until the beetle completes development and can transport the fungi to a new host. Enhanced understanding of the *G. clavigera* defence mechanism should therefore aid in the protection of conifers from these infestations. The genome of *G. clavigera* is approximately 30 Mb and consists of 18 supercontigs with 8314 protein-coding gene models (Diguistini et al., 2009; DiGuistini et al., 2011). In this study, this genome assembly was used as a reference for mapping the transcriptome reads and the follow-up calculating expression levels of the annotated as well as newly detected transcripts.

Several fungi have also been described as tolerating the toxic effect of monoterpenes during growth in rich media (Farooq, Atta-ur-Rahman, & Choudhary, 2004). However, until now there has been no evidence in the literature of fungi using monoterpenes as their sole carbon and energy source for growth. Recently, *G. clavigera* was shown to grow on a mixture of monoterpenes and diterpenes, containing α/β-pinene and 3-carene (DiGuistini et al., 2011). Previously, ABC efflux transporters, cytochrome P450 enzymes and other enzymes that are highly induced by monoterpenes and that metabolize or utilize monoterpenes were considered to be major mechanisms enabling *G. clavigera* resistance to monoterpenes (Lah, Haridas, Bohlmann, & Breuil, 2013; Wang et al., 2013, 2014).

To better understand the unique adaptation process of *G. clavigera* to terpenes, we conducted an RNA-seq analysis of *G. clavigera* transcriptome expression under several growing conditions (GEO accession number GSE43006). We then compared these expression data to processed RNA-seq expression data from DiGuistini et al. (2011). Our results showed similarity in major transcription regulation mechanisms between these two data sets and also revealed new transcripts and processes putatively influencing *G. clavigera* tolerance to terpenes. In addition, the results led us to hypothesize that *G. clavigera* uses monoterpenes as an energy source when no other energy source is available.

2 | MATERIALS AND METHODS

2.1 | RNA-seq data

The next-generation sequencing (NGS) data were obtained from the public domain GEO database, GEO accession number GSE43006. We analysed eight samples: malt extract agar (MEA)-Control 12 hours (h); MEA-Terpene 12 h; *GcABC-G1* mutant-MEA-Control 12 h; *GcABC-G1* mutant-MEA-Terpene 12 h; yeast nitrogen base medium (YNB)-Mannose 5 days (d); YNB-Olive Oil (OO) 5 days; YNB-Oleic Acid (OA) 5 days; and YNB-Terpene 7 days. Three of these samples served as controls: MEA-Control 12 h served as a control for MEA-Terpene 12 h; mutant-MEA-Control 12 h served as a control for mutant-MEA-Terpene 12 h; and YNB-Mannose 5 days served as a control for YNB-Olive Oil 5 days, YNB-Oleic Acid 5 days and YNB-Terpene 7 days.

Another data set was taken from processed expression data from DiGuistini et al. (2011) (raw data were unavailable). The data set contains the four experimental samples (12-h terpene, 12-h LPPE, 36-h terpene and 36-h LPPE), along with an additional four samples of corresponding controls.

2.2 | Analysis of the NGS RNA-seq data set

Computational analysis was performed using several RNA-Seq pipelines of the T-BioInfo platform (http://tauber-data2.haifa.ac.il:3000/). The pipelines comprised of the following stages (Figure 1).

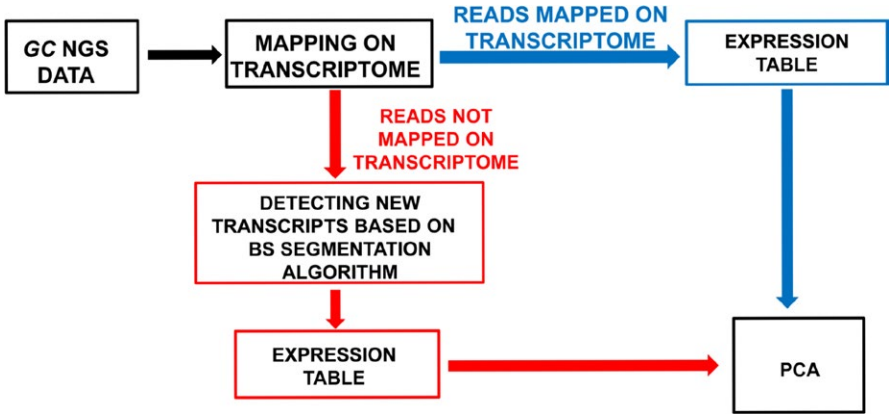

FIGURE 1 A logical scheme of RNA-seq data analysis of *Grosmannia clavigera* using the T-BioInfo platform. Two RNA-seq analysis pipelines were used. In one pipeline depicted by blue arrows, the NGS reads were mapped on known transcripts, and expression levels of transcripts were calculated. In the red arrow pipeline, the NGS reads were mapped on *G. clavigera* contigs, and then transcripts were detected as fragments enriched by the mapped reads. (tiff)

The NGS reads were mapped onto the *G. clavigera* transcriptome, followed by analysis using two pipelines in the T-BioInfo platform. The processed reads successfully mapped onto the known *G. clavigera* transcriptome are shown in blue; non-mapped processed reads are shown in red.

2.2.1 | Mapping of reads

The mapping of reads on the *G. clavigera* transcriptome was performed by the newly developed BSKLB-GPU mapping algorithm implemented in our T-BioInfo platform (http://tauber-data2.haifa.ac.il:3000/). The platform is licensed to Pine Biotech Inc. by the University of Haifa for educational and commercial use: http://tbioinfopb.pine-biotech.com/. The algorithm of the BSKLB-GPU mapping procedure starts by generating all *k*-mers, which are presented in the reference sequences. Each *k*-mer is used to produce a set of *k*-mers by sequential changing of one letter at every position across a *k*-mer. The extended set of all *k*-mers is used in mapping of reads. Thus, BSKLB-GPU allows in the exact matching between a read and the reference sequence because it generates both strictly corresponding and modified *k*-mers, and also permits a limited difference in offsetting the *k*-mers mapped on the reference sequence. The code of BSKLB-GPU was written using NVIDIA CUDA® technology.

2.2.2 | Detection of new putative transcripts

Detection of new putative transcripts (the analysis was performed only for samples of the GEO accession number GSE43006 data set) was based on the BS segmentation algorithm (Brodsky, Kogan, Benjacob, & Nevo, 2010), which detects genomic fragments significantly enriched by the read mapping counts. New putative transcripts were annotated using Blast2Go (with BLASTx as a search engine) (Conesa & Götz, 2008; Conesa et al., 2005; Götz et al., 2008). The annotated novel transcripts reported here do not overlap with genes from the existing *G. clavigera* genome annotation.

2.2.3 | Quantification of transcript expression levels

Transcript expression quantification in reads per kilobase per million mapped reads.

(RPKM) units was performed by the RSEM algorithm (Li & Dewey, 2011) after a new mapping of reads on an updated set of genomic transcripts.

2.2.4 | Post-processing of gene expression data

As a first step, the logarithmic values of transcript abundance were mutually normalized by the quantile method (Bolstad, Irizarry, Astrand, & Speed, 2003), and then these values were normalized according to the control-treatment expression data (per gene log-ratio of treatment vs. control). The statistical significance of differential expression of genes in experimental vs. control treatments was calculated by Kal's Z-test on expression proportions (the expression of a specific gene per sample as a fraction of the total expression of all genes in that sample) (Kal et al., 1999).

2.3 | Integration of two transcriptome expression data sets and detection of matching transcript regulations

Using PCA, we projected the normalized-by-control-treatment (NCT) expression data on a plane of the first two principal components (PC1 and PC2). The analysis of PCA was performed separately for the two data sets, and then PC1-PC2 planes for the two analyses were superimposed (Appendix S1) (the same PCA procedure was carried out separately for new putative transcripts). As raw data were processed in the GEO NGS data set, and this data set contains more samples, the GEO NGS data set was designated as the leading one for selecting the genes differentiated by D1 and D2 directions. To perform the selection of significantly differentiating genes, projections on D1 and D2 of the gene profiles were standardized, and genes with D1 or D2 projections deviating from null differentiation (the centre) by more than two standard deviations (below -2SD or above 2SD) were selected under the additional condition that the ratio between D scores (separately for D1 or D2) in comparison with the corresponding D scores for the DiGuistini et al. (2011) data set was less than two.

Classification of gene functions for the genes differentiated by D1 and D2 was carried out using Blast2Go (Conesa & Götz, 2008; Conesa et al., 2005; Götz et al., 2008).

