Predicting functions of putative fungal sesquiterpene synthase genes based on multiomics data analysis

Tetyana Nosenko[a*], Ina Zimmer[a], Andrea Ghirardo[a], Tobias G. Köllner[b], Baris Weber[a], Andrea Polle[c], Maaria Rosenkranz[a,1], Jörg-Peter Schnitzler[a]

[a]Helmholtz Zentrum München, Research Unit Environmental Simulation, 85764 Neuherberg, Germany

[b]Max Planck Institute for Chemical Ecology, Department of Natural Product Biosynthesis, 07745 Jena, Germany

[c]Forest Botany and Tree Physiology, University of Göttingen, 37077 Göttingen, Germany

[1]Present affiliation: Institute of Plant Sciences, Ecology and Conservation Biology, University of Regensburg, 93053 Regensburg, Germany

*Corresponding author: Tetyana Nosenko

E-mail address: tetiana.nosenko@helmholtz-muenchen.de

Abstract

Sesquiterpenes (STs) are secondary metabolites, which mediate biotic interactions between different organisms. Predicting the species-specific ST repertoires can contribute to deciphering the language of communication between organisms of the same or different species. High biochemical plasticity and catalytic promiscuity of sesquiterpene synthases (STSs), however, challenge the homology-based prediction of the STS functions.

Using integrated analyses of genomic, transcriptomic, volatilomic, and metabolomic data, we predict product profiles for 116 out of 146 putative *STS* genes identified in the genomes of 30 fungal species from different trophic groups. Our prediction method is based on the observation that STSs encoded by genes closely related phylogenetically are likely to share the initial enzymatic reactions of the ST biosynthesis pathways and, therefore, produce STs *via* the same reaction route. The classification by reaction routes allows to assign STs known to be emitted by a particular species to the putative *STS* genes from this species. Gene expression information helps to further specify these ST-to-STS assignments. Validation of the computational predictions of the STS functions using both *in silico* and experimental approaches shows that integrated multiomic analyses are able to correctly link cyclic STs of non-cadalane type to genes. In the process of the experimental validation, we characterized catalytic properties of several putative *STS* genes from the mycorrhizal fungus *Laccaria bicolor*. We show that the STSs encoded by the *L. bicolor* mycorrhiza-induced genes emit either nerolidol or α-cuprenene and α-cuparene, and discuss the possible roles of these STs in the mycorrhiza formation.

Keywords: functional prediction, fungi, multiomics, mycorrhiza, sesquiterpene synthase genes, sesquiterpenes

Abbreviations: ST, sesquiterpene; STS, sesquiterpene synthase; TPS, terpene synthase; RC, reaction class; VOC, volatile organic compound; DGE, differential gene expression; GC-MS, gas chromatography–mass spectrometry; HGT, horizontal gene transfer.

1. Introduction

In the process of co-evolution, fungi have developed diverse strategies for defense and communication with other organisms. Similar to plants and bacteria, fungi produce a large group of semi-volatile organic compounds sesquiterpenes (STs) and their derivatives sesquiterpenoids, which play important roles in mediating their mutualistic and antagonistic interactions. STs and sesquiterpenoids have received major attention for their application for human use: as fragrance ingredients in perfumery, food production, and cosmetics, as well as pharmaceutical compound precursors for the development of anti-inflammatory and anti-cancer drugs (Breitmaier, 2006; Dai et al., 2021; Moujir et al., 2020). The discovery of biomedical properties of fungal STs and their derivatives has promoted growing interest into their original ecological function (Kramer and Abraham, 2012). In particular, many fungal STs exhibit antifungal and antibacterial properties and serve as chemical weapons against parasitic microorganisms and competitors (Dichtl et al., 2010; Hansson et al., 2012; Quintana-Rodriguez et al., 2018; Shen et al., 2019; Presley et al., 2020). Along with defensive functions, STs can act as signaling molecules in the establishment and regulation of symbiosis between fungal species and their insect and plant hosts (Burkhardt et al., 2019; Ditengou et al., 2015). STs represent a class of terpenes characterized by 15-carbon skeleton derived from the linear precursor farnesyl pyrophosphate (FPP); they can be linear or contain different number of rings. The structural diversity of STs is achieved *via* reactions catalyzed by sesquiterpene synthases (STSs; Fig. 1). Downstream chemical modifications of the 15-carbon skeleton (*e.g.*, oxygenation, acetylation, lipidation) and further rearrangements result in a variety of sesquiterpenoid compounds (Dai et al., 2021).

Depending on their conformational flexibility, STSs exhibit different degrees of catalytic fidelity/promiscuity (Christianson, 2008). While high-fidelity STSs produce unique STs, promiscuous STSs may convert the substrate FPP into dozens of various side products through a cascade of hydroxylation, elimination, cyclization, and rearrangement reactions. A single substitution of an amino acid residue involved in protein folding may alter the STS product profile (López-Gallego et al., 2010b). These features explain differences often observed between the ST profiles of STSs encoded by within-species paralogs or homologous genes from closely related species (Wawrzyn et al., 2012; Ichinose and Kitaoka, 2018). Information on fungal *STS* genes, their product profiles, and the dynamics of ST biosynthesis and emission may help to decipher the language of interspecies communication and provide deeper insights into species ecology. The 1000 Fungal Genomes Project (2018) opens an opportunity to investigate the evolution of enzymatic machinery of ST biosynthesis in fungi using comparative genomics approaches. Detection of different classes of terpene synthase (*TPS*) genes, including *STS*s, mainly relies on the presence of two metal-binding motifs: an aspartate-rich DDXXD/E and NSE (ND/EXXSXXXE; Leeper and Vederas, 2000; Rynkiewicz et al., 2001). Although enzymatic activity of many bacterial, fungal, and plant STSs have been characterized experimentally, high sequence diversity and catalytic promiscuity of these proteins (Zhang et al., 2020a) challenge a homology-based detection of the *STS* genes and their product profile prediction. Recently, Durairaj et al. (2021) have developed a machine learning approach combining sequence and structure information for functionally characterized and putative *STS*s to accurately predict STS product specificity across plant species. In plants, enzymatic activity has been characterized for nearly 300 *STS* genes from over one hundred species (Durairaj et al., 2021). Only 97 *STS* genes from 33 species have been characterized experimentally (Table S1), which is not sufficient for applying similar computational approaches for predicting STS functions in this group of organisms (Mischko et al., 2018).

In this study, we use a multiomics approach to assess the patterns of the *STS* gene family evolution and reconstruct the most probable product profiles for putative *STS* genes identified in genomes of 30 species representing three fungal phyla: Mucoromycota, Ascomycota, and Basidiomycota. To this end, we combine four types of data: (i) primary sequence information (i.e., phylogenetic analyses), (ii) species-specific ST profiles, (iii) ST profiles of experimentally characterized fungal STS, and (iv) gene expression data. Our ST-to-STS assignment approach is based on the observation that plant and fungal *STS* genes, which share high sequence similarity and thus belong to the same phylogenetic clade, are likely to share their primary cation selectivity and initial cyclization reactions. Furthermore, we use gene expression information to discriminate between predicted ST profiles of the closely related fungal *STS* genes arisen *via* recent gene duplications. To assess results of the bioinformatic predictions, we experimentally characterized four *STS* genes involved in the symbiosis between the mycorrhizal fungus *Laccaria bicolor* and its plant hosts.

2. Materials and Methods

2.1. Sesquiterpene synthase gene identification and phylogenetic analysis

We used 30 species (31 strains) from three fungal phyla (Mucoromycota, Ascomycota, and Basidiomycota) and four different trophic groups (saprophytes, plant parasites, mycoparasites, and mycorrhizal fungi; Table S2). These taxa were selected based on the availability of whole-genome sequence data from the list of species, for which we recently characterized species-specific volatilomes (Guo et al., 2021). Genome sequence, gene annotation, and predicted protein model data are available for all 30 species from publicly accessible databases (Table S2).

To detect putative terpene synthase (*TPS*) genes, whole-genome protein models were screened using HMMER v.3.2.2 (2020) and two Pfam domains: PF03936 and PF06330 (TPS metal-binding and trichodiene synthase TRI5 domains, respectively). In addition, amino acid (aa) sequences of known fungal (Tables S1 and S3), and selected plant and bacterial *STS* genes (Durairaj et al., 2021; Reddy et al., 2020; Yamada et al., 2015) were obtained from publicly accessible databases and used as queries for sequence similarity searches (BLASTp; e-value threshold e10^{-6}) against the whole-genome protein models of the species of interest. To filter out fungal di- and triterpene synthase genes, a set of *TPS* genes identified using these two methods were clustered together with the set of known fungal *STS* genes using CLASS2 v.1.0 (Kelil et al., 2008). The CLASS2 gene clusters containing among others functionally characterized *STS* genes were selected for further analyses. Sequences within each *STS* cluster were aligned using MUSCLE (Edgar, 2004). Sequence alignments were refined manually using Mesquite v.3.61 (Maddison and Maddison, 2018) and subjected to Maximum Likelihood (ML) phylogenetic analyses using RAxML (Stamatakis, 2014) with the bootstrapping method (100 bootstrap replicates) and LG aa substitution model. A two- and three-letter code was assigned to the monophyletic clades with bootstrap support above 50%. Because our analysis includes species from different fungal phyla, the clade nomenclature used in this study differs from those developed in previous publications on the Basidiomycota STSs (Ichinose and Kitaoka, 2018; Quin et al., 2013; Wawrzyn et al., 2012; Zhang et al., 2020b). In addition, gene composition of the Basidiomycota clades resulting from our analysis may differ from previously established systems as we excluded highly divergent sequences potentially derived *via* interkingdom Horizontal Gene Transfer (HGT; Flynn and Schmidt-Dannert, 2018; Jia et al., 2019).

