

Investigation of the Potential Mechanisms of Limonene Resistance in *Meyerozyma Caribbica* by Comparative Genomics

Filipe Augusto Teixeira

Universidade Federal de Viçosa

Gustavo Molina

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Fernanda Matias Albuini

Universidade Federal de Viçosa

Ananda Aguilar

Universidade Federal de Viçosa

Patrick Neves Squizato

Institute of Biosynthetic and Fiber Innovation

Tiago Antônio de Oliveira Mendes

tiagoaomendes@ufv.br

Universidade Federal de Viçosa

Research Article

Keywords: Limonene, *Meyerozyma caribbica*, comparative genomics, MFS transporter, molecular docking, bioconversion

Posted Date: February 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8732531/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Terpene bioconversion presents a sustainable alternative to chemical synthesis for producing high-value compounds, yet its application is constrained by the inherent toxicity of terpenes to microbial biocatalysts. Bioprospecting offers a cost-effective strategy to overcome this limitation and through this approach a strain of *Meyerozyma caribbica*, with a profile of resistance to the monoterpene limonene, was isolated from *Solanum lycocarpum*. This study aimed to elucidate the molecular mechanisms underlying this resistance through comparative genomics. These *in silico* analyses against five food-industry relevant yeasts (*Saccharomyces cerevisiae*, *Kluyveromyces marxianus*, *Spathaspora passalidarum*, *S. arborariae*, and *Scheffersomyces stipitis*) identified 58 orthologs gene clusters and highlighted 26 proteins belonging to the Major Facilitator Superfamily (MFS) transporter in *M. caribbica*. Homology modeling and molecular docking analyses demonstrated a satisfactory binding affinity between limonene and the generated MFS protein models, supporting their potential role as specific efflux pumps. Quantitative PCR analysis identified a specific clade within the MFS transporter superfamily that was significantly upregulated under limonene stress. However, experimental validation revealed that intracellular limonene concentrations in *M. caribbica* following exposure to 4% (v/v) limonene were not significantly lower than those measured in susceptible yeast strains. The confluence of *in silico* and experimental evidence strongly indicates that the limonene resistance in *M. caribbica* is mediated by the active efflux of the compound, facilitated by specific MFS transporters. These findings establish *M. caribbica* as a promising biological chassis for the industrial bioconversion of limonene into value-added products.

1. Introduction

The production of aroma compounds is traditionally achieved through natural extraction or chemical synthesis, both of which are typically associated with limiting factors such as low productivity, dependence on climatic conditions, extreme process requirements, and the generation of racemic mixtures[1]. Biotechnological processes offer a viable alternative for the production of aroma compounds, as they circumvent the need for seasonal raw materials, operate under milder conditions, exhibit high specificity, and can be considered environmentally sustainable[2].

Among the substrates extensively studied and employed in bioprocesses, monoterpenes hold particular significance[3]. In this group, limonene (empirical formula: $C_{10}H_{16}$) stands out due to its industrial relevance, attributed to its widespread availability, low cost, and role as a precursor for various high-value-added compounds[1]. However, terpene bioconversion is constrained by several factors, particularly the inherent toxicity of these compounds to microorganisms when used as substrates in biotechnological processes[2].

A potential strategy to overcome the challenges associated with the physicochemical properties of limonene in bioconversion processes is the use of robust microorganisms exhibiting resistance phenotypes to terpene toxicity and other organic solvents[4]. Through bioprospecting efforts, a strain of

Meyerozyma caribbica displaying limonene resistance was isolated from a plant-derived source and identified via genetic markers by the research group at the Food Biotechnology Laboratory of the Federal University of Vales do Jequitinhonha and Mucuri, in collaboration with the Synthetic Biology and Biological Systems Modeling Laboratory at the Federal University of Viçosa.

The yeast *M. caribbica* demonstrates significant biotechnological potential, as recent studies have shown that isolates of this species are capable of bioremediating aqueous media containing manganese ions[5]. Furthermore, it has been identified as a promising candidate for ethanol production from various carbohydrate sources[6].

The increasing availability of genomic data and bioinformatics tools enables the discovery of novel biological insights and facilitates the establishment of correlations between organisms and their phenotypic traits[7]. In this context, the present study aims to investigate the potential biological mechanisms underlying the limonene resistance phenotype observed in *M. caribbica* through comparative genomics.

2. Material and methods

2.1 Microorganisms and culture medium

The *Meyerozyma caribbica* yeast strain used in this study was isolated by the research group at the Food Biotechnology Laboratory (LBA) of the Institute of Science and Technology at the Federal University of Vales do Jequitinhonha and Mucuri (ICT-UFVJM) from the fruit of *Solanum lycocarpum*.

The *Saccharomyces cerevisiae*, *Kluyveromyces marxianus*, *Spathaspora passalidarum*, *Spathaspora arborariae*, and *Scheffersomyces stipitis* strains were kindly provided by Dr. Luciano Gomes Fietto from the Department of Biochemistry and Molecular Biology at the Federal University of Viçosa.

For cultivation experiments, an adapted Yeast Malt (YM) medium was used, consisting of bacteriological peptone (0.5%), glucose (1%), yeast extract (0.3%), and distilled water. When a solid medium was required, 2% bacteriological agar was added to the formulation.

2.2 Limonene resistance assay

The cultivation strategies employed in this study were based on the methodology used in the isolation and selection of *Meyerozyma caribbica* by the LBA research group, as described by Bicas & Pastore[8]. Microorganisms were activated in YM medium on Petri dishes for 48 hours at 30°C in a bacteriological incubator (Thelga, TE64CB).

Following the activation period, a bacteriological loop was used to transfer the isolates to a tube of 16 × 100 mm (13.5 mL) containing 6 mL of sterile YM medium. The tubes were incubated in an orbital shaker (Tecnal, model TE-421) at 30°C with agitation at 150 rpm for 48 hours.

Following cultivation, microbial growth rates were indirectly assessed by measuring absorbance at 600 nm (UA) using 200 µL aliquots of culture medium containing biomass in a digital spectrophotometer (Versa Max Microplate Reader, Molecular Devices). The absorbance values obtained were used to normalize the initial inoculum for yeast cultivation in the presence of limonene.