2.4 | Individual genes as classifiers for the early terpenes, late terpenes and non-terpene classes of samples

Using the control normalized expression values of all genes from Table S3, the per-sample posterior log-odds of Naïve Bayesian Classifiers (Duda, Hart, & Stork, 2000) were calculated for samples of three classes in three contrasts: early terpenes vs. late terpenes, early terpenes vs. no terpenes and late terpenes vs. no terpenes. A gene is a classifier for one of the contrasts if it provides perfect log-odd classifications for all samples to two corresponding classes of the contrast. Information about each gene as a classifier across three contrasts and corresponding log-odds of the samples can be found in Table S4. The same Naïve Bayesian techniques for finding classifiers were used earlier (Douglass et al., 2016).

For every gene, the control normalized log-expression values in nine samples were divided into three biological groups: (class C0) no terpenes in the media, (class C1) early-time terpenes and (class C2) late-time terpenes. This grouping of samples into three classes across all genes was supported by results of the Principal Component Analysis (Figure 2e). Thus, for every gene, the distributions of log-signals were considered as normal ones, in each class with corresponding mean and variance. The posterior probabilities of the Naïve Bayesian Classifier to belong to a given class were calculated for each log-expression value of the gene in contrasts between the classes as follows: C0–C1, C0–C2 and C1–C2. Under the null hypothesis on independence

of measurements, the posterior probability for all log-expressions of a particular biological group was calculated as the product of posterior probabilities of the individual log-expressions. The biological group log-odds in the above contrasts were calculated from posterior probabilities of the groups. If every biological group of log-expressions of a gene having positive and high log-odd values is placed in its own class in all three contrasts, then the gene is considered a good classifier—the gene differentiates all three biological groups by its log-expression values. If gene expression values of a biological group receive high positive log-odds in its own class against two other classes and do not receive good enough positive log-odds in the third contrast, this gene will be the class-specific discriminator.

3 | RESULTS AND DISCUSSION

3.1 | Relationships between the different RNA-seq samples of *Grosmannia clavigera*

In the present study, to understand the adaptation of *G. clavigera* to monoterpene, an analysis of RNA-Seq data of *G. clavigera* (GEO accession number GSE43006) growing under different conditions (MEA + terpene 12-h wild type, MEA + terpene 12-h *GcABC-G1* mutant, YNB + terpene 7 days, YNB + OO 5 days, YNB + OA 5 days) was performed and compared with the processed RNA-Seq data from DiGuistini et al. (2011). In the GEO project, the PCA of expression profiles of the five samples, normalized by corresponding control treatments (NCT data; see Materials and Methods), showed biologically meaningful separation of samples into three groups on the PC1–PC2 plane (which covers 77.8% of the total data variability) (Figure 2c). The three groups separated according to treatment time-point as well as type of treatment are as follows: early-time-point terpene (terpene treatment, 12-h growth, wild type and mutant); late-time-point terpene (terpene, 7 days) and media without terpenes (OA and OO, 5-days growth). The PCA of the data set from DiGuistini et al. (2011) showed a separation into three similar groups of samples on the PC1–PC2 plane (which covered 58.6% of the total data variability): early-time-point terpene (terpene, 12 h), late time-point terpene (terpene, 36 h) and media without terpenes (LPPE, 12 h and 36 h) (Figure 2a). After superimposing PC1–PC2 planes for the two analyses, the positions of samples from the two projects on the united PC1–PC2 plane appear to be similarly grouped: (1) early-time-point terpene treatment, (2) late-time-point terpene treatment and (3) media without terpenes. As depicted in Figure 2, a diagonal between the negative direction of PC1 and the positive direction of PC2 (designated D1) is a differentiating direction: early- and-late terpene groups from both projects were separated along this D1 direction. A diagonal between the positive direction of PC1 and the positive direction of PC2 was designated D2: all terpene samples were separated from terpene-free samples along the D2 direction. Selection of the significantly differentiating genes in the early vs. late and terpene vs. other contrasts was based on scores of projection of each gene profile onto the D1 and D2 directed lines in two data sets separately (see Materials and Methods) (Figure 3).

3.2 | Putative biological processes differentiating between groups of growth conditions

Using Blast2Go (Conesa & Götz, 2008; Conesa et al., 2005; Götz et al., 2008), we ascribed putative biological processes to the genes from the four separated groups (D1 positive, D1 negative, D2 positive and D2 negative). As shown in Figure 4, all four groups were characterized by the processes of oxidation–reduction and transmembrane transport; however, the genes that belonged to each process were mostly different between the groups (Table S2).

Additionally, several other processes were typical for specific groups. Namely, genes that were upregulated at early-time terpene treatment (Figure 4a, D1 positive) were related to regulation of the transcription processes (and indications that the terpene treatment induced *transcriptome reprogramming* were also shown by DiGuistini et al., 2011), as well as developmental processes and response to stress processes, which include genes such as heat shock proteins and transcription factors that may act as regulators of all these processes. Furthermore, using GO analysis, we found that the clusters of genes associated with early-time-point terpene treatments are involved in *phosphorylation* and *cellular protein modification* processes. These early-time-point genes also include ones, such as serine threonine protein kinases, which may also serve as *stress signalling regulators*, enabling the fungus to cope with the stress of growing on a medium with terpenes at an early time-point. This observation is supported by several previously published articles, where it has been shown for the filamentous fungus *Neurospora crassa* that calcium/calmodulin-dependent kinases participate in growth, thermotolerance and oxidative stress survival (Kumar & Tamuli, 2014). Also, it is known that, in *Saccharomyces cerevisiae*, Cbk1 kinase regulates MAP kinase activation/inactivation during heat shock (Kuravi, Kurischko, Puri, & Luca, 2011).

Genes that were upregulated during terpene treatment at a late-time-point (D1 negative) and genes upregulated in other samples (D2 positive) were related to different metabolic processes (Figure 4b,c). One of these metabolic processes was a carbohydrate metabolic process in the group that included fatty acid and triglyceride samples. This observation matches with already known regulatory mechanisms: indeed, previously, in *Ophiostoma piceae*, a wood-staining fungus that grows on conifers, and it was shown that genes predicted to be involved in the breakdown of carbohydrates and sugars were upregulated during growth on triglycerides and fatty acids (Haridas et al., 2013). Genes from the D2 negative group (generally, upregulated with terpene and downregulated in other samples) were, as expected, related to pathways involved in terpene treatment at both early and late time-points (Figure 4d).

3.3 | Linkage between genes participating in the same biological processes

We classified the genes into eight groups according to their biological processes ascribed by Blast2Go (Conesa & Götz, 2008; Conesa et al., 2005; Götz et al., 2008). As can be seen from the Table S2, one