2.2. Assignment of sesquiterpenes and sesquiterpenoids produced by fungi to the putative STS *genes*

Information on the production of STs and ST derivatives was collected for the fungal species included in this study from publications referenced in Supplementary Table S4[1] and used as additional evidence for predicting functions of the putative *STS* genes. Briefly, STs emitted by fungi of interest and those produced by experimentally characterized fungal STSs (Tables S1 and S4, respectively) were classified based on biosynthesis pathway according to the classification scheme of Durairaj *et al.* (2021) adapted for fungi (Fig. 1). According to this scheme, STs were divided into 14 reaction classes (RCs) depending on the type of primary cation (farnesyl or nerolidyl) and initial cyclization reaction on the ST biosynthesis pathway. In addition to the classes defined in Durairaj *et al.* (2021), we split STs branching out of the bisabolyl cation into two RCs: N131 including STs derived *via* a transformation of bisabolyl cation to the cuprenyl cation (*e.g.*, trichodiene, cuprenene, barbatene, widdradiene, chemigrene, and thujopsene) and N132 consisting of all other STs branching out of the bisabolyl cation (Hong and Tantillo, 2014). Catalytic reaction pathways were determined for each ST using schemes available from IUBMB Enzyme Nomenclature Supplement 24 (2018) and verified using publications listed in Supplementary Table S1. RC codes of experimentally characterized *STS* genes were protracted to their putative homologs. For each species, STs were matched to the putative *STS* genes based on RCs using custom R scripts (R Core Team, 2020). For each species, the same STs were assigned to all genes that belong to one clade. More than one RC were allowed for those STs that are known to be synthesized *via* several alternative pathways (*e.g.*, cadalanes; Quin et al., 2014).

2.3.   In silico differential expression analyses of sequiterpene synthase genes in fungi-plant interactions

To assess the functionality of the two putative *STS* genes recently duplicated in Basidiomycota, we analyzed RNAseq data available from the NCBI Sequence Read Archive (SRA) for six species, including *Cortinarius glaucopus*, *Heterobasidion annosum*, *L. bicolor*, *Piloderma croceum*, *Postia placenta*, and *Trametes versicolor*. Accession numbers of the respective SRA BioProjects are provided in Supplementary Table S2. RNA-Seq reads were filtered, quality trimmed using Trimmomatic v.0.35 (Bolger et al., 2014), and aligned to the reference genomes of corresponding species using STAR v2.5.2a (Dobin et al., 2013). Read pairs aligned to exonic regions were summarized per gene using featureCounts (Liao et al., 2014).

Differential gene expression (DGE) analysis was conducted using seven previously published RNA-Seq datasets to assess changes in the *STS* gene expression in (1) *H. annosum* due to the HetPV13-an1 virus infection (SRA Bioproject PRJNA362289, Vainio et al., 2018), (2) *P. placenta* during the wood colonization (single species culture; PRJNA422777; Zhang et al., 2019) and (3) competition with another wood-colonizing fungal species (PRJNA634380, Presley et al., 2020), (4) *T. versicolor* during its interaction with an algae *Microcystis aeruginosa* PCC7806 (PRJNA427843, Dai et al., 2018) and (5) wood colonization (single-species culture; PRJNA422777, Zhang et al., 2019), (6) *P. croceum* and (7) *L. bicolor* during the establishment of mycorrhiza with *Quercus robur* and *Populus tremula* x *alba*, respectively (PRJNA269500, Kohler et al., 2015; PRJNA443987, Basso et al., 2020). Count matrices constructed for these projects (see above) were analyzed using the R package DESeq2 v1.11.15 (Love et al., 2014) and negative binomial generalized linear model fitting and Wild statistics (nbGLM-Wald; for comparing the stress condition to the control in the projects 1, 6 & 7). For

---

[1] The Microbial volatile organic compound database mVOC was not accessible during the period when the analyses presented in this paper were conducted. After these analyses were complete, the new version of the mVOC database has been released (https://bioinformatics.charite.de/mvoc/), which contains partial ST information for several species included in Table S4.

the data resulting from the time-series experiments (projects 2-5), we conducted Likelihood Ratio Test (LRT), in which every subsequent time point was compared to the reference time point zero. In the project 4, between-time-points expression differences in the control samples were taken into account. All protein-coding genes represented by at least three counts in at least two libraries were included in the analyses. The thresholds for DGE were set to the fold change $FC \geq 2$ and $P < 0.05$ after the FDR adjustment for multiple testing.

2.4.   Sesquiterpene synthase gene cloning, heterologous expression, and in vitro enzymatic activity assays

Gene models encoding four putative *L. bicolor* STSs were verified and corrected using transcript-to-genome sequence alignments and IGV v.2.4 (Robinson et al., 2011). CDS sequences encoding mature *LbSTS4a, b, c, LbSTS6,* and a cuprenene synthase gene from *Coprinus cinereus* (*Cop6*; Agger et al. 2009) were optimized for expression in *Escherichia coli* using GeneArt portal software (Invitrogen-ThermoFisher Scientific). The resulting sequences were synthesized, cloned into the Gateway donor vector pENTR221 (Invitrogen), and subsequently subcloned into the Gateway destination vector pDEST17 (Invitrogen) by Life Technologies-ThermoFisher Scientific (Darmstadt, Germany). Genes were expressed in the chemically competent *E. coli* cells BL21(DE3) (ThermoFisher Scientific). Crude *E. coli* protein extraction and purification of the His-tagged STS proteins by affinity chromatography on Pure-Cube Ni-NTA agarose (Cube-Biotech, Hessisch-Oldendorf, Germany) were conducted as described in Schnitzler *et al.* (2005). Analysis of *in vitro* enzyme activities from His-tagged STS proteins and crude protein extracts with the substrate FPP were performed using stir the bar sorptive extraction (SBSE) technique coupled with thermal desorption–gas chromatography–mass spectrometry (TD-GC-MS). Two enzymatic activity assays were conducted independently at the Helmholtz Zentrum München (HMGU) and Max-Planck Institute in Jena, Germany. Details of these assays are provided in Supplementary Methods S1.

To distinguish between the α-cuprenene and β-himachalene GS-MS spectra, we used a commercially available (+)-cuparene standard (Sigma-Aldrich, Deisenhofen, Germany) and two vectors containing control genes: β-himachalene synthase from *Cryptosporangium arvum* (*HcS*, provided by J. Dickschart, Bonn, Germany; Rinkel and Dickschart, 2019) and α-cuprenene synthase *Cop6* from *Coprinus cinereus* (Agger et al., 2009). The annotation of individual sesquiterpenes was conducted by comparing the obtained mass spectra with those of the commercially available authentic standards ((+)-cuparene, nerolidol; Sigma-Aldrich) and NIST05 and Wiley library spectra. The representative *m/z* and retention indices of the STs were calculated according to van den Dool and Kratz (1963).

3. Results

3.1.   *Identification and phylogeny of fungal sesquiterpene synthase genes*

Sequence similarity- and HMMER-based searches and subsequent CLASS2 (Kelil et al., 2008) clustering of the identified *TPS* sequences with 94 functionally characterized fungal *STS* genes (Table S1) resulted in 147 putative *STS* genes from 25 out of 31 fungal genomes (24 out of 30 species) included in these analyses. Together, the putative and functionally characterized *STS* genes formed three large gene clusters including 53 species (55 strains) of Asco- and Basidiomycota (Table S3, Fig. 2). No *STS* genes have been detected in genomes of the Mucoromycota species (*Mucor circinelloides* f. *lusitanicus* and *Rhizopus oryzae*) included in this study. Except for the short metal- and substrate-binding motifs typical for the class I *TPS* genes (Zhou and Peters, 2009), the three clusters share no aa sequence similarity and therefore each is represented as a separate phylogenetic tree (Table S3, Figs. 2 and S1). The phylogenetic

tree constructed for the largest cluster (A; Figs 2A and S1A) consists of 177 genes from 42 species (43 strains): 23 species of Basidiomycota and 19 species (20 strains) of Ascomycota, - which form eight monophyletic clades. The second cluster (B; Figs 2B and S1B) consists of 53 genes from 22 species (eight Basidiomycota and 14 Ascomycota), including functionally characterized genes encoding trichodiene (Pfam family TRI5; B-I.2), longiborneol (B-II.1), and cuprenene (B-II.2) synthases. The largest clades of Asco- and Basidiomycota in Cluster B include both putative and experimentally characterized *STS* genes. The two phylogenetic trees show a clear separation between Asco- and Basidiomycota, except clades A-1.1 and A-III.1, which contain species from both phyla. The most parsimonious explanation for the occurrence of the Ascomycota-like *STS* genes in the basidiomycetes *P. croceum* (Fig. S2) and the Basidiomycota-type of genes in several species of ascomycetes (A-I.1 and A-III.1, respectively) is an interphylum HGT. HGT from Basidiomycota has been proposed as a source of δ-6-protoilludene synthase gene in the ascomycetes *Diaporthe* sp. by de Sena Filho et al. (2016). Unlike Clusters A and B, Cluster C is specific to Ascomycota (Figs. 2C and S1C). Sequences within this cluster form two monophyletic clades: C-I represented by putative *STS* genes only and C-II, which along with sequences from our species of interest includes functionally characterized aristolochene and β-caryophyllene synthase genes (*Ari1* and *CaTPS*, respectively).

We observed high variation in the number of *STS* genes among the 31 genomes included in this study (from zero in, e.g., *U. hordei* to 19 in *T. versicolor*). In general, the number of *STS* genes is significantly higher in Basidiomycota species compared to Ascomycota. The distribution of *STS* genes across phylogenetic trees suggests multiple recent gene duplications in Basidiomycota. The number of *STS* genes showed a weak correlation with genome size (Pearson correlation 0.51, p-value < 0.01) and no significant correlation with the total number of protein-coding genes in a genome and species trophic modes (Fig. 3).

Our phylogenetic analyses demonstrate that *STS* genes that belong to the same monophyletic clade may have very different product profiles (Table S1). Thus, two functionally characterized *P. placenta* paralogs from Clade A-III.1, *PpSTS08* and *PpSTS14*, encode δ-6-protoilludene- and pentalene synthases, respectively. In Clade B-II.2, STSs encoded by closely related genes *Cop6* from *C. cinereus* and *Omp9* and *Omp10* from *Omphalotus olearius* provide another example of enzymes with different ST profiles. To establish an affiliation between the phylogeny and enzymatic products of fungal STSs, STs synthesized by functionally characterized STSs were assigned to RCs based on the primary cation and initial cyclization reactions as shown in Figure 1. The resulting RC distribution across the phylogenetic trees shows that STSs belonging to the same monophyletic clade tend to share the initial catalytic steps of the ST biosynthesis and thus belong to the same RCs (Figs. 2 and S1). This observation suggests that enzymes encoded by putative *STS* genes are likely to share biosynthesis pathways with their functionally characterized homologs from the same phylogenetic clade and produce STs within RC assigned to this clade.