Finally, microorganisms were cultivated in liquid YM medium in 13.5 mL test tubes containing 0, 2, and 4% limonene ((R)-(+)-limonene (Sigma-Aldrich, ≥ 93% purity)) (v/v), with the initial inoculum adjusted to 0.6 UA. Cultures were maintained at 30°C with agitation at 150 rpm in an orbital incubator. Absorbance measurements were taken every 24 hours for up to 72 hours. All assays were performed in technical triplicates.

2.3 Comparative genomics

The genomes used in this study are available in the National Center for Biotechnology Information (NCBI) database, and the accession numbers for each assembly are listed in Table 1. In general, the comparative genomics methodology can be summarized in three main steps: (i) prediction of all genes from the assembly of a reference genome for each microorganism, (ii) identification and clustering of orthologous genes, and (iii) annotation of the potential proteins encoded by the genes of interest. These steps were conducted in a Linux virtual environment, and both the software and the resulting analysis data were hosted on the Júpiter server of the Information Technology Directorate at UFV.

Table 1
Species and NCBI accession numbers of the
genomes analyzed.

Microorganism	Accession Number
<i>Saccharomyces cerevisiae</i>	GCF_000146045.2
<i>Kluyveromyces marxianus</i>	GCF_001417885.1
<i>Spathaspora passalidarum</i>	GCF_000223485.1
<i>Spathaspora arborariae</i>	GCA_000497715.1
<i>Scheffersomyces stipitis</i>	GCF_000209165.1
<i>Meyerozyma caribbica</i>	GCA_000755205.1

Gene prediction for each selected genome was performed using Augustus software (version 3.2.2, available at: <http://augustus.gobics.de/>)[9], with parameters set for protein and coding sequence (CDS) retrieval. Orthologous gene clustering followed the methodology described by Mendes et al.[10] and was conducted using OrthoMCL[11].

Subsequently, annotation of the clustered genes classified as orthologous to *M. caribbica* was performed using InterProScan 5.30[12], which predicts protein domains and families based on the CDS obtained from Augustus, correlating them with the InterPro database[13].

2.4 Sample preparation for cell viability test and volatile compound extraction

Initially, the yeasts *S. cerevisiae*, *K. marxianus*, *S. passalidarum*, and *M. caribbica* were cultivated in 13.5 mL test tubes containing approximately 5.8 mL of YM medium at 30°C with orbital shaking at 150 rpm until the cultures reached an optical density of 1 AU. A 300 µL aliquot of each culture was reserved for cell viability testing through serial dilution and spot plating. Subsequently, 240 µL of limonene (final concentration: 4% (v/v)) was added to each tube, followed by vortexing for 1 minute. The tubes were then returned to the incubator at 30°C with orbital shaking at 150 rpm. After 15 minutes, another 300 µL aliquot of each culture was collected for a second cell viability test.

Two 600 µL aliquots of each culture exposed to limonene were reserved for volatile compound extraction in an organic solvent, adapting the methodology described by Molina et al.[14].

2.5 Cell viability test

The cell viability assay was conducted at two points, before and after exposure to 4% (v/v) limonene for 15 minutes. Serial dilutions 1:10, 1:100, 1:1,000, 1:10,000, and 1:20,000 were prepared from the 300 µL aliquots and spotted (10 µL per spot) onto petri dishes containing solid YM medium. The plates were incubated at 30°C for 48 hours and photographed using a gel documentation system (UVB Transilluminator LTB-20X20HE, Loccus do Brasil, Cotia, São Paulo, Brazil).

2.6 Volatile compound extraction

For volatile compound extraction, 600 µL of culture medium exposed to limonene for 15 minutes was transferred to 1.5 mL microcentrifuge tubes and centrifuged at 3,000 rpm for 5 minutes (Nova Instrument, NI1802). The supernatant was separated from the precipitate into new 1.5 mL tubes. Ethyl acetate was used as the solvent, added at a 1:1 ratio to the initial sample volume (600 µL) in each tube (supernatant and precipitate), followed by vortexing for 1 minute. A second centrifugation was performed under the same conditions, and the organic phase (clear layer) was transferred to new 1.5 mL tubes containing 100 mg of anhydrous sodium sulfate. This process (solvent addition, vortexing, centrifugation, and organic phase transfer) was repeated three times. The final volume of the extracted sample in ethyl acetate was approximately 1.8 mL.

A calibration curve was generated to correlate limonene peak areas with biological sample concentrations, using eight reference points (0, 0.1, 0.2, 0.5, 1, 2, 3, and 4 µL of limonene per mL of ethyl acetate).

2.7 Limonene quantification by gas chromatography-mass spectrometry (GC-MS)

Limonene quantification was performed using a gas chromatograph coupled with a mass spectrometer (SHIMADZU, GCMS-QP2010 Ultra) equipped with a Restek Rtx-5MS fused silica column (30 m length,

0.25 mm internal diameter, 0.25 μm film thickness), composed of 5% diphenyl and 95% dimethyl polysiloxane. Helium was used as the carrier gas at a constant flow rate of 1.8 mL/min. The split-splitless injector was maintained at 220°C and operated in split mode with a split ratio of 40:1. The detector was set at a transfer line temperature of 240°C, an ionization energy of + 70 eV, and a mass range of 35–400 m/z. The gas chromatograph oven temperature was initially held at 40°C for 2 minutes, then increased to 200°C at a rate of 10°C/min, where it was maintained for an additional 5 minutes.

2.8 Gene expression quantification

To design primers targeting the Major Facilitator Superfamily (MFS) transporters in *M. caribbica*, we first identified nucleotide sequences of annotated transporter orthologs. These sequences were clustered and a phylogenetic tree was constructed using MEGA software[15]. Consensus sequences from the resulting clades were used to design 18 primer pairs, ensuring coverage of all 26 annotated MFS transporter genes. The nucleotide specificity of all primers was subsequently evaluated using Primer-BLAST[16].

M. caribbica cells used for RNA isolation were grown in YM medium overnight at 30°C and stirring at 120 rpm. Each inoculum was divided into three 125 mL Erlenmeyer flasks containing 20 mL of fresh medium, and the final monoterpene concentrations of 0 (control), 2 and 4% (v/v). The initial OD₆₀₀ was adjusted to 0.6. The flasks were incubated for 72 h under the conditions described above. The samples were collected at 0, 24 and 72 hours of stress exposure to perform the expression analysis. Experiments were conducted in three biological replicates.