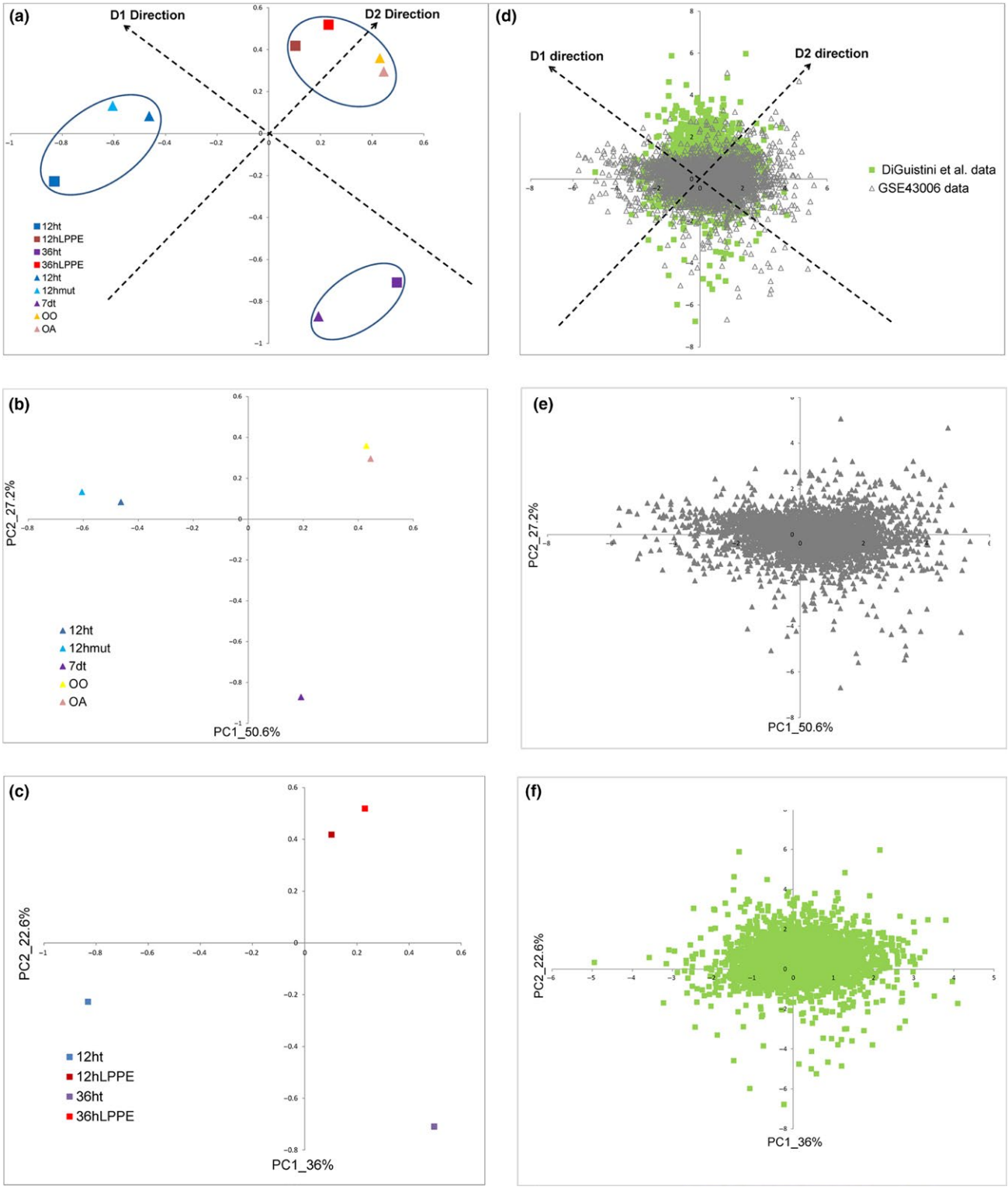

FIGURE 2 Separation between groups of *Grosmannia clavigera* RNA-seq samples from two data sets. (a) A projection of the normalized expression profiles from the data set of DiGuistini et al. (2011) on the plane of the first two principle components prepared for this data set. (b) DiGuistini et al. (2011) data set: gene profile positions on the plane of the first two principal components. (c) A projection of the normalized expression profiles from the GSE43006 data set on the plane of the first two principle components prepared for this data set. (d) GSE43006 data set: gene profile positions on the plane of the first two principal components. (e) The PC1–PC2 planes of two data sets were superimposed by corresponding PCs, and samples of of data sets were denoted on the superimposed plane with the same positions as on the original PCA planes. The samples of the DiGuistini et al. (2011) data are denoted as squares; the samples of the GSE43006 data are denoted as triangles. The samples of two data sets from similar biological condition are clustered together. (f) Gene profile positions on the superimposed PCA plane for genes from the two data sets. (tiff)

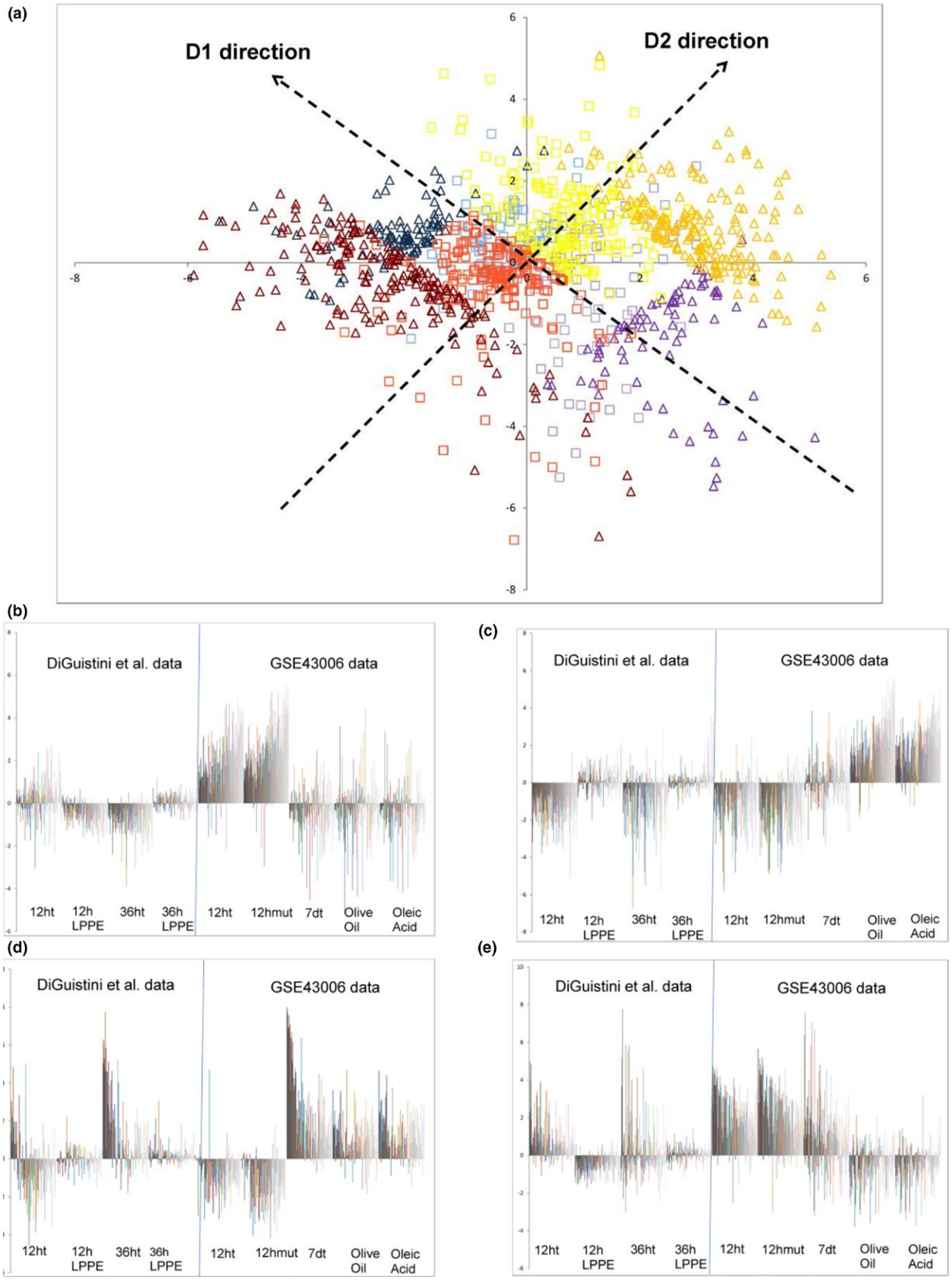

FIGURE 3 Selection of genes significantly differentially expressed between different growth states of *Grosmannia clavigera*. (a) Positions of gene profiles differentiating between groups of growth conditions on the united PCA plane for the two data sets. Gene profiles from the DiGuistini et al. (2011) data set are denoted as squares (light colours), and gene profiles from the GSE43006 data set are denoted as triangles (dark colours). Gene groups are highlighted in specific colours as follows: blue, genes upregulated under early-time-point terpene treatments; purple, genes upregulated under late-time-point terpene treatments; yellow, genes upregulated under non-terpene conditions; and red, genes upregulated under all terpene treatments. (b) Expression profiles for genes that were upregulated under early-time-point terpene treatment, across samples from both data sets (D1 significantly positive). (c) Upregulation of genes in fatty acid and LPPE treatments (D2 significantly positive). (d) Upregulation of genes in late-time-point terpene treatments (D1 significantly negative). (e) Upregulation of genes in all terpene treatments (D2 significantly negative). (tiff)

of the largest groups was a group of genes related to the *transporter process* (group 1). Indeed, the transport process activation was seen in all of the growth states of *G. clavigera* that we investigated. However, different genes participated in this process at different growth states. In a previous study, Wang and coauthors showed that fungal ABC efflux transporters (mainly *GcABC-G1*) play a major role in how *G. clavigera* copes with the diffusion of monoterpenes into its cells. When *GcABC-G1* was deleted, the mutant transcriptome showed an increased stress response to monoterpene exposure and upregulation of other ABC transporters, but none of them was expressed at the same level as wild-type *GcABC-G1*, and also none of them seemed to substitute the functionality of *GcABC-G1* in the mutant (Wang et al., 2013). In our analysis, high expression levels of *GcABC-G1*, both in early- and late-time-point terpene treatments were also observed, while in the mutant, their expression levels were very low. Other ABC transporters were also upregulated, but not as highly upregulated as the wild-type *GcABC-G1* expression levels.