Recently, Guo *et al.* (2021) showed an association between the patterns of volatile organic compounds (VOCs) emitted by fungal species and their trophic modes, lifestyle, and host type. These results from *in vivo* VOC measurements suggest that species-specific composition of genes encoding STSs may depend on the ecological features as well. To test this hypothesis, we analyzed the distribution of species characterized by different trophic modes across phylogenetic clades and RCs assigned to these clades. We found, however, that the trophic mode does not affect the *STS* gene clustering and gene number; each phylogenetic clade and each RC are represented by species of several trophic modes (Fig. 2). To summarize, fungi of different trophic modes share repertoires of their *STS* genes, which allow them to produce a very broad spectrum of STs in a condition-specific manner. Clustering of these *STS* genes is based exclusively on taxonomy and not on species ecology.

3.2.   Function prediction for the putative sesquiterpene synthase genes

Twenty-five out of 30 species included in this study have been reported to produce STs (total, 202 STs and ST derivatives, including 46 cadalanes; Table S4). One or more putative *STS* genes were identified in the genomes of 23 sesquiterpene-emitting species (17 Asco- and six Basidiomycota). We utilized these data to find possible products of the putative *STS* genes. Similarly to the products of functionally characterized STSs, we classified the STs emitted by the species of interest by RCs. Classification by RCs of the ST and ST derivatives reported for these species is provided in Supplementary Data S1. Using this classification, STs were assigned to 116 putative *STS* genes from 19 species (45 and 71 genes from 13 and 6 Asco- and Basidiomycota species, respectively; Table S5, Fig. 4). Out of total 166 unique ST-species combinations, 120 combinations were assigned to a single phylogenetic clade and 46 STs received ambiguous assignments. As expected, the majority of the ambiguously assigned STs were of the cadalane type. The number of STs assigned to one gene varied from one to 14. Predicted product profiles of 50% of these genes consist of 1-3 STs. The number of STs that have not been assigned to any gene is larger among Ascomycota. One factor that may contribute to this result is a strain-specific composition of both *STS* genes and STs emitted by some Ascomycota species (e.g., *T. harzianum*; Guo et al., 2021). For example, in the present study, STs emitted by six different *T. harzianum* strains were matched to the *STS* genes predicted for reference genome assemblies available for only two strains. A between-strain difference in the ST profiles (Guo et al., 2021) and *STS* gene portfolio (Table S3) may explain a high number of unassigned *T. harzianum* STs.

Eight *STS* genes from genomes of three Basidiomycota genomes analyzed in this study were previously functionally characterized (Ichinose and Kitaoka, 2018; Zhang et al., 2020b), including *Hetan454193* (*H. annosum*), *Pilcr825684* (*P. croceum*), and *PpSTS01*, *PpSTS03*, *PpSTS06*, *PpSTS08*, *PpSTS14*, and *PpSTS25* (*P. placenta*; Table S1). To assess the quality of our bioinformatic prediction, we compared the putative and experimentally established ST profiles of these genes. The result of this comparison demonstrates that, at the level of phylogenetic clades, all cyclic STs of non-cadalane type received correct and unambiguous assignments (Table S5). Cadalanes and acyclic terpenes, which can be synthesized *via* several pathways (Fig. 1), received ambiguous assignments to genes from several phylogenetic clades. The limitation of this method is that it does not discriminate between multiple gene copies from the same species within one clade. Thus, in *P. croceum*, viridiflorene (RC F112) was assigned to both the functionally characterized gene *Pilcr825684* and its sister copy *KIM73468*, while in *H. annosum*, δ-6-protoilludene and its four derivatives (RC F12) affiliated not only with the functionally characterized *Hetan454193* but also with other five putative *STS* genes from the clade A-III.1c.

3.3.   Discriminating functions of closely related sesquiterpene synthase genes based on their expression profiles

The ST-to-gene assignment method described above allowed affiliating STs with phylogenetic clades for each species. However, it did not differentiate between the within-clade paralogs derived *via* recent gene duplications, which are abundant in, for example, Basidiomycota (Figs. 2 and S1). The fate of duplicated genes includes the following scenarios: loss of function by one copy, acquisition of a new biological function, retention of the ancestral function, and, rarely, sub-functionalization of the two copies (Lynch and Conery, 2000; Ohno, 1970). To discriminate between these scenarios and to further specify functional predictions for the putative *STS* genes, we analyzed gene responses to different types of biotic interactions in the five Basidiomycota species, which received the ST-to-STS assignments. This approach

is based on the assumption that closely related genes which respond similarly to environmental triggers are likely to have the same biological function, while the expression pattern divergence may indicate functional diversification. *P. placenta* genes from Clade A-III.3 *PpSTS01*, *PpSTS03*, and *PpSTS06*, which encode proteins characterized experimentally as α-muurolene, γ-cadinene, and α-gurjunene synthases respectively (Ichinose and Kitaoka, 2018), exhibit different expression profiles during the three stages of wood colonization (Fig. 5; PRJNA634380, PRJNA422777) and thus support the aforementioned assumption. Among the six closely related *H. annosum STS* genes from A-III.1, *Hetan148791* and *Hetan446121* were upregulated, *Hetan454193* and *Hetan42859* were downregulated, and *Hetan51706* and *Hetan48772* did not show any expression change in response to the HetPV13-an1 virus infection. Hetan454193 is known to produce δ(6)-protoilludine (Zhang et al., 2020b). Accordingly, the expression patterns suggest that δ(6)-protoilludine can be excluded from the product lists predicted for the upregulated and no-response *H. annosum STS* genes from this clade. *L. bicolor* genes *LbSTS6* and *XP001886282* (B-II.2) provide another example of a potential functional diversification of *STS* gene copies after duplication. In contrast, upregulation of all three *L. bicolor* paralogs from the clade A-III.2 (*LbSTS4a*, *b*, and *c*) during the mycorrhiza development suggests that these genes share their function. Interestingly, in *P. croceum*, a contact with the host roots only slightly induced the γ-cadinene/viridiflorene synthase gene *Pilcr825684* (Zhang et al., 2020b) closely related to the *LbSTS4* genes, while expression of *KIM84489*, *KIM74248*, and *KIM73930* genes, which probably originate *via* interphylum HGT (Fig. S2) and have no homologs in *L. bicolor,* were strongly upregulated during the mycorrhiza development. This observation suggests that nonrelated *LbSTS4* and *KIM84489*, *KIM74248*, and *KIM73930* may nevertheless have similar product profiles, while products of closely related *Pilcr825684* and *LbSTS4* may differ. The comparison of the *Pilcr825684* and *LbSTS4* product profiles (Zhang et al., 2020b, this study) confirms this prediction (see next section) and suggests a possible neofunctionalization of the *LbSTS4* paralogs. These examples prove DGE analysis as a useful method not only for detecting fungal *STS* genes involved in the interspecies interactions, but also for specifying their product profile predictions.

3.4.   *Functional characterization of the Laccaria bicolor sesquiterpene synthase genes induced during the establishment of mycorrhiza with poplar*

In addition to the *in silico* verification of the results of our bioinformatic predictions, we experimentally assessed functions of four *L. bicolor* genes significantly induced during the mycorrhiza formation: *LbSTS4a, b, c*, and *LbSTS6* (Figs. 5 and S3; Basso et al., 2020). Genes *LbSTS4a*, *b*, and *c* share high sequence similarity (100% node bootstrap support) and are located on different chromosomes. The three genes belong to the phylogenetic clade A-III.2a, which includes characterized fungal STSs producing the cadalane STs *via* the nerolidyl cation and C1,C10 cyclization reaction (N:C1,C10; Figs. 2 and S1). Out of the STs emitted by *L. bicolor*, the best candidates for the *LbSTS4a*, *b*, and *c* products are eight cadalane STs, including α-amorphene, γ-cadinene, α-ylangene, (+)-sativen, cadina-1,4-diene, epizonarene, α-muurolene, and δ-cadinene (Table S5). Primary sequence conservation, including active site residues in the three functionally important motives, and similar expression patterns shared by *LbSTS4a*, *b*, and *c* (Table S3, Figs. 5 and S1A; Basso et al., 2020; Plett et al., 2015) suggest that STSs encoded by these genes are likely to synthesize the same product(s). This prediction was confirmed experimentally. However, instead of the cadalane STs predicted, heterologously expressed LbSTS4 proteins produced an acyclic terpene. After three hours of incubation with FPP, formation of the sesquiterpene alcohol nerolidol (calculated RI 1551) was detected for all three proteins) but not for any of the control samples, including FPP substrate, protein extracts

from the wild type *E. coli*, and *E. coli* containing empty pDEST17 vector (Table S6, Figs. 6A and S4). Nerolidol has not been detected among VOCs emitted by the axenic cultures of *L. bicolor* (Ditengou et al., 2015; Guo et al., 2021; Müller et al., 2013). This fact suggests that nerolidol biosynthesis may occur specifically upon the contact of the *L. bicolor* mycelium with the plant-symbiont roots. The *LbSTS4* gene expression patterns support this hypothesis (Fig. 5).