The harvested cells were treated with Lyticase from *Arthrobacter luteus* (Sigma-Aldrich) for 1 hour at 30°C under gentle stirring. Total RNA was extracted using the RNeasy Mini kit (Qiagen). RNA concentration was quantified with the Qubit 3.0 Fluorometer (Invitrogen). Samples were treated with RNase-free DNase I (Promega M6101), and cDNA synthesis was performed with the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems 4374967) according to the following cycle: 10 min at 25°C, 120 min at 37°C, and 5 min at 85°C. Genomic DNA was extracted using the lithium acetate/single-stranded carrier DNA/polyethylene glycol (LiAc/ss-DNA/PEG)[17].

The qPCR analysis was performed using GoTaq qPCR Master Mix (Promega A6002) with StepOne™ Real-Time PCR System (Applied Biosystems). The assays were carried out in technical duplicate using the conditions as follows: 10 min at 95 °C and 40 cycles of 30 s at 95 °C, 1 min at 60 °C. Relative mRNA quantification was performed using the standard curve method, in which genomic DNA was serially diluted to obtain a standard curve for each target by plotting the average Ct versus the log10 of DNA concentrations from the control sample (1.56–100.0 ng. μL^{-1}). The results were normalized using actin as internal control.

2.9 Molecular docking

The nucleotide sequences of predicted proteins from clades B and I were selected for molecular docking analysis and translated *in silico* into their corresponding polypeptides using the ExPASy translate

tool[18], thereby generating the amino acid sequences employed in all subsequent analyses. As a first step in structural characterization, the predicted proteins were analyzed for the presence of signal peptides and transmembrane helices. Signal peptide prediction was carried out using the SignalP 5.0 server[19], configured for the "Eukarya" organism group with an extended output format. Concurrently, the prediction of transmembrane helices was conducted using the TMHMM 2.0 server[20], which provided an extensive output that included graphical representations of topology. Following this initial characterization, three-dimensional protein structures were modeled by homology using the Phyre2 server[21] in intensive mode with 20 references templates models, including the crystal structure of *E. coli* multidrug transporter MdfA (PDB accession number 4ZP0), and then subjected to energy minimization for structural refinement. This optimization step was performed using the YASARA energy minimization server[22] with the YASARA force field. Following refinement, the optimized structures underwent a rigorous quality assessment. This comprehensive evaluation included analysis of residue stereochemistry via Ramachandran plots, as well as validation of the three-dimensional models using the ProSA, ERRAT, and Verify 3D programs. For the docking simulations, the ligand structure of (+)-limonene was acquired from the LigandBox chemical database[23]. Finally, protein-ligand interactions were evaluated through molecular docking simulations performed with AutoDock4[24, 25] and AutoDock Vina[26] within the PyRx platform. This integrated environment enabled the streamlined preparation of both the protein and ligand for comprehensive binding analysis. All three-dimensional resultant models were visualized with PyMOL.

2.10 Statistical analysis

To assess and determine whether there were significant differences in limonene quantification values and gene expression quantification, an Analysis of Variance (ANOVA) with Bonferroni correction for multiple hypothesis testing was performed using GraphPad Prism software (version 8.0.1). A significance level of 5% ($p < 0.05$) was adopted.

3. Results

3.1 Comparative genomics

One of the most successful strategies used to investigate phenotypic characteristics of living organisms is comparative genomics[27, 28]. Through computational methods, it is possible to identify groups of genes from different organisms that share ancestry and biological functions, as well as similarities and differences at the genetic level. After predicting the genes and classifying them into orthologous groups based on the genomes listed in Table 1, a total of 58 orthologous gene clusters were found in the genome of *M. caribbica*; 39 for *S. cerevisiae*; 33 for *S. arborariae*; 30 for *S. passalidarum*; 25 for *K. marxianus* and none for *S. stipitis* (Fig. 1).

Seeking to understand which biological processes might be involved in the phenotypic resistance to limonene observed in *M. caribbica* and considering that this characteristic could be related to genes exclusively present in this microorganism, the annotation of the proteins within the 58 clusters unique to

this yeast was performed. The Fig. 1 displays the categorization of all these 153 proteins present in the 58 orthologous gene groups of *M. caribbica*. Seven categories of proteins were identified: cell adhesion (8), autophagy (3), enzymes (22, with 15 related to carbohydrates), transcription factors (10), leucine-rich repeat (24), redox proteins (16), and transporters (26), along with hypothetical proteins (44), which are of unknown function.

The category with the highest representation corresponded to transporters, comprising 26 proteins. All of these belong to the MFS, the largest family of secondary transporters, which is widely conserved across all forms of life[29]. This finding is particularly relevant, as efflux pumps are known to contribute to reduced toxicity and increased tolerance to organic solvents in microorganisms. The toxicity of limonene toward microbial cells is largely attributed to its lipophilic nature, which leads to its accumulation in the cell membrane, thereby compromising plasma membrane integrity and inhibiting specific cellular functions. Thus, the high number of transporter proteins identified in the *M. caribbica* genome suggests that resistance mechanisms in this microorganism may be associated with limonene transporters.

3.2 Limonene quantification

Based on the hypothesis that the resistance phenotype observed in *M. caribbica* may be associated with transporter proteins, intracellular limonene quantification was performed using gas chromatography and compared with three other yeasts (*S. cerevisiae*, *K. marxianus*, and *S. passalidarum*). Cell viability assays were conducted before and after exposure of the cultures to 4% (v/v) limonene for 15 minutes (Fig. 2). Prior to exposure, cell growth was observed for all yeast strains across all dilutions, confirming their viability. After exposure, microbial growth was detected only in *K. marxianus* and *M. caribbica*.

Statistical analysis of intracellular limonene quantification couldn't reach a significant difference between *M. caribbica* and the other yeast species, although the absolute intracellular limonene levels measured in this microorganism were consistently lower than those detected in the remaining yeast strains under investigation. While a bacteriostatic effect was observed for *K. marxianus*, *M. caribbica* demonstrated continued viability after the 15-minute limonene treatment.