Besides putative transporter genes that were found previously as upregulated under terpene treatment (Wang et al., 2013), we found additional putative transporter genes, which were upregulated under one of the studied treatments (Table S2). Namely, these genes were upregulated mainly during terpene treatment at an early time-point (p-value < .01, FDR < 0.05, D1 positive and/or D2 negative), such as *CMQ_1821* and *CMQ_6705* (Table 1, Table S3). Under late-time-point terpene treatment, there was upregulation (p-value < .01, FDR < 0.05, D1 negative and/or D2 negative) of several putative transporters: *CMQ_746*, *CMQ_5680*, *CMQ_2588*, *CMQ_1489*, *CMQ_6042* and *CMQ_3736* (Table 1, Table S3). Other putative transporters, such as *CMQ_4535*, *CMQ_5114*, *CMQ_1983*, *CMQ_2990* and *CMQ_5389*, were especially upregulated under fatty acid treatment (Table S3). Interestingly, several putative transporter genes were especially upregulated in the mutant strains *CMQ_2315*, *CMQ_5080* and *CMQ_4374* (Table S3).

These results indicated that the transporters (especially *GcABC-G1*) did indeed play a major role in the adaptation of *G. clavigera* to monoterpenes, both during early- and late-time-point terpene treatments. As shown by Wang et al. (2013), when *GcABC-G1* was deleted, the mutant survived the terpene treatment, but required 2-3 days longer than the wild-type *G. clavigera* to adapt to the chemical, and then grew more slowly than the wild type. Based on our results, one can hypothesize that other putative transporter genes that are upregulated in the mutant might partly complement the functionality of *GcABC-G1*, and thus enable survival of the mutant in the terpene media, but with a slower growth rate.

Another large group of genes was related to the *oxidation–reduction process* (Table S2, group 2). Many genes of this group, which were upregulated during late-time-point terpene treatments, have been identified previously (DiGuistini et al., 2011; Wang et al., 2014) as being related to terpene modification or degradation (examples of these genes: *CMQ_6740*, *CMQ_6956*, *CMQ_7007*, *CMQ_5369*, *CMQ_8234*, *CMQ_6887*, *CMQ_6536*, *CMQ_6937*, *CMQ_6612*, *CMQ_3621* and *CMQ_8299*). Intriguingly, several of these genes (such as *CMQ_3621*, *CMQ_8234* and *CMQ_8299*) were described in a previous study (Wang et al., 2014) as being upregulated during late-time-point terpene treatment only under minimal media (monoterpenes as a sole carbone source) and not under complete media. However, in our analysis, we noticed upregulation of those genes both 36-h (complete media) and 7-days (minimal media) samples. Furthermore, we found additional genes in the group *oxidation–reduction process* that may function in terpene modification or degradation; these genes are *CMQ_7160* and *CMQ_764*, and they both were upregulated during terpene treatment (p-value < .01, FDR < 0.05, D2 negative) (Table 1, Table S3).

Another mechanism that was described as allowing *G. clavigera* to deal with monoterpenoid toxicity was *modification or degradation of monoterpenes*. It has been suggested (DiGuistini et al., 2011; Wang et al., 2014) that some steps of monoterpene degradation occur through the beta-oxidation pathway, and mainly through genes located in the mitochondria, in contrast to fatty acid degradation that required genes located in peroxisomes and mitochondria. Also, cytochromes P450 have been suggested (Lah et al., 2013) to be involved in terpene metabolism.

In our analysis, we also noticed overexpression of cytochrome P450 gene (*CYP530A13*, *CMQ_3491*) at late-time-point terpene treatments (Table S3). Two other cytochrome P450 genes (*CMQ_5365* and *CMQ_5370*) that were previously observed as over-expressed in fatty acid and triglyceride media (Novak et al., 2015) were also upregulated in our analysis under the fatty acid and triglyceride treatments (Table S2, Table S3). In addition, the results from that previous study suggested the importance of W-oxidation when the main B-oxidation pathway may be overwhelmed by other substrates such as the host's terpenoid compounds. An additional group in the Table S2 (group 3) contained genes which were upregulated mainly under late-time-point terpene treatments and related to several *metabolic processes*, such as *carboxylic acid catabolic process*, *lipid metabolic process* and *small molecule metabolic process*. Among those genes were genes that previously were described as putatively related to the beta-oxidation pathway or valine catabolic pathway (*CMQ_3456*, *CMQ_6959*, *CMQ_1282*,

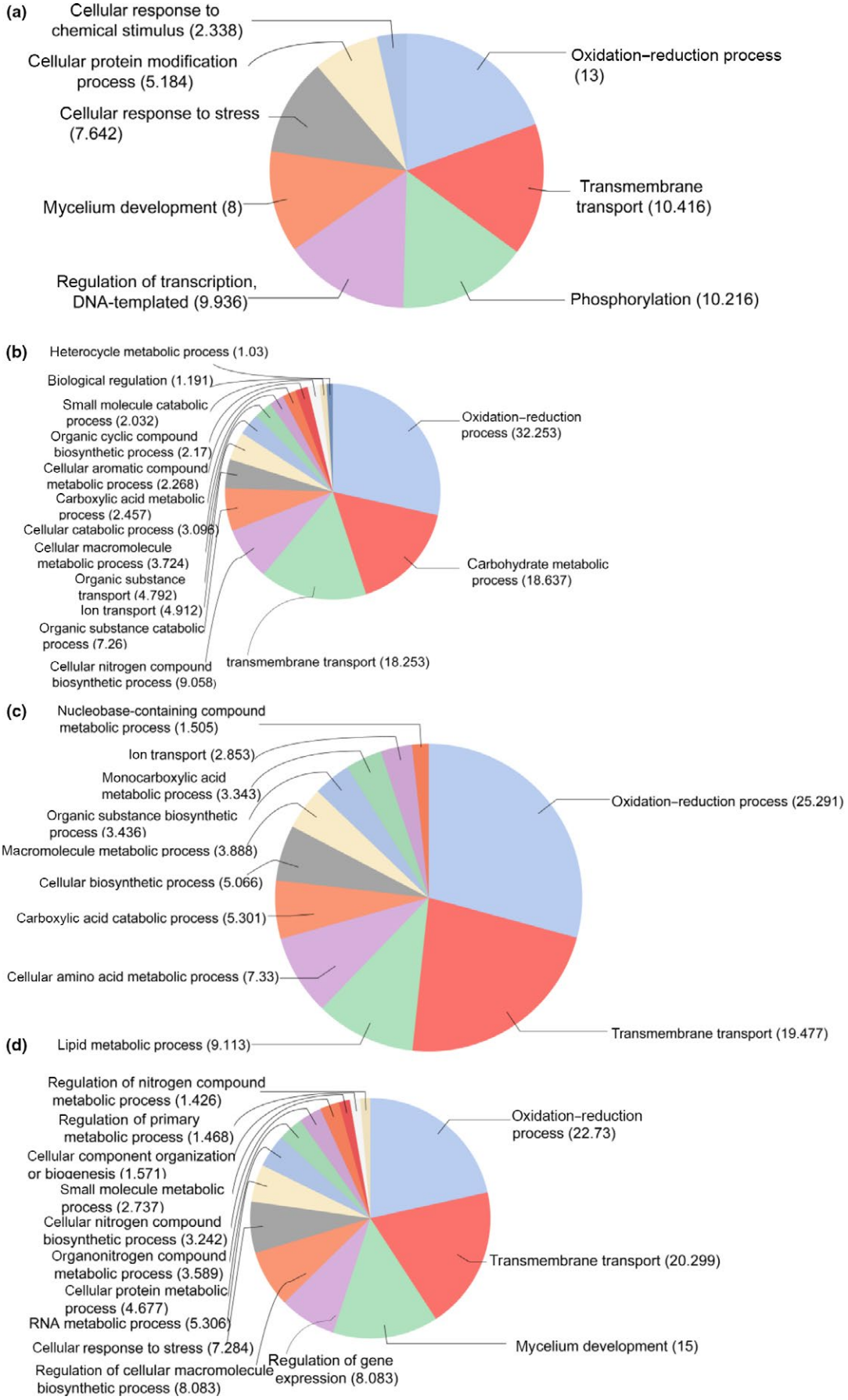

FIGURE 4 Putative biological processes enriched during different *Grosmannia clavigera* growth states. Each sector area represents the score of the corresponding processes determined by Blast2Go annotation for sets of differentiating genes: (a) for genes upregulated at early-time-point terpene treatment; (b) for genes upregulated at fatty acid/triglyceride and LPPE treatments; (c) for genes upregulated at late-time-point terpene treatments; and (d) for genes upregulated at terpene treatments. (tiff)

CMQ_4029 and *CMQ_8094*) (Wang et al., 2014). Here, we found additional genes upregulated during late-time-point terpene treatments (*p*-value < .01, FDR < 0.05, D1 negative and/or D2 negative); these genes could work as terpene modification or degradation agents (*CMQ_4724*, *CMQ_4999* and *CMQ_8300*) (Table 1, Table S3).