Another gene that shows a significant expression upregulation in *L. bicolor* during the host colonization, *LbSTS6* (Figs. 5 and S3), belongs to the clade B-II, which includes STSs of the N:C1,C6 reaction route characterized by high product specificity (Agger et al., 2009; López-Gallego et al., 2010b; Wawrzyn et al., 2012). *LbSTS6* shares high sequence similarity with its sister gene *XP001886282*. According to our bioinformatic prediction, these two genes are likely to be in charge of the biosynthesis of four STs emitted by *L. bicolor*, including (-)-α-cuparene, thujopsene, cis-γ-bisabolene, and acoradiene (Table S5; Ditengou et al., 2015; Müller et al., 2013). The fact that the two genes showed different expression patterns (Fig. 5; Basso et al., 2020; Miyauchi et al., 2020) suggests that they are likely to produce different STs from the list above. Because α-cuparene is a product of α-cuprenene oxidation (Dauben and Oberhänsli, 1966; López-Gallego et al., 2010a), one would expect *L. bicolor* to emit both of these STs. However, α-cuprenene emission has not been reported for *L. bicolor in vivo*. In our experiments, two ST peaks were detected when the purified LbSTS6 protein was incubated with FPP: a major peak (87.6 ± 1.3 %) at RT of 21.41 min (calculated RI=1512) and a minor one (12.4 %) at RT of 21.5 min (calculated RI 1516). The minor peak corresponds to α-cuparene, as shown by comparison to an authentic standard, while the major one was tentatively annotated by spectra match to the NIST20 database as either β-himachalene or α-cuprenene, as both STs have identical MS spectra (Table S6). Using two control genes, a canonical α-cuprenene synthase *Cop6* from *C. cinereus* (Agger et al., 2009) and β-himachalene synthase *HcS* (WP035852539) from *Cryptosporangium arvum* (Rinkel and Dickschat, 2019), we discriminated between the two STs and confirmed α-cuprenene as the major product of LbSTS6 (Fig. 6B). Notably, the list of STs detected in the *L. bicolor* culture using GC-MS includes β-himachalene (Müller et al., 2013). However, our bioinformatic analyses did not assign β-himachalene to any *L. bicolor* gene (Table S5). Taken together, results of our experiments and bioinformatic analyses confirmed that α-cuprenene is one of the major ST compounds in the *L. bicolor* emission profile, which was miss-annotated in our previous study by Müller *et al.* (2013) as β-himachalene due to the absence of an authentic standard and reliance solely on spectra match identification.

Results of these experiments confirmed our method prediction capability and limitations inferred using *in silico* data analyses.

4. Discussion

4.1.   Fungal sesquiterpene synthases that belong to one phylogenetic clade are likely to share their primary cation selectivity and initial cyclization reaction

Predicting gene functions based on the sequence similarity to their homologs from closely related species is a fundamental genomics approach. However, in the case of STSs, promiscuous enzymes characterized by high catalytic plasticity, the homology-based approach has limited prediction power. In this study, we evaluated the utility of different types of data and developed a method for the computational prediction of the *STS* gene functions in fungi based on the primary gene sequence information, species-specific ST profiles, and gene expression information. Phylogenic trees including putative and functionally characterized fungal *STS* genes (Figs. 2 and S1) provide the main framework for these predictions. Analyses of distribution across the phylogenetic trees of the STs produced by functionally characterized

STSs show that the fungal STSs encoded by genes sharing high aa sequence similarity (and therefore belonging to one monophyletic clade) may produce different STs. Nevertheless, such STSs tend to be related by the parent cation and primary cyclization reaction on their ST biosynthetic pathways (i.e., reaction route) – the main criteria for ST classification by reaction classes (RCs) in Figure 1. For example, four different STs produced by the functionally characterized STSs from Clade B-II.2: α-cuprenene (Cop6, Fompi1, and LbSTS6 from *C. cinereus*, *Fomitopsis pinicola*, and *L. bicolor,* respectively), α- and γ-barbatene, and trans-dauca-4(11),8-diene (Omp9 and Omp10 from *O. olearius,* respectively), - derive from the nerolidyl cation *via* the C1,C6 primary cyclization reaction (Agger et al., 2009; Wawrzyn et al., 2012; this study). This rule remains true even in the case of the *STS* genes acquired *via* the interphylum HGT, such as *Dia1* gene of the Basidiomycota origin in the Ascomycota species *Diaporthe* sp. (clade A-III.1; de Sena Filho et al., 2016). This trend is consistent with the hypothesis that a small change in the aa sequence or even in the microenvironment may be sufficient to cause an alteration in the STS product profile within one reaction route (López-Gallego et al., 2010a), while a structural STS transformation leading to the reaction route switch at the initial steps of the ST biosynthesis requires a greater mutation accumulation. Based on this principle, we extended RCs inferred from the functionally characterized fungal STSs to their putative homologs within corresponding monophyletic clades (Table S1). Such classification restricts the list of candidate STs for each clade to those that belong to the RC(s) assigned to it.

4.2.   Species-specific sesquiterpene profiles and gene expression data complement the phylogeny-based predictions of the sesquiterpene synthase gene functions

The joined phylogenetic analysis of putative and functionally characterized *STS* genes allowed us to predict the *STS* gene products only to the level of specific RCs, each of which consists of dozens of STs. Information on the species-specific ST profiles apperas to be particularly useful to further narrow lists of candidate STs for each gene within each RC to those that are actually known to be produced by a particular species. Using the RC-based classification described above, we matched the STs emitted by a species to its *STS* genes. From this, 73% of putative *STS* genes included in our phylogenetic analyses were affiliated with at least one ST reported for the species of interests (Tables S4 and S5). Moreover, 72% of the STs affiliated with the putative *STS* genes were assigned to a single phylogenetic clade. Most STs that received ambiguous assignments to genes from several phylogenetic clades were either of the cadalane type or acyclic. Since the cadalane type STs may be synthesized *via* several alternative pathways, the ambiguous predictions generated could be accurate. For example, in *Clytocybe aegerita*, muurolene and cadalane isomers are produced by functionally characterized STSs from four different phylogenetic clades, including A-III.1b, A-III.2a, and A-III.3a and c (*via* N:C1,C10 and F:C1,C10 reaction routes, respectively; Zhang et al., 2020b). Notably, the size of product profiles predicted for the putative *STS* genes based on the RCs and species-specific STs fall exactly within the normal product range of fungal STSs (1-14 STs; Table S1). Verification of the ST-to-STS assignments using the actual ST profiles for previously functionally characterized *STS* genes from *H. annosum*, *P. croceum*, and *P. placenta* and *STS* genes from *L. bicolor* functionally characterized in this study showed that all cyclic STs of non-cadalane type known to be produced by these enzymes were correctly linked to the clades containing corresponding *STS* genes.

This method, nevertheless, has certain limitations. Firstly, this method cannot unambiguously predict STSs producing acyclic STs. For example, according to our bioinformatic predictions, enzymes encoded by the *L. bicolor* genes *LbSTS4a-c* (A-III.2) were expected to produce cadalane STs *via* the N:C1,C10 route. Experimentally however, we found

that these enzymes produce acyclic nerolidol *in vitro* (Figs. 6 and S4). The acyclic STs farnesene and nerolidol are derived from primary cations *via* proton loss or a reaction with water molecules at the early stage of the FPP conversion, which is shared by multiple ST biosynthetic pathways (Fig. 1). Because of this, acyclic STs are often present in the product profiles of promiscuous STS from different RCs along with cyclic STs, such as cadalanes (Table S1). We speculate that the biosynthesis of an acyclic ST as the main or only product of an STS may occur in STSs from different RCs as the result of complete termination of the cyclization reactions due to mutations affecting protein folding. The occurrence of the STSs catalyzing the conversion of FPP to nerolidol as the main product has been previously reported for fungal STSs *ApLNS*, *GmLNS*, *HsLNS*, and *AaLNS* (*Agrocybe pediades*, *Galerina marginata*, *Hebeloma cylindrosporum*, and *C. aegerita*, respectively) from Clade A-III.1b (N:C1,C10 reaction route; Ichinose and Kitaoka, 2018; Zhang and Fernie, 2021). Such distribution of the acyclic STSs across the *STS* gene tree confirms that the STS shift from cyclic to acyclic products occurred in several phylogenetic clades independently.

Another limitation is that this method does not differentiate between the product profiles of gene copies which belong to the same species and form a single phylogenetic clade. In this study, we demonstrate that gene expression data may partially compensate for this limitation. We reconstructed the *STS* gene expression profiles in response to different types of biotic interactions in Basidiomycota species, in which recent within-clade gene duplications are especially abundant. The comparison of expression profiles of the functionally characterized *STS* genes resulting from such duplications showed that closely related genes that have similar expression profiles are likely to share functions. The three *LbSTS4* genes from *L. bicolor,* which according to the results of our experiments produce the same ST, showed similar expression responses to the mycelium interaction with the plant-symbiont roots, whereas *P. placenta* STSs known to produce different STs showed different expression profiles during several stages of the wood colonization (Fig. 5). Therefore, the comparison of expression profiles of closely related putative and functionally characterized *STS* genes may enable further specification of predicted ST profiles.

Gene expression data was found to be particularly useful in identifying *STS* genes and STs that may play a role in the antagonistic or mutualistic interactions between fungi and other species. In this study, we paid particular attention to the *STS* genes which, according to the DGE analyses, are involved in the symbiotic interactions between the mycorrhizal fungus *L. bicolor* and its host plants.

4.3.   Nerolidol and α-cuprenene may play different roles in the Laccaria bicolor-plant symbiosis

Ditengou *et al.* (2015) demonstrated that at the early, pre-colonization stage of mycorrhiza development, (-)-thujopsene has a signaling role in the establishment of a dialog between the mycorrhizal fungus *L. bicolor* and its host plants. This study largely relied on the VOC emission profiles of the axenic fungal cultures. Once a direct contact between the fungus and plant roots is established, it is difficult to discriminate between the changing VOC profiles of the interacting partners. In this situation, gene expression data can be particularly useful. In the present study, we characterized functions of four *L. bicolor STS* genes, which according to their expression patterns (Basso et al., 2020; Miyauchi et al., 2020; Plett et al., 2015) are important players in mycorrhiza development at the root colonization stage. Using GC-MS analyses, we found that three of these genes, *LbSTS4a*, *b*, and *c*, encode STSs catalyzing biosynthesis of nerolidol (Fig. 6A). These genes showed up to 16-fold expression induction during the first weeks of symbiotic interactions between *L. bicolor* and three plant species: *P. tremula* x *alba* (Basso et al., 2020), *P. trichocarpa*, and *Pseudotsuga menziesii* (Plett et al., 2015). Induction

of the *LbSTS4* genes upon contact with the host-plant roots and absence of nerolidol from the VOC profiles of free-living *L. bicolor* mycelium (Ditengou et al., 2015; Guo et al., 2021; Müller et al., 2013) implies that the STS4 enzymatic products may specifically be involved in the plant-fungus dialog during mycorrhiza formation. In *Arabidopsis*, an exogenous application of nerolidol was shown to inhibit primary and lateral root growth and induce morphological changes in the root system by altering auxin balance (Landi et al., 2020). Interestingly, the establishment of mycorrhiza with *L. bicolor* has similar effects on the root morphology and auxin metabolism and signaling in poplar (Vayssieres et al., 2015). Based on the result of analyses of the mycorrhiza-induced changes in the *Populus* spp. Roots, Vayssières *et al.* (2015) predicted the existence of "an unknown signal from *L. bicolor* exudates or diffusible molecules" which regulate the symbiosis-specific auxin signaling by modifying the expression of genes associated with auxin biosynthesis. We speculate that nerolidol produced by the LbSTS4 enzymes represents a strong candidate for such a signal molecule.