3.3 Gene expression quantification

Nine clades were established (Fig. 3) based on the phylogenetic tree constructed from the protein sequence data of MFS-type transporters identified in *M. caribbica*. The clades were designated as A, B, C, D, E, F, G, H, and I.

The qPCR experiments were successfully performed for clades A, B, G, H, and I. The resulting data were normalized to establish relative expression across replicates and subsequently plotted after statistical analysis (Fig. 4). A statistically significant increase in gene expression was observed only for clade I when the yeast was cultivated in the presence of 4% (v/v) limonene for 72 hours.

3.4 Docking

The prediction of signal peptide presence yielded very low probabilities across all models, with values not exceeding 0.0046. Specifically, the model cluster5734_mc|g4624.t1 exhibited the highest probability of 0.0046, while all remaining models displayed a probability of 0.0005. Analysis of transmembrane helix topology revealed a degree of variation among the models. The cluster4139_mc|g4970.t1 model was predicted to possess 5 transmembrane helices, cluster5734_mc|g4624.t1 was predicted to have 12, and the remaining models consistently featured a predicted count of 10 transmembrane helices. Of the optimized models, all exhibited Ramachandran plot quality parameters exceeding 95% for the combined allowed and favored regions. Regarding the ProSA quality parameters, the models yielded z-scores proximate to established quality benchmarks, particularly models cluster4139_mc|4968.t1 (-4.53) and cluster4139_mc|4969.t1 (-3.99). In the ERRAT analysis, model cluster4139_mc|4971.t1 was the sole construct displaying an overall quality factor below 90%, whereas model cluster4139_mc|4968.t1 achieved a score of 95.693%. Conversely, for the Verify 3D assessment, none of the models attained the recommended threshold (3D-1D score ≥ 0.2) indicative of high quality. The highest performance in this category was demonstrated by model cluster5734_mc|g4624.t1, which reached 63.18%, a value substantially below the 80% benchmark associated with satisfactory structural integrity. Homology modeling yielded six three-dimensional protein structures, which exhibited an overall architecture highly similar to one of the templates models, the crystal structure of *E. coli* multidrug transporter MdfA. A structural superposition of the resulting models is presented in Fig. 7, with alpha-helices and beta-sheets depicted in red and yellow, respectively. Subsequent molecular docking analyses revealed binding affinity values for the protein-ligand complexes ranging from -5.4 kcal/mol to -6.7 kcal/mol. The model cluster4139_mc|g4968.t1 demonstrated the weakest binding affinity (-5.4 kcal/mol), while models cluster5734_mc|g4624.t1 and cluster4139_mc|g4970.t1 exhibited the strongest binding affinities, both at -6.7 kcal/mol.

4. Discussion

The present study was motivated by the critical need to overcome the inherent toxicity of limonene, a major bottleneck in the microbial valorization of this monoterpene into high-value compounds such as perillyl alcohol[30]. The isolation of a *Meyerozyma caribbica* strain exhibiting unprecedented resistance to limonene concentrations up to 4% (v/v) provided a pivotal starting point for this investigation. Consequently, we employed an integrative multi-methodological approach to elucidate the molecular underpinnings of this exceptional phenotype.

Our initial comparative genomic analysis suggested that membrane transporters could be key genetic determinants of limonene resistance, a hypothesis drawn from evidence in other microorganisms[31]. To empirically test the postulate that active efflux mitigates intracellular accumulation, we quantified limonene levels across several yeast strains. Although the results indicated no statistically significant difference between the mean intracellular limonene concentrations, the absolute values quantified in *M. caribbica* were consistently lower than those measured in the other yeast strains under investigation. A

plausible explanation for this outcome could involve the predetermined exposure time utilized in the experiment; a longer duration might have revealed a more pronounced disparity in limonene accumulation by the cells. *Kluyveromyces marxianus* also exhibited viability post-exposure, its high intracellular limonene load suggests a tolerance mechanism distinct from active efflux, potentially related to the duration of the assay or other intrinsic cellular defenses.

Subsequent transcriptomic analysis aimed to identify which, if any, of the putative transporters identified through genomics were functionally recruited under limonene stress. Notably, the expression levels of most candidate clades (A, B, G, H) remained low and inconclusive. However, a significant and specific upregulation of clade I was observed following prolonged exposure to 4% (v/v) limonene. This targeted transcriptional response strongly implicates the clade I encoded transporter could be directly involved in the cellular defense against limonene toxicity, offering crucial mRNA-level evidence for its role in the resistance phenotype.

To further interrogate the function of the leading candidates, molecular docking simulations were performed on the proteins encoded by clades B and I. The computational models revealed two critical findings: firstly, the predicted structures showed significant homology to known transporter proteins, and secondly, both models demonstrated a robust and favorable binding affinity for limonene. These *in silico* results provide compelling structural evidence that these putative transporters possess the necessary architecture to interact with and potentially facilitate the export of limonene molecules.

5. Conclusion

In conclusion, the convergence of genomic, empirical, transcriptomic, and computational evidence presented herein supports a coherent model for limonene resistance in *M. caribbica*. We propose that this phenotype could be related by the action of specific membrane transporters, most notably the protein encoded by clade I, which is transcriptionally induced by limonene stress. The function of this and other putative pumps is substantiated by their computationally predicted capacity for ligand binding. This work not only elucidates a key survival strategy in a robust yeast isolate but also establishes a foundational framework for future strain engineering. The identification and validation of these efflux transporters open promising avenues for developing superior microbial platforms capable of withstanding high concentrations of toxic monoterpenes, thereby enhancing the biotechnological production of valuable compounds.

Declarations

Acknowledgments

We gratefully acknowledge Dr. Sérgio Antônio Fernandes from the Department of Chemistry at UFV for his guidance in the methodology for intracellular limonene quantification using gas chromatography-mass spectrometry (GC-MS). We also extend our sincere thanks to Dr. Luciano Gomes Fietto from the Department of Biochemistry and Molecular Biology at the Federal University of Viçosa for generously

providing the yeast strains *Saccharomyces cerevisiae*, *Kluyveromyces marxianus*, *Spathaspora passalidarum*, *Spathaspora arborariae*, and *Scheffersomyces stipites*.