Group 4 also contained genes related to *several metabolic processes*, but they were mainly upregulated in the fatty acid and triglyceride media (Table S2).

Furthermore, group 5 contained general *transcription factors*, several of which are related to metabolic processes: some of these were upregulated under fatty acid treatment and others were upregulated under terpene treatment. Among them were genes that were previously mentioned as related to terpene modification or degradation (*CMQ_6803*, *CMQ_6601*, *CMQ_6820* and *CMQ_8045*) (DiGuistini et al., 2011; Wang et al., 2014). *CMQ_8045* was described previously as being upregulated only in response to late-time-point terpene treatments and minimal media conditions (Wang et al., 2014), but was observed in our analysis to be upregulated both in 36-h (complete media) and in 7-days (minimal media) samples. These results led us to an assumption that *G. clavigera* did not use monoterpenes as an energy source when other sources of energy were available. In early-time-point terpene treatments (12 h), as also mentioned in DiGuistini et al. (2011), it seems there were early steps of terpene modification (mostly including other genes from the later stages), while at 36 h (complete media), a little after the lag phase of the *G. clavigera* growth (24 h), the expression levels of the genes related to terpene modification were very high. Thus, it seems these genes took part in terpene detoxification. Later, in the 4-days sample (complete media), expression of many of these genes was reduced, which can be explained as reaching a steady state, in terms of regulation, that demanded less detoxification activity. However, in the 7-days sample (minimal media), the expression levels of these genes were again very high. This phenomenon could be due to the presence of additional functions of these genes; for example, not only in monoterpene detoxification but also in monoterpene utilization as an energy source when no other carbon source is available.

Besides the above-mentioned genes, we found additional putative transcription factors, which may be related to terpene modification or degradation processes, including *CMQ_1329*, which is upregulated under late-time-point terpene treatments (*p*-value < .01, FDR < 0.05, D1 negative); *CMQ_6445*, which is upregulated under early-time-point terpene treatment (*p*-value < 0.01, FDR < 0.05, D1 positive and D2 negative); and *CMQ_4220*, which is upregulated under terpene treatment in general (Table 1, Table S3).

Finally, the following putative processes, *cellular protein modification* and *phosphorylation* (group 6, e.g., gene *CMQ_523*), *mycelium development* (group 7, e.g., genes *CMQ_7872* and *CMQ_6292*) and *cellular response to stress* and *cellular response to chemical stimulus* (group 8, e.g., genes *CMQ_4400*, *CMQ_5857*, *CMQ_812* and *CMQ_7087*) were upregulated mainly during early-time-point terpene treatment (*p*-value < 0.01, FDR < 0.05, D1 positive and/or D2 negative) (Table 1, Table S3).

3.4 | Detection of new putative transcripts in the *Grosmannia clavigera* genome

The *G. clavigera* genome is not well annotated. Thus, we extracted the reads that were not mapped on the known *G. clavigera* transcriptome, mapped them on the entire genome, and found genome fragments enriched by mapped reads using the BS segmentation algorithm of the T-BioInfo platform (Brodsky et al., 2010). Well-annotated fragments were used in the analysis as new putative transcripts in the genome (see Materials and Methods). These new transcripts (Table 1) may indicate additional putative regulatory mechanisms used in *G. clavigera* growing on terpenes.

The PCA of expression values of the new putative transcripts' expressions (normalized by corresponding treatment controls; NCT data; see Materials and Methods) revealed the same grouping of samples as was seen before with known transcripts: meaningful separation of samples into three groups on the PC1–PC2 plane (covering 86.6% of total data variability). As seen previously, these groups were as follows: early-time-point terpene treatment (samples with terpene treatment, 12-h growth: wild type and mutant), late-time-point terpene treatment (sample with terpene treatment, 7-days growth) and other (samples with OA and OO, 5-days growth) (Fig. S1A). Similar to the procedure of plotting on the PC1–PC2 plane for known transcripts, the same diagonals, D1 and D2 directions, were selected to separate clusters. The early- and late-time-point terpene treatment groups were separated along the D1 direction, and along the D2 direction, all terpene samples were separated from all samples without terpene treatment (Fig. S1B). As seen previously, selection of the significantly differentiated new transcripts was based on projection of each new transcript-profile onto a PC1–PC2 plane and separating along the D1 and D2 directions.

Among new transcripts upregulated under terpene treatment (mostly in early-time-point terpene treatments), we found transcripts annotated as retrotransposons, pectinesterase and cytochrome c oxidase copper chaperone. In previous studies (Chadha & Sharma, 2014; Raman et al., 2013; Santana & Queiroz, 2015; Stanley, Fraser, Stanley, & Chambers, 2010), retrotransposons were mentioned as being involved in genome rearrangement, and their role in stress adaptation in various fungal species was described. In addition, in plants, it is well known that transposable elements, specifically retrotransposons, are activated by biotic and abiotic stress (Belyayev et al., 2010; Hosid, Brodsky, Kalendar, Raskina, & Belyayev, 2012; Raskina, Brodsky, & Belyayev, 2011). Thus, we can hypothesize that

TABLE 1 Genes of *Grosmannia clavigera* with remarkable response to monoterpenes

Gene	Description	Ln(Fold Change) 12ht[a]	*p*-value	Ln(Fold Change) 36ht[a]	*p*-value	Ln(Fold Change 12ht[b]	*p*-value	Ln(Fold Change) 7dt[b]	*p*-value
Putative transporter genes									
CMQ_1821	major facilitator superfamily transporter monosaccharide	3.81*	0	2.87*	0	0.93*	2.22E-15	−3.2*	2.37E-12
CMQ_6705	amino acid permease	1.58*	1.04E-05	−0.03	0.76	2.76*	1.5E-07	1.06*	0.001
CMQ_746	major facilitator superfamily transporter multidrug resistance	−0.69*	1.50E-03	0.69*	1.11E-05	−1.47*	0.004	1.52*	0
CMQ_5680	major facilitator superfamily transporter	0.71	0.26	2.91*	3.11E-15	−0.80	0.11	3.60*	0
CMQ_2588	major facilitator superfamily transporter multidrug resistance	−0.36	0.51	3.89*	3.55E-15	−1.21	0.06	4.25*	0
CMQ_1489	major facilitator superfamily transporter sugar	0.95	0.14	2.51*	1.78E-15	−0.21	0.32	5.04*	0
CMQ_6042	major facilitator superfamily transporter	0.07	0.91	1.65*	4.49E-14	−1.09*	0.0001	2.88*	0
CMQ_3736	maltose permease	−0.13	0.29	2.07*	0	−1.05	0.07	2.57*	5.88E-15
Putative genes involved in terpene modification or degradation									
CMQ_7160	FAD binding domain protein	3.96*	1.95E-14	3.67*	0	2.36*	0	0.90*	4.7E-09
CMQ_764	zinc-binding alcohol	2.29*	6.66E-16	2.11*	8.88E-16	1.25*	0.00,016	1.36*	9.38E-05
CMQ_4724	acetyl-coenzyme a synthetase	0.23	0.23	0.99*	0	−1.24*	0	1.42*	0
CMQ_4999	citrate synthase	−0.08	0.27	2.66*	3.11E-14	0.68	0.13	4.24*	0
CMQ_8300	2-methylcitrate dehydratase	−0.15	0.01	1.86*	2.66E-15	−0.73*	0.00,085	1.92*	0
CMQ_1329	Fungal transcriptional regulatory protein	0.08	0.79	0.41*	3.8E-04	−2.17*	7.11E-15	0.75*	1.07E-07
CMQ_6445	c2 h2 transcription factor	1.09*	5.94E-05	−0.51	0.16	3.37*	0	0.13	0.19
CMQ_4220	fungal specific transcription factor domain protein	1.23*	2.69E-04	1.16*	2.62E-05	3.67*	0	2.58*	0
Putative gene involved in proteins modification and phosphorylation									
CMQ_523	oxidoreduction	0.75*	5.12E-03	0.11	0.27	3.14*	0	−0.28	0.25
Putative genes involved in mycelium development									
CMQ_7872	zinc finger protein zpr1	1.14*	3.55E-15	0.02	0.2	3.35*	0	1.89*	0
CMQ_6292	bax inhibitor family protein	0.64*	2.42E-07	−0.23	0.34	2.12*	0	0.58*	0
Putative genes involved in stress response									
CMQ_4400	c2 h2 transcription factor	0.97*	2.52E-06	−0.65	0.05	4.12*	0	3.09*	0
CMQ_5857	heat shock protein chaperonin	1.52*	7.74E-08	0.27	0.23	3.50*	0	1.98*	0
CMQ_812	bzip transcription factor idi-4	1.21*	0	−1.18*	0	1.35*	0	−0.82*	5.27E-06
CMQ_7087	heat shock protein hsp98 hsp104	0.66*	9.05E-04	0.10	0.3	3.05*	0	1.67*	0