The fourth gene functionally characterized in this study, *LbSTS6*, encodes a cuprenene synthase catalyzing biosynthesis of the cyclic STs α-cuprenene and α-cuparene (Fig. 6B). Analyses of the RNA-Seq data from Basso *et al.* (2020) showed about four-fold upregulation of the *LbSTS6* gene two weeks after the inoculation of *Populus tremula* x *alba* roots with *L. bicolor*. In contrast, in the microarray data from Plett *et al.* (2015), this gene is significantly downregulated during the same stage of the *P. trichocarpa* and *P. menziesii* root colonization. These contradictory results cast some doubt on the role of α-cuprenene and α-cuparene in the fungus-plant dialog. Recently, Stöckli *et al.* (2019) proved the role of the α-cuprenene synthase gene *Cop6*, a homolog of *LbSTS6* from *C. cinereus,* in the biosynthesis of the antibacterial ST lagopodin B and demonstrated that *Cop6* is induced specifically in hypha triggered by Gram-positive bacteria. Lagopodin B selectively inhibited growth of Gramm-positive (but not Gramm-negative) bacteria. These observations point to the possible role of the cuprenene synthase LbSTS6 in structuring the bacterial community in the mycorrhizosphere (Shirakawa et al., 2019). The possibility remains that the induction of *LbSTS6* expression at the mid-stage of mycorrhiza establishment with *P. tremula* x *alba* might have resulted from some unknown experimental factor which triggered defense mechanisms in the fungal symbiont.

The roles of nerolidol in the regulation of the fungus-plant symbiosis and α-cuprenene derivatives in the mycorrhiza defense have to be tested experimentally.

5. Conclusion

The quantity of genomic, transcriptomic, and metabolomics/ volatilomic data grows ever larger. Integrated computational analyses of the omics data provide scientists with an enhanced power to discover natural patterns and trends in biological systems. In the present study, we combined different types of omics data to improve the predictability of the products of highly promiscuous group of fungal enzymes sesquiterpene synthases. Despite some limitations, our approach proved to be a valuable tool for not only predicting functions of fungal sesquiterpene synthase genes, but also identifying genes and sesquiterpenes potentially involved in mutualistic or antagonistic interactions between fungi and other species. This method would benefit greatly from a curated database of fungal VOCs and non-volatile secondary metabolites. Unlike genomic and transcriptomic data, metabolomic and volatilomic data remain scattered across hundreds of publications. The present study focuses on 30 fungal species, for which we have previously generated volatilomic data (Guo et. al, 2021). A high-quality fungal VOC database would allow this analysis to be extended to many species included in the 1000 Fungal Genomes Project.

Credit authorship contribution statement

Tetyana Nosenko: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Ina Zimmer:** Investigation, Writing – review & editing. **Andrea Ghirardo:** Investigation, Formal analysis, Writing – review & editing**. Tobias G. Köllner:** Investigation, Writing – review & editing. **Baris Weber:** Investigation, Writing – review & editing. **Andrea Polle:** Conceptualization, Funding acquisition, Writing – review & editing. **Maaria Rosenkranz:** Conceptualization, Writing – review & editing. **Jörg-Peter Schnitzler:** Project administration, Conceptualization, Funding acquisition, Writing – review & editing.

Acknowledgments

Appendix A

Data supporting the results of this study are available as Supplementary Information files.

References

1000 Fungal Genomes Projects, 2018. http://1000.fungalgenomes.org/home/

Agger, S., et al., 2009. Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. Mol Microbiol. 72 (5)**,** 1181-95. doi: 10.1111/j.1365-2958.2009.06717.x.

Basso, V., et al., 2020. An ectomycorrhizal fungus alters sensitivity to jasmonate, salicylate, gibberellin, and ethylene in host roots. Plant Cell Environ. 43(4)**,** 1047-68. doi: 10.1111/pce.13702.

Bolger, A. M., et al., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15)**,** 2114-20. doi: 10.1093/bioinformatics/btu170.

Breitmaier, E., 2006. Terpenes: flavors, fragrances, pharmaca, pheromones. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. doi: 10.1021/np078143n.

Burkhardt, I., et al., 2019. Mechanistic characterization of three sesquiterpene synthases from the termite-associated fungus *Termitomyces*. Org Biomol Chem. 17(13)**,** 3348-55. doi: 10.1039/c8ob02744g.

Dai, W., et al., 2018. The algicidal fungus *Trametes versicolor* F21a eliminating blue algae via genes encoding degradation enzymes and metabolic pathways revealed by transcriptomic analysis. Front Microbiol. 9**,** 826. doi: 10.3389/fmicb.2018.00826.

Dai, Q., et al., 2021. Sesquiterpenoids Specially Produced by Fungi: Structures, Biological Activities, Chemical and Biosynthesis (2015-2020). J Fungi (Basel). 7. doi: 10.3390/jof7121026.

Dauben, W., Oberhänsli, P., 1966. Constituents of hiba wood oil. The isolation and synthesis of two isomeric cuprenenes. J. Org. Chem. 31(1)**,** 315-317. doi: 10.1021/jo01339a502.

de Sena Filho, J. G., et al., 2016. Genome of *Diaporthe* sp. provides insights into the potential inter-phylum transfer of a fungal sesquiterpenoid biosynthetic pathway. Fungal Biology 120(8)**,** 1050–1063. doi:10.1016/j.funbio.2016.04.001.

Dichtl, K., et al., 2010. Farnesol misplaces tip-localized Rho proteins and inhibits cell wall integrity signalling in *Aspergillus fumigatus*. Mol Microbiol. 76(5)**,** 1191-204. doi: 10.1111/j.1365-2958.2010.07170.x.

Ditengou, F. A., et al., 2015. Volatile signalling by sesquiterpenes from ectomycorrhizal fungi reprogrammes root architecture. Nat Commun. 6**,** 6279. doi: 10.1038/ncomms7279.

Dobin, A., et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1), 15-21. doi: 10.1093/bioinformatics/bts635.

Durairaj, J., et al., 2021. Integrating structure-based machine learning and co-evolution to investigate specificity in plant sesquiterpene synthases. PLoS Comput Biol. 17(3), e1008197. doi: 10.1371/journal.pcbi.1008197.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5), 1792-97. doi: 10.1093/nar/gkh340.

Flynn, C. M., Schmidt-Dannert, C., 2018. Sesquiterpene synthase-3-hydroxy-3-methylglutaryl coenzyme A synthase fusion protein responsible for hirsutene biosynthesis in *Stereum hirsutum*. Appl Environ Microbiol. 84(11). doi: 10.1128/AEM.00036-18.

Guo, Y., et al., 2021. Volatile organic compound patterns predict fungal trophic mode and lifestyle. Commun Biol. 4(1), 673. doi: 10.1038/s42003-021-02198-8.

Hansson, D., et al., 2012. Sesquiterpenes from the conifer root rot pathogen *Heterobasidion occidentale*. Phytochemistry. 82, 158-165. doi: 10.1016/j.phytochem.2012.06.024.

HMMER: biosequence analysis using profile hidden Markov models. v.3.2.2 (2020). http://hmmer.org/

Hong, Y. J., Tantillo, D. J., 2014. Branching out from the bisabolyl cation. Unifying mechanistic pathways to barbatene, bazzanene, chamigrene, chamipinene, cumacrene, cuprenene, dunniene, isobazzanene, iso-gamma-bisabolene, isochamigrene, laurene, microbiotene, sesquithujene, sesquisabinene, thujopsene, trichodiene, and widdradiene sesquiterpenes. J Am Chem Soc. 136(6), 2450-63. doi: 10.1021/ja4106489.

Ichinose, H., Kitaoka, T., 2018. Insight into metabolic diversity of the brown-rot basidiomycete *Postia placenta* responsible for sesquiterpene biosynthesis: semi-comprehensive screening of cytochrome P450 monooxygenase involved in protoilludene metabolism. Microb Biotechnol. 11(5), 952-65. doi: 10.1111/1751-7915.13304.

Jia, Q., et al., 2019. Terpene synthase genes originated from bacteria through horizontal gene transfer contribute to terpenoid diversity in fungi. Sci Rep. 9(1), 9223. doi: 10.1038/s41598-019-45532-1.

Kelil, A., et al., 2008. CLUSS2: an alignment-independent algorithm for clustering protein families with multiple biological functions. Int J Comput Biol Drug Des. 1(2), 122-40. doi: 10.1504/ijcbdd.2008.020190.

Kohler, A., et al., 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet. 47(4), 410-415. doi: 10.1038/ng.3223.

Kramer, R., Abraham, W., 2012. Volatile sesquiterpenes from fungi: what are they good for? Phytochem Rev. 11, 15-37. doi: 10.1007/s11101-011-9216-2.

Landi, M., et al., 2020. Phytotoxicity, morphological, and metabolic effects of the sesquiterpenoid nerolidol on *Arabidopsis thaliana* seedling roots. Plants (Basel). 9(10). doi: 10.3390/plants9101347.

Leeper, F., Vederas, J., 2000. Biosynthesis: aromatic polyketides, isoprenoids, alkaloids. Springer, Berlin.