Author contributions

Conceptualization: Filipe Augusto Teixeira, Gustavo Molina and Tiago Antônio de Oliveira Mendes; Methodology: Filipe Augusto Teixeira, Fernanda Matias Albuini, Ananda Aguilar and Patrick Neves Squizato; Formal analysis and investigation: Filipe Augusto Teixeira and Tiago Antônio de Oliveira Mendes; Writing - original draft preparation: Filipe Augusto Teixeira; Writing - review and editing: Filipe Augusto Teixeira, Gustavo Molina, Fernanda Matias Albuini, Ananda Aguilar, Patrick Neves Squizato and Tiago Antônio de Oliveira Mendes; Resources: Gustavo Molina and Tiago Antônio de Oliveira Mendes; Supervision: Gustavo Molina and Tiago Antônio de Oliveira Mendes.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Felipe L, de O, Oliveira AM, Bicas JL (2017) Bioaromas – Perspectives for sustainable development. Trends Food Sci Technol 62:141–153. <https://doi.org/10.1016/j.tifs.2017.02.005>
2. Vespermann KAC, Paulino BN, Barcelos MCS et al (2017) Biotransformation of α - and β -pinene into flavor compounds. Appl Microbiol Biotechnol 101. <https://doi.org/10.1007/s00253-016-8066-7>
3. Celestino Jdos, Carvalho R, de Lima LE et al (2014) M da P, Bioprospecting of Amazon soil fungi with the potential for pigment production
4. Ben Akacha N, Gargouri M (2015) Microbial and enzymatic technologies used for the production of natural aroma compounds: Synthesis, recovery modeling, and bioprocesses. Food Bioprod Process 94:675–706. <https://doi.org/10.1016/j.fbp.2014.09.011>
5. Amorim SS, Ruas FAD, Barboza NR et al (2018) Manganese (Mn^{2+}) tolerance and biosorption by *Meyerozyma guilliermondii* and *Meyerozyma caribbica* strains. J Environ Chem Eng 6:4538–4545. <https://doi.org/10.1016/j.jece.2018.06.061>
6. Sukpipat W, Komeda H, Prasertsan P, Asano Y (2017) Purification and characterization of xylitol dehydrogenase with L-arabitol dehydrogenase activity from the newly isolated pentose-fermenting yeast *Meyerozyma caribbica* 5XY2. J Biosci Bioeng 123:20–27. <https://doi.org/10.1016/j.jbiosc.2016.07.011>
7. Haft DH (2015) Using comparative genomics to drive new discoveries in microbiology. Curr Opin Microbiol 23:189–196. <https://doi.org/10.1016/j.mib.2014.11.017>
8. Bicas JL, Pastore GM (2007) Isolation and screening of D-limonene-resistant microorganisms. Brazilian J Microbiol 38:563–567. <https://doi.org/10.1590/S1517-83822007000300034>

9. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24:637–644.
<https://doi.org/10.1093/bioinformatics/btn013>
10. Mendes TAO, Lobo FP, Rodrigues TS et al (2013) Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. *Mol Biol Evol* 30:951–963.
<https://doi.org/10.1093/molbev/mst001>
11. Li L, Stoeckert CJJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes – Li 13 (9): 2178 – *Genome Research*. *Genome Res* 13:2178–2189.
<https://doi.org/10.1101/gr.1224503.candidates>
12. Jones P, Binns D, Chang HY et al (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
13. Mitchell AL, Attwood TK, Babbitt PC et al (2019) InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47:D351–D360.
<https://doi.org/10.1093/nar/gky1100>
14. Molina G, Bution ML, Bicas JL et al (2015) Comparative study of the bioconversion process using R-(+)- and S(-)-limonene as substrates for *Fusarium oxysporum* 152B. *Food Chem* 174:606–613.
<https://doi.org/10.1016/j.foodchem.2014.11.059>
15. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38:3022–3027. <https://doi.org/10.1093/molbev/msab120>
16. Ye J, Coulouris G, Zaretskaya I et al (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. <https://doi.org/10.1186/1471-2105-13-134>
17. Agatep R, Kirkpatrick RD, Parchaliuk DL et al (1998) Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. *Tech Tips Online* 3:133–137. [https://doi.org/10.1016/S1366-2120\(08\)70121-1](https://doi.org/10.1016/S1366-2120(08)70121-1)
18. Gasteiger E, Gattiker A, Hoogland C et al (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784–3788. <https://doi.org/10.1093/nar/gkg563>
19. Nielsen H, Tsirigos KD, Brunak S, von Heijne G (2019) A Brief History of Protein Sorting Prediction. *Protein J* 38:200–216
20. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹¹ Edited by, Cohen F (2001) *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>
21. Kelley LA, Mezulis S, Yates CM et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>
22. Krieger E, Joo K, Lee J et al (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function and Bioinformatics* 77:114–122

23. Kawabata T, Sugihara Y, Fukunishi Y, Nakamura H (2013) LigandBox: A database for 3D structures of chemical compounds. *Biophys (Japan)* 9:113–121. <https://doi.org/10.2142/biophysics.9.113>
24. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28:1145–1152. <https://doi.org/10.1002/jcc.20634>
25. Morris GM, Goodsell DS, Halliday RS et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14%3C1639::AID-JCC10%3E3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14%3C1639::AID-JCC10%3E3.0.CO;2-B)
26. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
27. Grueber CE (2015) Comparative genomics for biodiversity conservation. *Comput Struct Biotechnol J* 13:370–375. <https://doi.org/10.1016/j.csbj.2015.05.003>
28. Coutinho TJD, Franco GR, Lobo FP (2015) Homology-Independent Metrics for Comparative Genomics. *Comput Struct Biotechnol J* 13:352–357. <https://doi.org/10.1016/j.csbj.2015.04.005>
29. Yan N (2013) Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem Sci* 38:151–159. <https://doi.org/10.1016/j.tibs.2013.01.003>
30. Alonso-Gutierrez J, Chan R, Batth TS et al (2013) Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. *Metab Eng* 19:33–41. <https://doi.org/10.1016/j.ymben.2013.05.004>
31. Wang Y, Lim L, Diguistini S et al (2013) A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytol* 197:886–898