(Continues)

TABLE 1 (Continued)

Gene	Description	Ln(Fold Change) 12ht[a]	*p*-value	Ln(Fold Change) 36ht[a]	*p*-value	Ln(Fold Change 12ht[b]	*p*-value	Ln(Fold Change) 7dt[b]	*p*-value
New regions									
GL629738.1:1506-2246	related to pectinesterase					4.44*	0	0.97	0
GL629765.1:1,073,751-1,074,061	cytochrome c oxidase copper chaperone					2.97*	0	−1.37*	0
GL629740.1:1551-2295	retrotransposon nucleocapsid protein					3.69*	0	2.71*	0
GL629772.1:198-996	retrotransposon nucleocapsid protein					3.58*	0	1.73*	5.27E-06
GL629735.1:73,020-75,666	gag polymerase env					3.10*	0	−1.59*	0
GL629729.1:1,830,212-1,830,956	HET-R					−2.05*	0	−0.01	0.55
GL629766.1:2384-2611	HET-R					−2.08*	0	0.76*	0
GL629974.1:2183-2481	HNWD2 protein vegetative incompatibility HET-E					−1.77*	3.04E-12	0.37*	1.68E-08
GL630006.1:2,227,792-2,229,106	Vegetative incompatibility protein HET-E-1					−2.15*	0	−0.74	0
GL629765.1:1,513,650-1,514,179	ribonuclease P					−0.85*	3.22E-13	−1.3*	0
GL630006.1:1,366,882-1,367,567	integral membrane protein					0	0.32	9.99*	0
GL630006.1:1,367,626-1,368,204	integral membrane protein					0	0.32	10.26	0

[a]DiGuistini et al. data [1]
[b]GEO accession number GSE43006 Data
*p-value < .01, FDR < 0.05 (Kal's Z-test)

activation of these putative retrotransposons may play a role in the adaptation of *G. clavigera* to terpene treatment, mainly in the early stages.

Pectinesterase is essential for making pectin more susceptible to enzymes of pathogens (Lionetti, Cervone, & Bellincampi, 2012). Here, we showed the upregulation of new transcripts with putative "pectinesterase" annotation under *G. clavigera* growing on the terpene media at early time-points. As early-time-point terpene media simulates growth of the fungus during early stages of tree infection under terpene attack, activation of this enzyme may have a role in the virulence of the fungus.

In late-time-point terpene treatment, we found that the upregulation of two transcripts was annotated as integral membrane proteins. These two transcripts were located close to each other on the genome, and also their expression levels were similar. Thus, they might be two exons of the same gene. In addition, these two transcripts have high-sequence similarity to a known integral membrane protein (*CMQ_8029*) of *G. clavigera*, also with high-upregulation mainly in response to late-time terpene treatment (Table S3) but found elsewhere in the genome. Therefore, these two transcripts probably represent another integral membrane protein of *G. clavigera* that may have a regulatory effect during late-time-point terpene treatment.

Besides the above-mentioned newly detected transcripts, we found additional transcripts annotated as HET-R, HET-E and ribonuclease that were downregulated in response to terpene treatment (mostly in the early-time-point terpene treatments) (Figure 5, Table S1).

Previously, it was suggested that *het* genes activate a cell-death pathway (Paoletti & Clavé, 2007; Saupe, Clavé, & Bégueret, 2000). Thus, repression of these transcripts may indicate repression of the cell-death pathway under early-time-point terpene treatment, enabling the fungus to survive in the presence of terpenes in the medium. Previously, it was mentioned that terpenoid treatment induces transcriptome reprogramming (DiGuistini et al., 2011). As HET-R is a serine-type protease (Dong, Huang, Wu, & Zhao, 2000), repression of this transcript, as well as repression of ribonuclease, may indicate repression of ribonuclease and protease activity.

3.5 | Individual genes as classifiers of each group of samples according to their growth conditions

The samples of both sets of experiments were grouped into three biological classes: class 0 – no-terpene treatment; class 1 – early terpene treatment; and class 2 – late terpene treatment. As the biological classes of samples match clusters of samples on the PCA plane (Figure 2e), the normal distribution based on Naïve Bayesian Classifier techniques was applied for determining genes that are classifiers for all classes according to their log-expression values in samples of the class and in determining gene-discriminators that separate one class from all other classes (see Materials & Methods). As a result, gene-classifiers and gene-discriminators were found for classes of samples associated with growth conditions and therefore with *G. clavigera*'s response to terpene intoxication as well as with using terpenes as the carbon source (Table S4). With high probability, the gene-classifier allows the assigning of a sample to its class (no-terpene treatment, early terpene treatment, or late terpene treatment) based on this gene's log-expression level. For a gene-discriminator of a class, its expression in a sample of this class allows the association of the sample to this specific class. However, expressions in samples of other classes reveal that the sample is not from the class of interest and does not discriminate between other classes.

Class 1 early-time terpene treatment is characterized by detoxification and stress-response processes as well as the onset of metabolic processes. We found one gene-discriminator for this class of samples, *CMQ_5191*. This gene is annotated as *amine oxidase flavin-containing superfamily protein* and may participate in detoxification (Figure 6b). A number of genes were also found to be putative discriminators of class 1; however, they are less significant Bayesian log-odds of the class than *CMQ_5191*. These genes, *CMQ_2132*, *CMQ_3613*, *CMQ_5935*, *CMQ_6342*, *CMQ_4038*, *CMQ_2490*, *CMQ_4617*, *CMQ_4844* and *CMQ_643*, were annotated as *efflux* and *ABC transporters* and were discussed in previous sections of the manuscript (see section: "Linkage between genes participating in the same biological processes"). This group of genes is associated with stress response. The gene CMQ_3264 (annotated as *3,4-dihydroxy-2-butanone 4-phosphate synthase*), also a discriminator of this class, is associated with the onset of metabolic processes.

Class 2 late-time terpene treatment is characterized by the expression of gene-discriminators involved in organic decomposition processes, in particular, the decomposition of terpenes for utilization by the fungus as a carbon source. Indeed, as a rule, fungi decompose organic compounds by introducing a hydroxyl group (Mikami, 1988) and

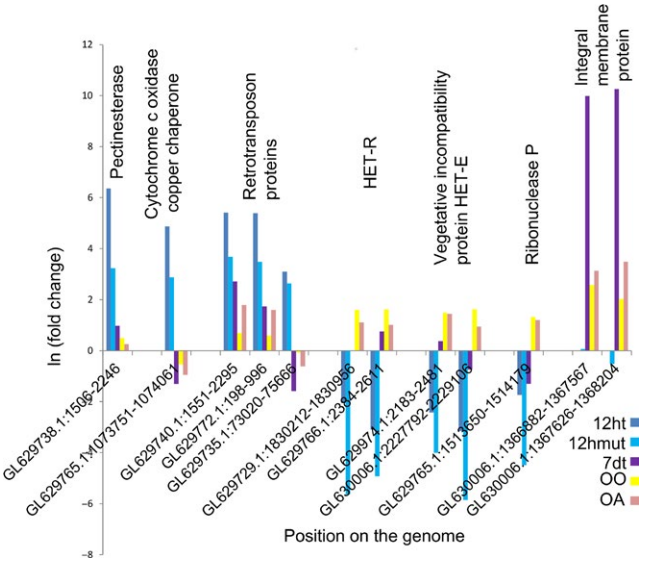

FIGURE 5 New putative transcripts significantly differentially expressed between different growth states of *Grosmannia clavigera*. Bar height represents the natural logarithm of a fold change in transcript expression under each treatment condition versus control. The name of the transcript indicates its location in the genome sequence. (tiff)