Liao, Y., et al., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30(7), 923-30. doi: 10.1093/bioinformatics/btt656.

López-Gallego, F., et al., 2010a. Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus:* Catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric Isomers. Chembiochem. 11(8), 1093-1106. doi: 10.1002/cbic.200900671.

López-Gallego, F., et al., 2010b. Selectivity of fungal sesquiterpene synthases: Role of the active site's H-1 alpha loop in catalysis. Applied and Environmental Microbiology. 76(23), 7723-33. doi: 10.1128/Aem.01811-10.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12)**,** 550. doi: 10.1186/s13059-014-0550-8.

Lynch, M., Conery, J. S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290(5494)**,** 1151-55. doi: 10.1126/Science.290.5494.1151.

Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3.61. http://www.mesquiteproject.org

Mischko, W., et al., 2018. Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective beta-copaene and cubebol production in *E. coli*. Microb Cell Fact. 17(1)**,** 164. doi: 10.1186/s12934-018-1010-z.

Miyauchi, S., et al., 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat Commun. 11(1)**,** 5125. doi: 10.1038/s41467-020-18795-w.

Moujir, L., et al., 2020. Applications of sesquiterpene lactones: a review of some potential success cases. Applied Sciences. 10**,** 3001. doi: 10.3390/app10093001.

Müller, A., et al., 2013. Volatile profiles of fungi-chemotyping of species and ecological functions. Fungal Genet Biol. 54**,** 25-33. doi: 10.1016/j.fgb.2013.02.005.

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme nomenclture. Supplement 24*,* 2018, https://iubmb.qmul.ac.uk/enzyme/supplements/sup2018/

Ohno, S., 1970. *Evolution by gene duplication.* Springer, Berlin (Germany).

Plett, J. M., et al., 2015. The mutualist *Laccaria bicolor* expresses a core gene regulon during the colonization of diverse host plants and a variable regulon to counteract host-specific defenses. Mol Plant Microbe Interact. 28(3)**,** 261-73. doi: 10.1094/MPMI-05-14-0129-FI.

Presley, G. N., et al., 2020. Functional Genomics, Transcriptomics, and Proteomics Reveal Distinct Combat Strategies Between Lineages of Wood-Degrading Fungi With Redundant Wood Decay Mechanisms. Front Microbiol. 11**,** 1646. doi: 10.3389/fmicb.2020.01646.

Quin, M. B., et al., 2014. Traversing the fungal terpenome. Nat Prod Rep. 31(10)**,** 1449-73. doi: 10.1039/c4np00075g.

Quin, M. B., et al., 2013. Mushroom hunting by using bioinformatics: application of a predictive framework facilitates the selective identification of sesquiterpene synthases in basidiomycota. Chembiochem. 14(18)**,** 2480-91. doi: 10.1002/cbic.201300349.

Quintana-Rodriguez, E., et al., 2018. Shared weapons in fungus-fungus and fungus-plant interactions? Volatile organic compounds of plant or fungal origin exert direct antifungal activity in vitro. Fungal Ecology. 33**,** 115-121.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2020. http://www.R-project.org/

Reddy, G. K., et al., 2020. Exploring novel bacterial terpene synthases. PLoS One. 15(4)**,** e0232220. doi: 10.1371/journal.pone.0232220.

Rinkel, J., Dickschat, J. S., 2019. Mechanistic investigations on multiproduct beta-himachalene synthase from *Cryptosporangium arvum*. Beilstein J Org Chem. 15**,** 1008-19. doi: 10.3762/bjoc.15.99.

Robinson, J. T., et al., 2011. Integrative genomics viewer. Nat Biotechnol. 29(1)**,** 24-26. doi: 10.1038/nbt.1754.

Rynkiewicz, M. J., et al., 2001. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. Proc Natl Acad Sci U S A. 98(24)**,** 13543-48. doi: 10.1073/pnas.231313098.

Schnitzler, J. P., et al., 2005. Biochemical properties of isoprene synthase in poplar (*Populus* x *canescens*). Planta. 222(5), 777-786. doi: 10.1007/s00425-005-0022-1.

Shen, X. T., et al., 2019. Unusual and highly bioactive sesterterpenes synthesized by *Pleurotus ostreatus* during coculture with *Trametes robiniophila* Murr. *Appl Environ* Microbiol. 85(14). doi: 10.1128/AEM.00293-19.

Shirakawa, M., et al., 2019. Mycorrhizosphere bacterial communities and their sensitivity to antibacterial activity of ectomycorrhizal fungi. Microbes Environ. 34(2), 191-98. doi: 10.1264/jsme2.ME18146.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9), 1312-13. doi: 10.1093/bioinformatics/btu033.

Stöckli, M., et al., 2019. Bacteria-induced production of the antibacterial sesquiterpene lagopodin B in *Coprinopsis cinerea*. Mol Microbiol. 112(2), 605-19. doi: 10.1111/mmi.14277.

Vainio, E. J., et al., 2018. *Heterobasidion* partitivirus 13 mediates severe growth debilitation and major alterations in the gene expression of a fungal forest pathogen. J Virol. 92(5). doi: 10.1128/JVI.01744-17.

van den Dool, H., Kratz, P. D., 1963. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. J Chromatogr. 11, 463-71. doi: 10.1016/s0021-9673(01)80947-x.

Vayssieres, A., et al., 2015. Development of the poplar-*Laccaria bicolor* ectomycorrhiza modifies root auxin metabolism, signaling, and response. Plant Physiol. 169(1), 890-902. doi: 10.1104/pp.114.255620.

Wawrzyn, G. T., et al., 2012. Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. Chem Biol. 19(6), 772-83. doi: 10.1016/j.chembiol.2012.05.012.

Yamada, Y., et al., 2015. Terpene synthases are widely distributed in bacteria. Proc Natl Acad Sci U S A. 112(3), 857-62. doi: 10.1073/pnas.1422108112.

Zhang, F., et al., 2020a. Enzyme promiscuity versus fidelity in two sesquiterpene cyclases (TEAS versus ATAS). ACS Catalysis. 10, 1470-1484. doi: 10.1021/acscatal.9b05051.

Zhang, C., et al., 2020b. *Agrocybe aegerita* serves as a gateway for identifying sesquiterpene biosynthetic enzymes in higher fungi. ACS Chem Biol. 15(5), 1268-77. doi: 10.1021/acschembio.0c00155.

Zhang, J., et al., 2019. Gene regulation shifts shed light on fungal adaption in plant biomass decomposers. mBio. 10(6). doi: 10.1128/mBio.02176-19.

Zhang, Y., Fernie, A. R., 2021. Metabolons, enzyme-enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. Plant Commun. 2(1), 100081. doi: 10.1016/j.xplc.2020.100081.

Zhou, K., Peters, R.J., 2009. Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. Phytochemistry. 70, 366-69.