Figures

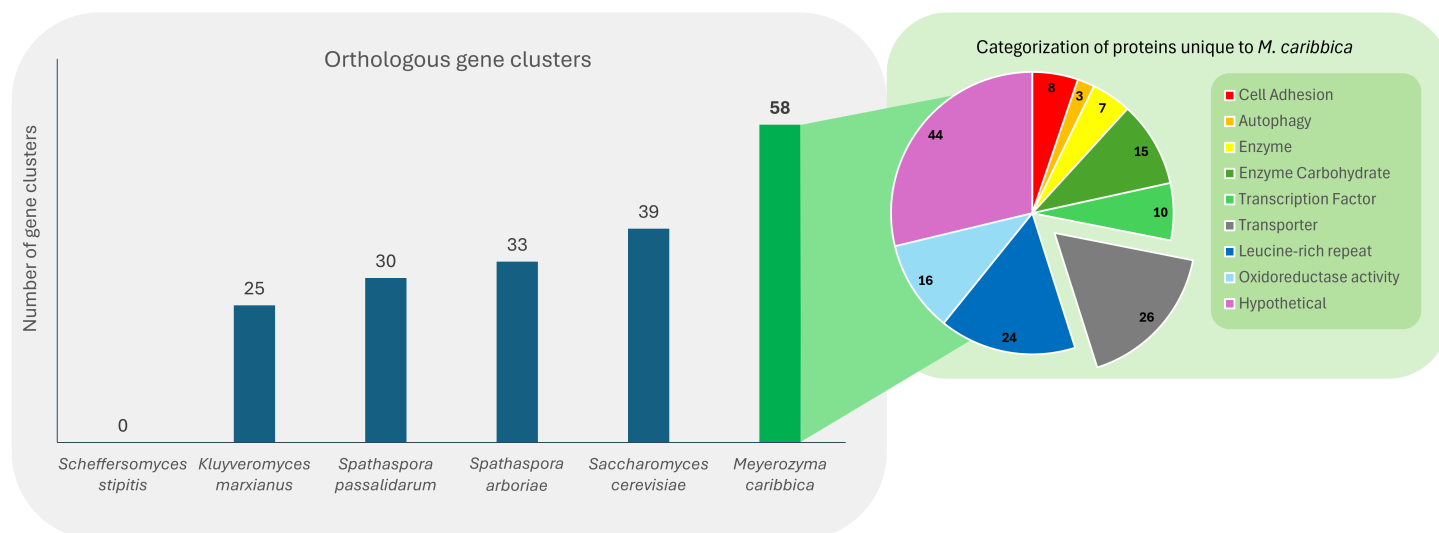


Figure 1

Number of orthologous gene clusters found for each yeast submitted to comparative genomics and categorization of proteins unique to *M. caribbica*.

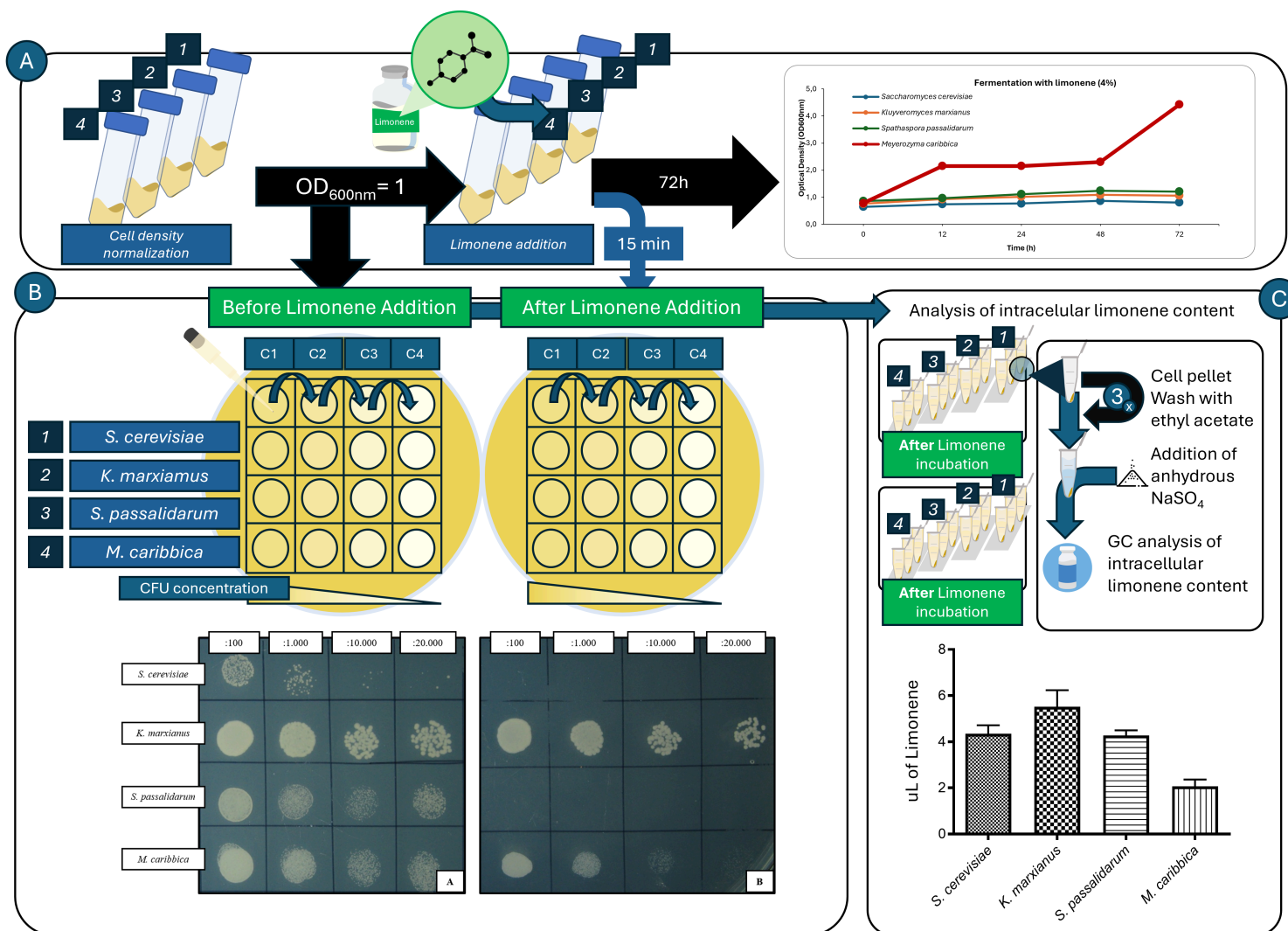


Figure 2

Intracellular limonene quantification. A: Sample preparation methodology and growth curves of the four tested yeast strains following 72-hour cultivation. B: Cell viability assay before and after 15-minute limonene exposure, performed via serial dilution of the cultures. C: Intracellular limonene quantification by gas chromatography.