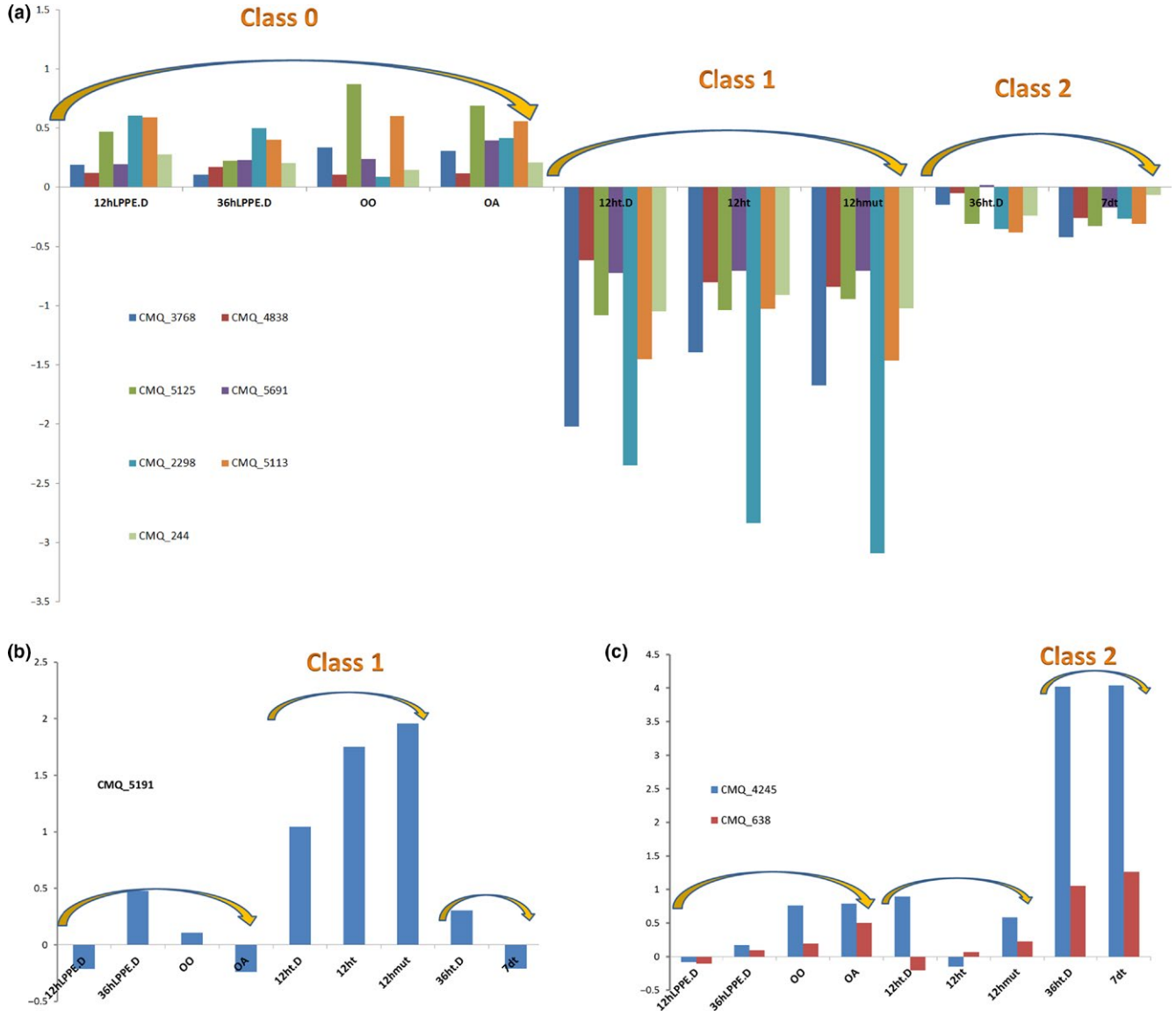

FIGURE 6 Log-expression levels of genes determined as classifiers and discriminators by the Naïve Bayesian Classifier techniques. For all three biological classes, the gene-classifier allows to assigning of a sample to its class (no-terpene treatment, early-terpene treatment, or late-terpene treatment) with high probability based on the gene's log-expression level. A gene-discriminator of a biological class separates samples of this class from samples of two other classes but cannot discriminate between two alternative classes. (a) Gene-classifiers of all three classes of samples. (b) Gene-Discriminator of class 1. (c) Two gene-discriminators of class 2. (tiff)

transform primary alcohol to carboxylic acid via aldehyde and other organic compounds. Several terpene treatment differentiating genes are connected to specific processes of redox: niacin or "niacin equivalents" can ultimately be converted to nicotinamide adenine dinucleotide (NAD), which exists in both oxidized (NAD+) and reduced (NADH) forms as well as the phosphorylated form of NAD+ (NADP+) and also NADP+ reduced to NADPH. The most significant gene-discriminator of this class, *CMQ_4245*, is annotated as major facilitator superfamily nicotinic acid (niacin) transporter. This transporter is the key part of cofactor in the above redox process. Another gene-discriminator, *CMQ_638* (electron transfer flavoprotein-ubiquinone oxidoreductase), is an enzyme responsible for the cleavage of carbon rings—the most stable terpene substructures (Petasch et al., 2014).

CMQ_6937 that is encoding mitochondrial aldehyde dehydrogenase is also a gene-discriminator of class 2. Aldehyde dehydrogenases are important enzymes in limonene and pinene degradation pathway. It is known that these enzymes are highly expressed in several organisms, which are feeding on a monoterpene-rich diet (Marmulla & Harder, 2014).The normalized log-expressions of these two discriminators of class 2 are depicted on Figure 6c.

A number of genes were found to be classifiers for **all three classes** simultaneously: it means that according to a log-expression of the gene in any sample of the set, this sample will be associated with one particular class (Figure 6a).

Four C6 zinc finger domain containing proteins (*CMQ_3768*, *CMQ_4838*, *CMQ_5125* and *CMQ_5691*) demonstrate the same

classifier pattern of the normalized log-expression. As four copies of this protein are very significant three-class classifiers, this type of protein is a "robust classifier": one can expect correct classification for a sample in a new experiment based on normalized log-expression values of these proteins.

Two monooxydases (*CMQ_2298*, *CMQ_5113*), known to participate in the *oxidation–reduction process*, are also good gene-classifiers. Similar to the C6 zinc finger containing domain proteins, these monooxydases are significantly downregulated regarding corresponding control in the early-time-point of terpene treatment, weakly downregulated in the late terpene time-point and weakly upregulated in no-terpene media.

4 | CONCLUSIONS

By performing RNA-seq analysis of raw NGS data from *G. clavigera* under different growth conditions and then integrating the results with the RNA-seq processed data set from similar experiments, our discovery of transcriptome regulatory elements was supported by two independent experimental studies, increasing the robustness of the new findings. Our results support previous findings and also reveal new transcriptome regulatory mechanisms and processes that may be important in the adaptation of *G. clavigera* to terpenes. In addition, an analysis of newly detected putative transcripts in the *G. clavigera* genome revealed additional suggested regulatory mechanisms that may play an important role in the tolerance of the fungus to terpenes.

Overall, it seems that under stress caused by terpenes in the 12-h terpene treatment, there was a lag phase in *G. clavigera* growth and induction of transcriptome reprogramming (followed also by retrotransposon activity). This led to activation of stress-response pathways, upregulation of transporters (removing terpenes from the cells) and initiation of terpene detoxification through activation of genes that modify terpenes or cause degradation of terpenes. Later, under terpene treatment growth, there was an increase in terpene detoxification, followed by an increase in expression of genes related to terpene modification and degradation. During the next growth stage, under 4-days terpene treatment, there was an adaptation of the fungus growing on the terpene media, followed by a reduction in the expression of the terpene modification and degradation genes that had increased in response to the 36-h treatment. Finally, in the 7-days treatment on minimal media, under stress mainly caused by starvation, once again there was an increase in expression levels of the terpene modification and degradation genes that led to the utilization of terpenes as an energy source.

Annotations of gene-discriminators for the early- and late-terpene groups corroborate with the above hypotheses about major processes in *Grosmannia clavigera* under these biological conditions.