Figures and figure legends
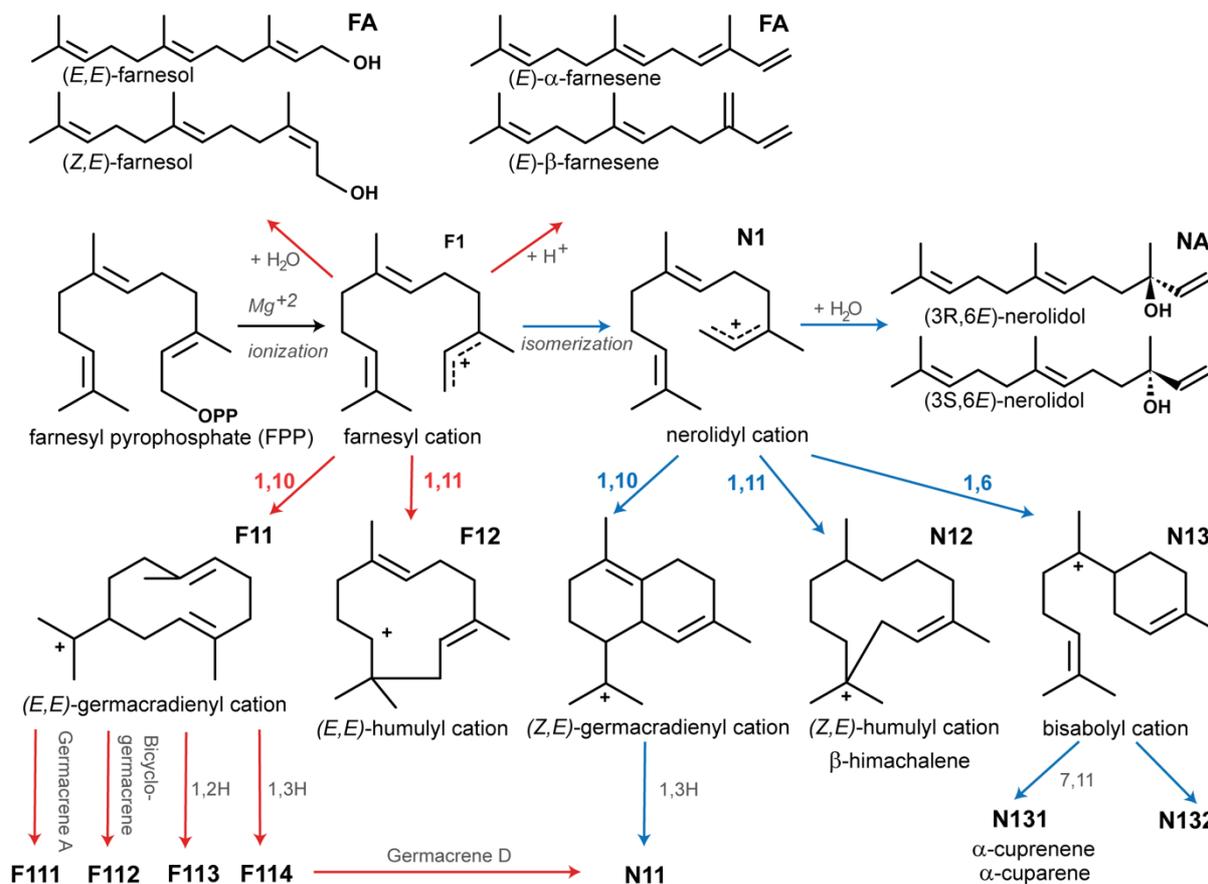
Fig. 1. Scheme of fungal sesquiterpene (ST) classification. Reaction mechanisms of ST biosynthesis are shown starting from farnesyl pyrophosphate (FPP). Loss of the FPP diphosphate moiety leads to the formation of the farnesyl cation (F1), which consequently can be converted to the nerolidyl cation (N1). F1 and N1 can form either acyclic STs (FA and NA, respectively) or cyclic cations: (*E,E*)-germacradienyl and (E,E)-humulyl (F11 and F12) and (*Z,E*)-germacradienyl, (Z,E)-humulyl, and bisabolyl (F11, F12 and N11, N12, N13, respectively) cations, - which in turn undergo complex modifications to form cyclic STs. STs are classified in several Reaction Classes (RCs) based on the initial steps of their biosynthesis depicted in this figure. This classification was adapted from Durairaj et al. (2021) with some modifications specific to fungi (see Materials and Methods). Red and blue arrows correspond to pathways branching from F1 and N1, respectively. Red and blue numbers indicate the initial cyclization reaction. Classification of ST and ST derivatives reported for the fungal species analyzed in this study is provided in Supplementary Data S1.
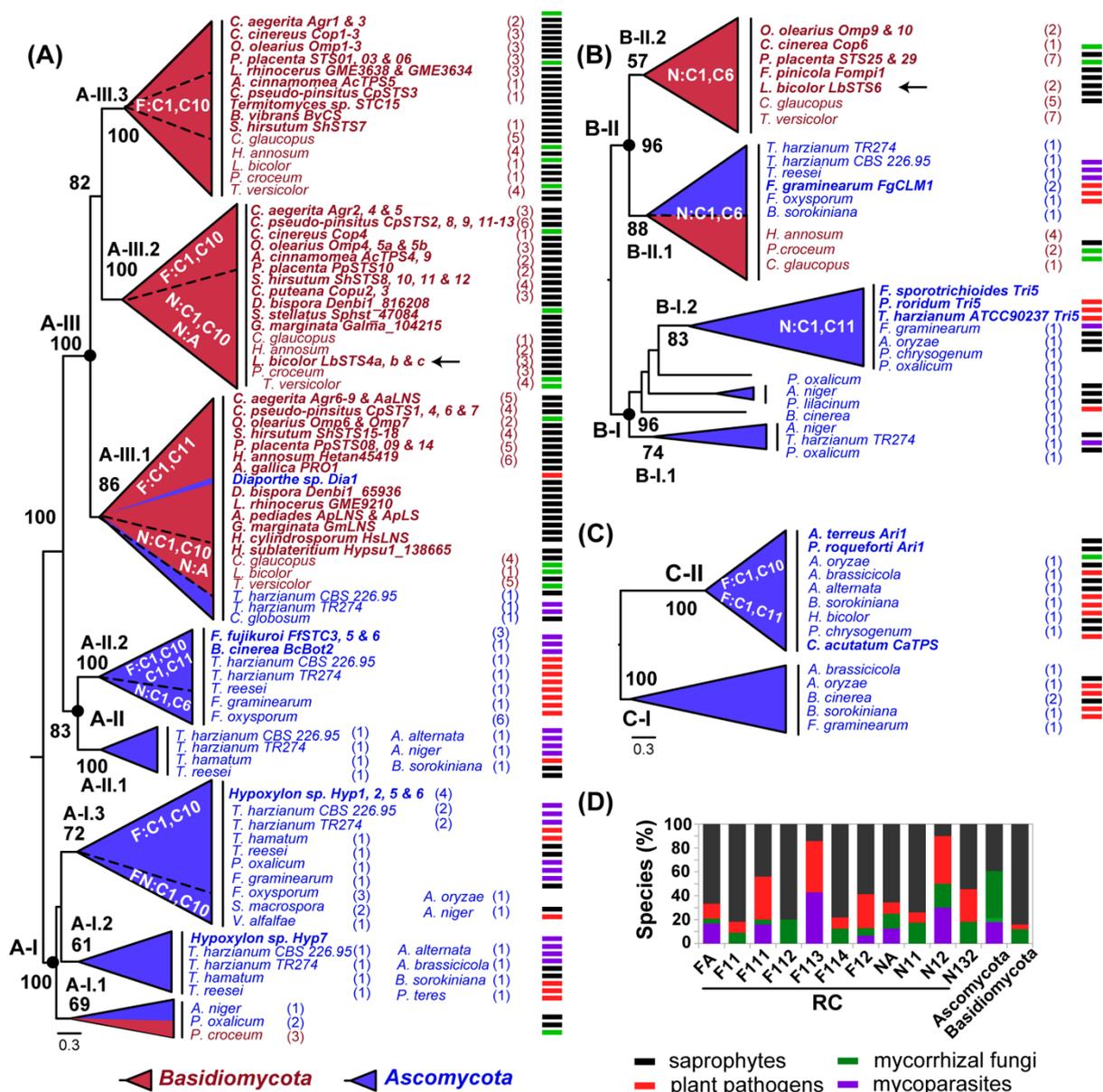
Fig. 2. Schematic representation of the Maximum Likelihood (ML) phylogenetic trees consisting of putative sesquiterpene synthase (*STS*) genes from 30 species included in this study and functionally characterized fungal *STS* genes. Phylogenetic trees were constructed for the three CLASS2 gene clusters that contained genes from more than three species and, among them, at least one functionally characterized *STS* gene. The tree (A) consists of multiple phylogenetically related clades of the class I *STSs*; tree (B) consists of trichodeine synthase and trichodeine synthase-like genes and tree (C) is composed of genes encoding aristolochene synthases and aristolochene synthase-like proteins. A nonredundant list of species is provided for each phylogenetic clade denoted by a vertical line. Numbers in brackets indicate within-clade numbers of genes for each species and strain included in the analysis. Blue clade shading indicates Ascomycota; brown clade shading shows Basidiomycota. Clades that consist of species from both taxonomic groups are bicolored. Phylogenetic clade labels are shown at the corresponding tree nodes. Different sub-clades are separated by dashed lines. Primary cation ('N' for nerolidyl and 'F' for farnesyl cations) and the type of primary cyclization reaction are shown for each sub-clade containing experimentally characterized STS genes. 'A' stands for

acyclic sesquiterpene. Names of the functionally characterized genes are in bold. Arrows point to the genes characterized experimentally in this study. Numbers at the tree nodes indicate bootstrap node support. Only significant node support (≥ 50) is shown. The sequence of colors in the color bars on the right of the trees correspond to the species order in the original trees (Fig. S1) with black color indicating saprophytes, green – mycorrhizal fungi, red – plant parasites, and purple – mycoparasites. Detailed gene and species information is given in Tables S1, S2, and S3. (D) Distribution of *STS* genes from species characterized by different trophic modes across reaction classes (RCs). This distribution plot was constructed using all species and genes from the phylogenetic trees A, B, and C. Colors correspond to trophic modes as described above. RC codes are explained in Figure 1.
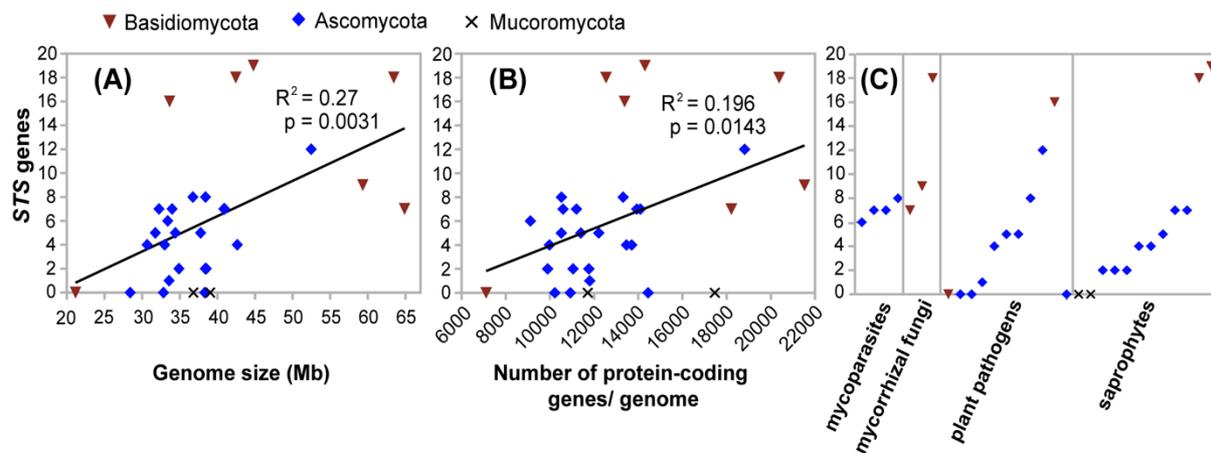