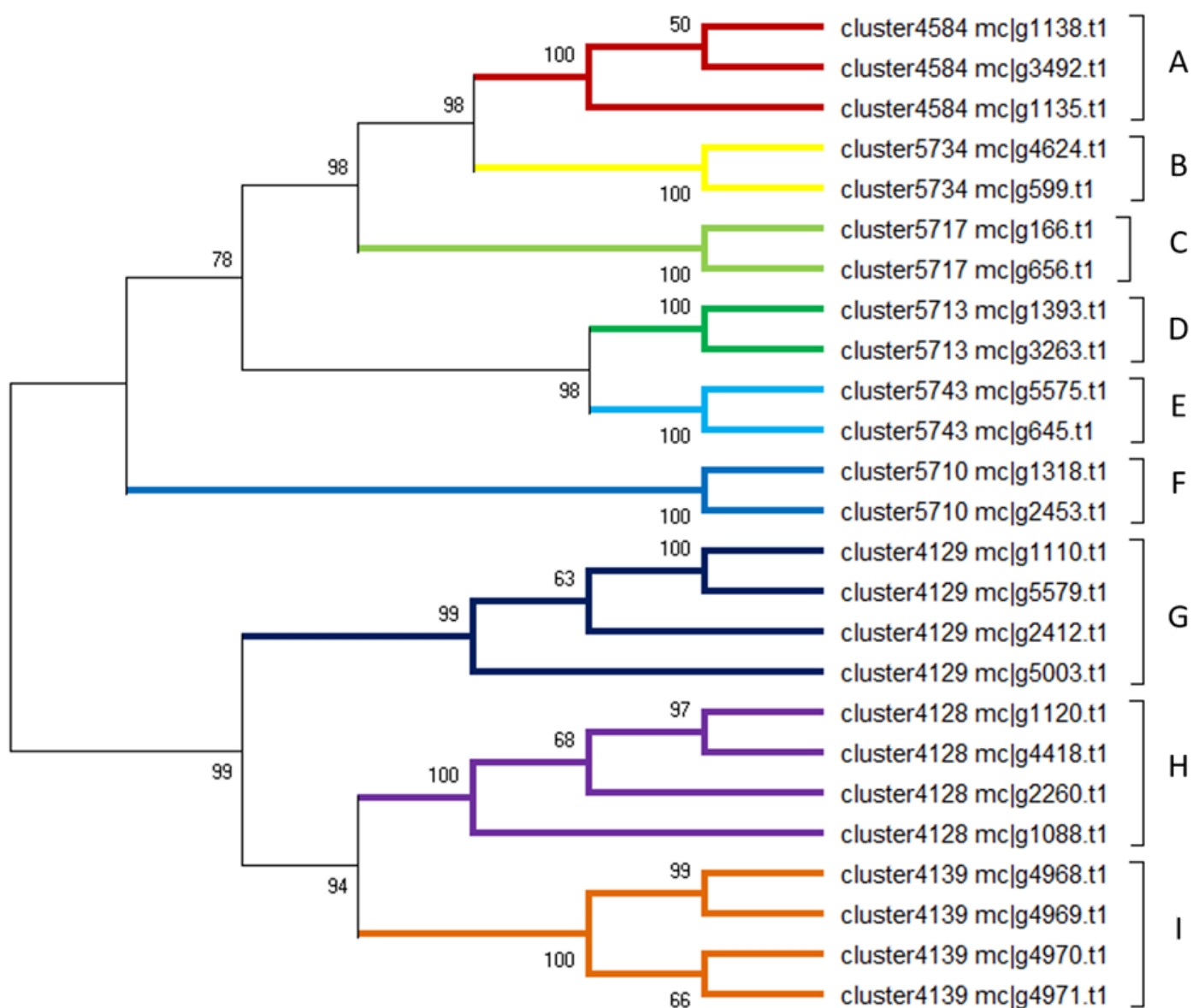


Figure 3

Phylogenetic distribution of the nine clades formed from the 26 MFS-type transporter sequences identified in *M. caribbica*.