In summary, these findings may shed light on the mechanisms of tolerance of the *G. clavigera* pathogen to terpenoids synthesized by the conifer hosts of this fungus.

ACKNOWLEDGEMENTS

We thank Dina Mukha for helpful discussions about the biological interpretation of results. We are very thankful to Dr. Claudia Copeland for the English editing of this manuscript. We thank Dr. Maxim Krakhmalnyi for help with the *G. clavigera* fungus taxonomy and thank Elia Brodsky for helpful discussions. We are thankful to all members of the Tauber Bioinformatics Research Center for their support and valuable comments. We also appreciate Dr. Breuil and the team of the Department of Wood Science, University of British Columbia, Vancouver, Canada for help in accessing digital data analysed in this project.

REFERENCES

Alamouti, S. M., Wang, V., Diguistini, S., Six, D. L., Bohlmann, J., Hamelin, R. C., … Breuil, C. (2011). Gene genealogies reveal cryptic species and host preferences for the pine fungal pathogen *Grosmannia clavigera*. *Molecular Ecology*, *20*, 2581–2602.

Belyayev, A., Kalendar, R., Brodsky, L., Nevo, E., Schulman, A. H., & Raskina, O. (2010). Transposable elements in a marginal plant population: Temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mobile DNA*, *1*, 6.

Bohlmann, J. (2012). Pine terpenoid defences in the mountain pine beetle epidemic and in other conifer pest interactions: Specialized enemies are eating holes into a diverse, dynamic and durable defence system. *Tree Physiology*, *32*, 943–945.

Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, *19*, 185–193.

Brodsky, L., Kogan, S., Benjacob, E., & Nevo, E. (2010). A binary search approach to whole-genome data analysis. *Proceedings of the National Academy of Sciencesof the United States of America*, *107*, 16893–16898.

Chadha, S., & Sharma, M. (2014). Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen *Magnaporthe oryzae*. *PLoS ONE*, *9*, e94415.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *vol. 2008*, 619832.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676.

Diguistini, S., Liao, N. Y., Platt, D., Robertson, G., Seidel, M., Chan, S. K., … Jones, S.J.M. (2009). De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology*, *10*, R94.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., … Breuil, C. (2011). Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceedings of the National Academy of Sciencesof the United States of America*, *108*, 2504–2509.

Dong, Y., Huang, X., Wu, X. Y., & Zhao, J. (2000). Identification of the active site of HetR protease and its requirement for heterocyst differentiation in the cyanobacterium Anabaena sp. strain PCC 7120. *Journal of Bacteriology*, *182*, 1575–1579.

Douglass, S., Hsu, S.-W., Cokus, S., Goldberg, R. B., Harada, J. J., & Pellegrini, M. (2016). A naïve Bayesian classifier for identifying plant microRNAs. *The Plant Journal*, *86*, 481–492.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*, 2nd ed.. New York, NY: John Wiley & Sons.

Farooq, A., Atta-ur-Rahman, & Choudhary, M.I. (2004). Fungal transformation of monoterpenes. *Current Organic Chemistry*, *8*, 353–367.

Franceschi, V. R., Krokene, P., Christiansen, E., & Krekling, T. (2005). Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist*, *167*(2), 353–375.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*, 3420–3435.

Haridas, S., Wang, Y., Lim, L., Massoumi Alamouti, S., Jackman, S., Docking, R., … Breuil, C. (2013). The genome and transcriptome of the pine saprophyte *Ophiostoma piceae*, and a comparison with the bark beetle-associated pine pathogen *Grosmannia clavigera*. *BMC Genomics*, *14*, 373.

Hosid, E., Brodsky, L., Kalendar, R., Raskina, O., & Belyayev, A. (2012). Diversity of long terminal repeat retrotransposon genome distribution in natural populations of the wild diploid wheat *Aegilops speltoides*. *Genetics*, *190*(1), 263–274.

Kal, A. J., van Zonneveld, A. J., Benes, V., van den Berg, M., Koerkamp, M. G., Albermann, K., … Tabak, H.F. (1999). Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Molecular Biology of the Cell*, *10*(6), 1859–1872.

Kumar, R., & Tamuli, R. (2014). Calcium/calmodulin-dependent kinases are involved in growth, thermotolerance, oxidative stress survival, and fertility in *Neurospora crassa*. *Archives of Microbiology*, *196*, 295–305.

Kuravi, V. K., Kurischko, C., Puri, M., & Luca, F. C. (2011). Cbk1 kinase and Bck2 control MAP kinase activation and inactivation during heat shock. *Molecular Biology of the Cell*, *22*, 4892–4907.

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., … Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, *452*, 987–990.

Lah, L., Haridas, S., Bohlmann, J., & Breuil, C. (2013). The cytochromes P450 of *Grosmannia clavigera*: Genome organization, phylogeny, and expression in response to pine host chemicals. *Fungal Genetics and Biology*, *50*, 72–81.

Li, B., & Dewey, C.N (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Lionetti, V., Cervone, F., & Bellincampi, D. (2012). Methyl esterification of pectin plays a role during plant-pathogen interactions and affects plant resistance to diseases. *Journal of Plant Physiology*, *169*, 1623–1630.

Marmulla, R., & Harder, J. (2014). Microbial monoterpene transformations-a review. *Frontiers in Microbiology*, *5*, 346.

Mikami, Y. (1988). Microbial conversion of terpenoids. *Biotechnology and Genetic Engineering Reviews*, *6*, 271–320.

Novak, M., Lah, L., Šala, M., Stojan, J., Bohlmann, J., & Komel, R. (2015). Oleic Acid Metabolism via a Conserved Cytochrome P450 System-Mediated ω-Hydroxylation in the Bark Beetle-Associated Fungus *Grosmannia clavigera*. *PLoS ONE*, *10*, e0120119.

Paoletti, M., & Clavé, C. (2007). The fungus-specific HET domain mediates programmed cell death in *Podospora anserina*. *Eukaryotic Cell*, *6*, 2001–2008.

Petasch, J., Disch, E.-M., Markert, S., Becher, D., Schweder, T., Hüttel, B., … Harder, J. (2014). The oxygen-independent metabolism of cyclic monoterpenes in *Castellaniella defragrans* 65Phen. *BMC Microbiology*, *14*, 164. https://doi.org/10.1186/1471-2180-14-164

Phillips, M., & Croteau, R. (1999). Resin-based defenses in conifers. *Trends in Plant Science*, *4*, 184–190.

Raman, V., Simon, S. A., Romag, A., Demirci, F., Mathioni, S. M., Zhai, J., … Donofrio, N.M. (2013). Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, *Magnaporthe oryzae*. *BMC Genomics*, *14*, 326.

Raskina, O., Brodsky, L., & Belyayev, A. (2011). Tandem repeats on an eco-geographical scale: Outcomes from the genome of *Aegilops speltoides*. *Chromosome Research*, *19*, 607–623.

Santana, M., & Queiroz, M. (2015). Transposable elements in fungi: A genomic approach. *Scientific Journal of Genetics and Gene Therapy*, *1*, 012–016.

Saupe, S. J., Clavé, C., & Bégueret, J. (2000). Vegetative incompatibility in filamentous fungi: *Podospora* and *Neurospora* provide some clues. *Current Opinion in Microbiology*, *3*, 608–612.

Six, D. L., & Wingfield, M. J. (2011). The role of phytopathogenicity in bark beetle–fungus symbioses: A challenge to the classic paradigm. *Annual Review of Entomology*, *56*, 255–272.

Stanley, D., Fraser, S., Stanley, G. A., & Chambers, P. J. (2010). Retrotransposon expression in ethanol-stressed *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, *87*, 1447–1454.

Wang, Y., Lim, L., DiGuistini, S., Robertson, G., Bohlmann, J., & Breuil, C. (2013). A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist*, *3*(197), 886–898.

Wang, Y., Lim, L., Lina, M., Ljerka, L., Joerg, B., & Colette, B. (2014). Gene discovery for enzymes involved in limonene modification or utilization by the mountain pine beetle-associated pathogen *Grosmannia clavigera*. *Applied and Environmental Microbiology*, *80*, 4566–4576.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Kligun E, Ostretsov B, Titievsky A, Farkov M, Alamouti SM, Brodsky L. Adaptation of the pine fungal pathogen *Grosmannia clavigera* to monoterpenes: Biochemical mechanisms revealed by RNA-seq analysis. *For Path*. 2017;47:e12372. https://doi.org/10.1111/efp.12372