Fig. 3. Correlation between the number of sesquiterpene synthase genes and genome size (A), total number of protein-coding genes (B), and a trophic mode (C) in fungi. Blue diamonds indicate Ascomycota, brown upside-down triangles - Basidiomicota, and black crosses - Mucoromycota species.
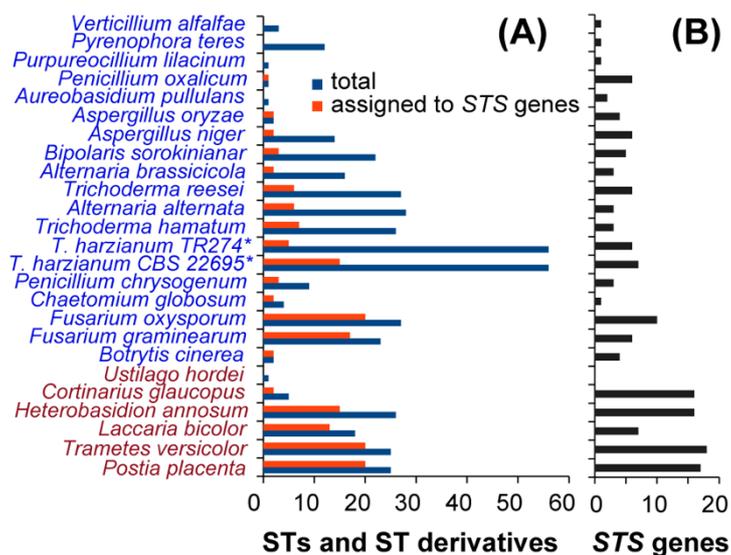
Fig. 4. Distribution of fungal sesquiterpenes (STs) and ST derivatives across sesquiterpene synthase (*STS*) gene models. (A) Blue bars indicate total number of STs and ST derivatives reported for a corresponding species (Table S4); red bars show number of STs assigned to the predicted *STS* genes based on their Reaction Code (RC; Table S5); (B) gray bars in the right panel show total number of *STS* genes in the genome of corresponding species. The names of Ascomycota species are in blue, Basidiomycota – in brown. The data are shown only for the 25 out of 30 fungal species for which information on the ST emission is available (Table S4). For *T. hatzianum,* all STs reported for six different strains were assigned to *STS* genes of the two strains, for which genome sequence assemblies are available (labeled with a star).
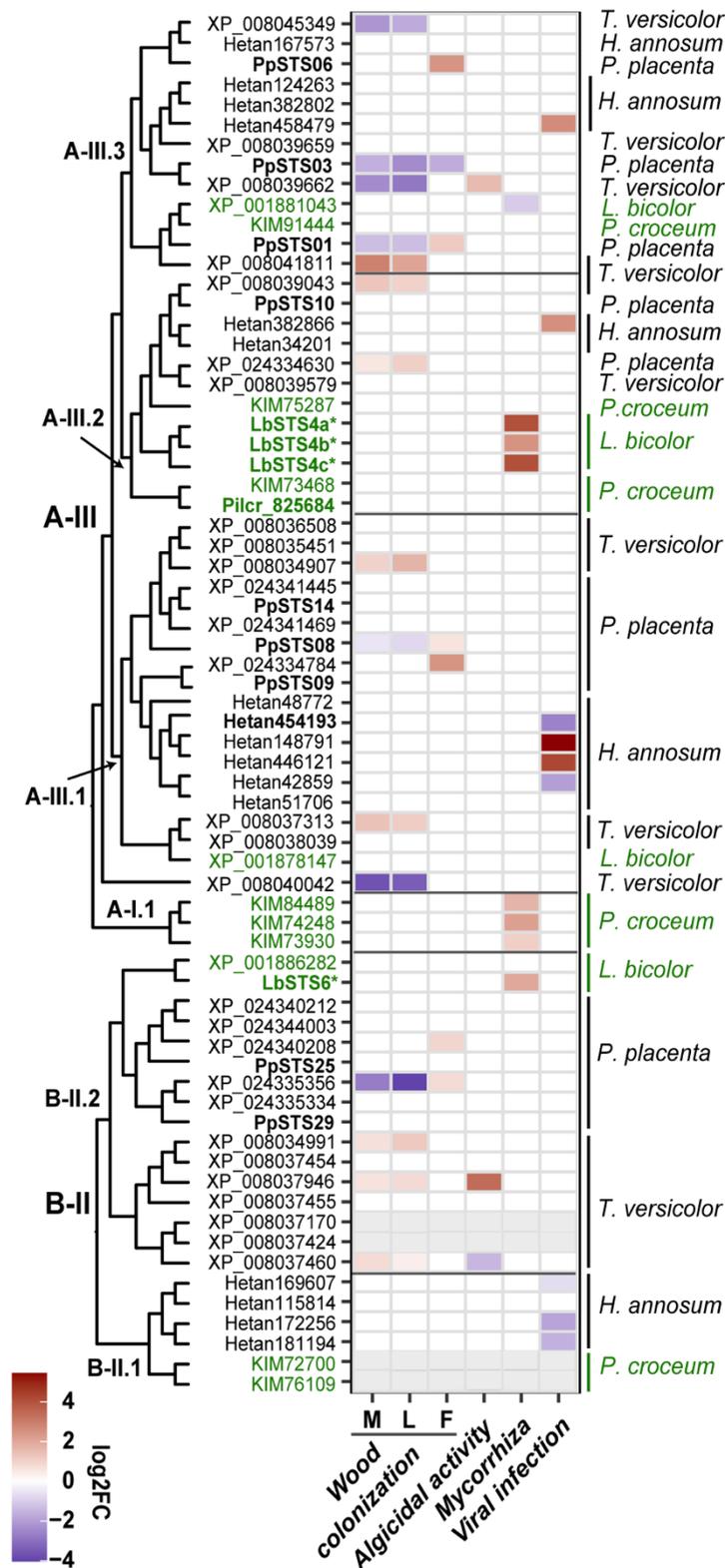
Fig. 5. Response of putative sesquiterpene synthase (*STS*) genes to different types of biotic interactions in Basidiomycota. The Maximum Likelihood phylogenetic tree and expression heatmap are shown for the *STS* genes predicted for the five Basidiomycota species: *C. glaucopus*, *H. annosum*, *L. bicolor*, *P. croceum*, *P. placenta*, and *T. versicolor*. The heatmap shows log2 expression fold change (log2FC) for the *STS* genes significantly ($P_{adj} < 0.05$) up- or down-regulated (red and blue, respectively) in response to: wood colonization (M: middecay, L: late decay; *P. placenta* and *T. versicolor*), fungal competition for the wood colonization (F: *P. placenta*), algacidal activity (*T. versicolor*), establishment of mycorrhiza (*P. croceum* and *L. bicolor*), and viral infection (*H. annosum*). Empty cells indicate that either genes are not significantly differentially expressed under the given experimental conditions or data are not available. Genes with no transcript support are shaded gray. Protein accession numbers in green are shown for the mycorrhizal fungi; black indicates saprophytes. Labels at the tree nodes correspond to the clade and sub-clade labels in the Figures 2 and S1. Protein IDs of functionally characterized STSs are in bold. Stars indicate genes functionally characterized in this study. Information on the RNA-Seq data used for these analyses is provided in Materials and Methods.
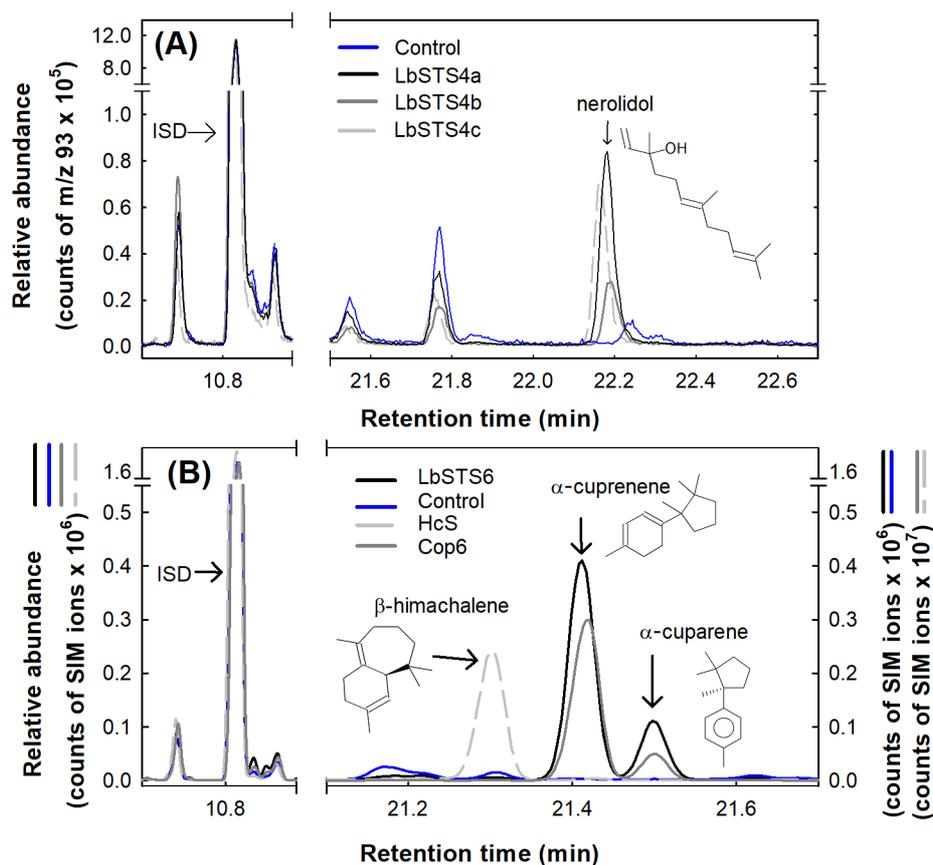
Fig. 6. Functional characterization of four sesquiterpene synthase (*STS*) genes from *L. bicolor* (*LbSTS*). (A) Production of the acyclic sesquiterpenoid nerolidol (RT 22.2) from lbSTS4a-c (solid black, solid and dashed gray lines, respectively) compared to control (blue line). (B) Production of the cyclic α-cuprenene and α-cuparene from LbSTS6 (black line) and Cop6 (dark grey line) and β-himachalene from HcS (gray dashed line) compared to control (blue line). Enzymes heterologously expressed in *E. coli* were purified using the His-Tag purification protocol, assayed *in vitro* by incubation with FPP, and analyzed by SBSE and TD-GC-MS. (a), Extracted ion chromatogram (EIC) at *m/z* 93; (b), single ion mode (SIM) monitored at *m/z* 93+121+136 (until RT 14.9) and *m/z* 119+132+134+145+202+204 until the end of the chromatographic run. Abr.: ISD, internal standard (860 pmol of δ-2-carene). HcS β-himachalene synthase (WP035852539) from *Cryptosporangium arvum* (Rinkel and Dickschat, 2019); Cop6, α-cuprenene synthase from *C. cinereus* (Agger et al., 2009). Analysis of HcS and Cop6 demonstrate that LbSQT6 produces α-cuprenene and α-cuparene. The products of LbSTS4a and LbSTS6 were confirmed independently in the experiments conducted at the MPE (Fig. S4).