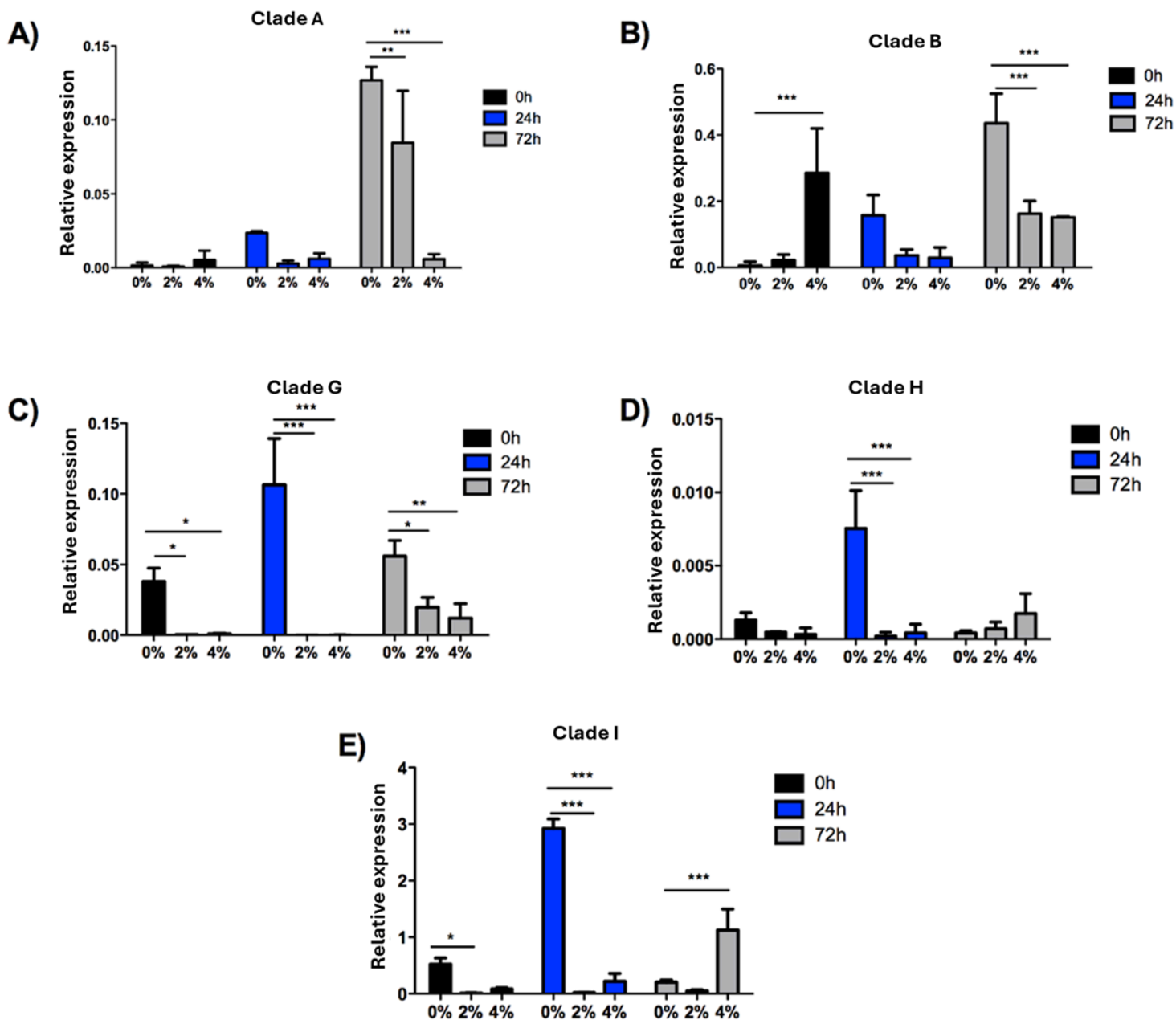


Figure 4

Relative expression of genes from clades A, B, G, H, and I. Significant relative expression was observed solely for clade I when *M. caribbica* was cultivated for 72 hours in the presence of 4% (v/v) limonene.

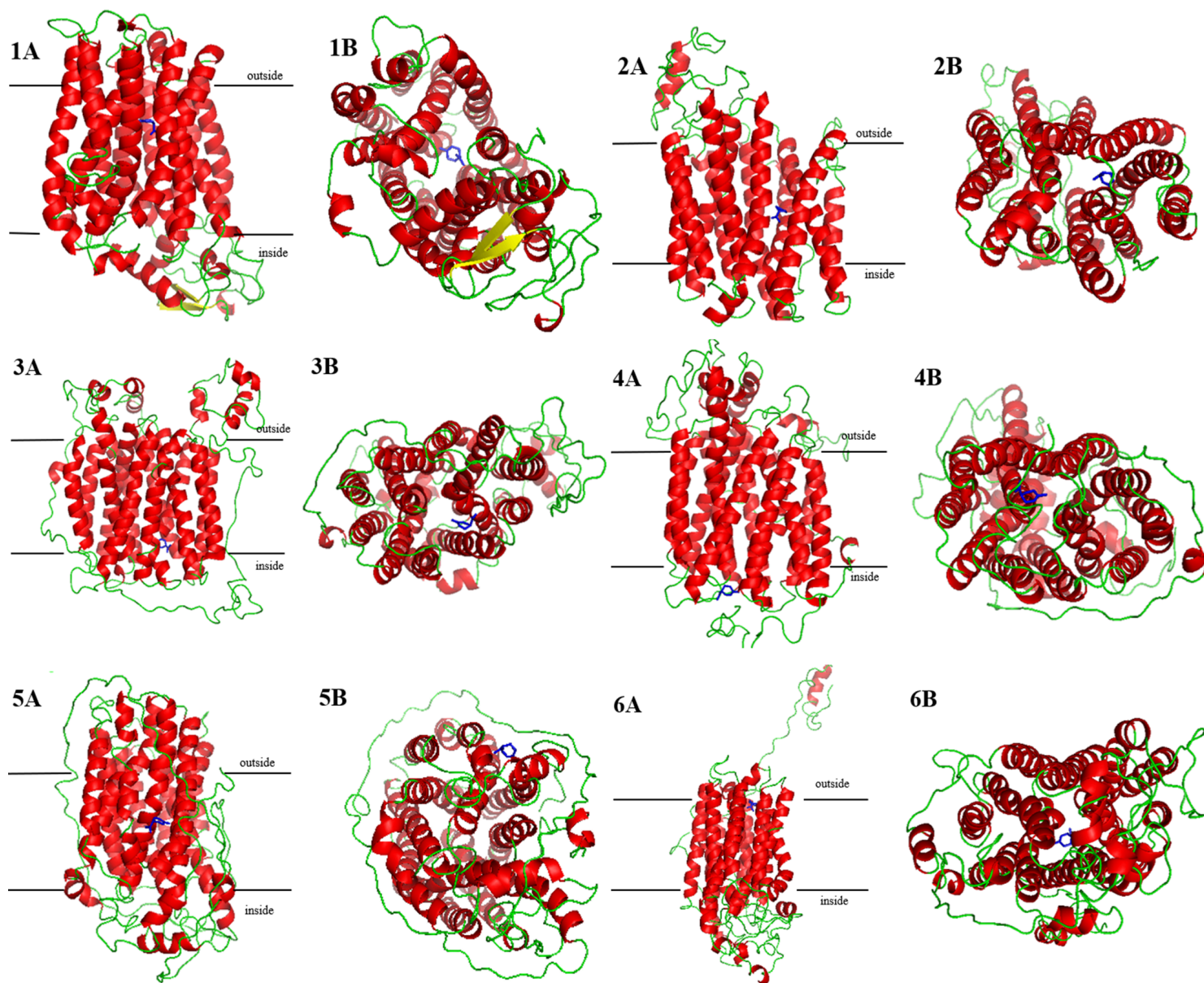


Figure 5

Optimized predicted models from molecular docking. A) Top view and B) cross-sectional view of each optimized model. The limonene molecule is highlighted in blue.