Short Thesis for the degree of doctor of philosophy (PhD)

Investigation of *Armoracia rusticana* endophytes in *in vitro* and *in vivo* models

by Tamás Plaszkó

Supervisor: Dr. Sándor Gonda

University of Debrecen
Doctoral School of Pharmaceutical Sciences

Debrecen, 2023

Investigation of *Armoracia rusticana* endophytes in *in vitro* and *in vivo* models

By Tamás Plaszkó, Biology MSc

Supervisor: Sándor Gonda, PhD

Doctoral School of Pharmaceutical Sciences, University of Debrecen

Head of the **Defense Committee**:	Prof. Dr. Árpád Tósaki, PhD, DSc
Reviewers:	Dr. Szabolcs Béni, PhD
	Dr. Éva Julianna Leiter, PhD
Members of the Defense Committee:	Dr. Ida Gálné Miklós, PhD
	Dr. József Jekő, PhD

The PhD Defense takes place at the F202 room of Life Sciences Laboratory building, University of Debrecen, from 13:00 on 13 February, 2024

1. Introduction

"You are what you eat." The originator of this now cliché proverb probably didn't realize at the time that by consuming a single part of a plant, one could consume an entire ecosystem. In fact, research over the last few decades has shown that, under natural conditions, the inner tissues and cells of plants can be home to a diverse community of bacteria and fungi. Typically, these organisms colonize the host plant without any visible symptoms and may even help it to grow, survive pathogens, etc.

In our research work, which is the basis of this thesis, we have investigated these organisms, in particular to answer the question of the molecular level relationships involved in the interactions between endophytic filamentous fungi and the host plant. Our studies focused on the fungal community of horseradish (*Armoracia rusticana*), a member of the Brassicaceae plant family. This plant is not only important for its food and molecular biology (the enzyme horseradish peroxidase is used in a number of immunohistochemical and ELISA methods) uses, but may also be of pharmacological importance due to its chemopreventive properties. The plant is also capable of producing diverse antimicrobial compounds, which raises a number of intriguing questions about how a diverse and complex microbial community can develop in such a harsh chemical environment.

During our work we had the opportunity to investigate these interactions in both in vitro and in vivo models. We have isolated and long-term maintained a number of endophytic filamentous fungi from horseradish roots, and used headspace GC-MS techniques to investigate their ability to produce volatile organic compounds (VOCs) when incubated on horseradish root extracts. In addition, over the past few years we have had the opportunity to collect more than 100 root samples from horseradish cultivars grown in field conditions on an experimental agricultural field to explore in vivo interactions between the fungal community and horseradish metabolites using metabolomics and metagenomics methods.

Among the methods we used, metabolomics, like the other "omics", is an interdisciplinary field, applying tools from several disciplines (instrumental analysis, informatics, data science) to study the complete set of metabolites of a biological system. In contrast to classical analytical methods, which mainly only allow the analysis of a few well-defined compounds, metabolomics allows the relative amounts of up to hundreds of chemical "features", to be determined and even approximately identified by means of computer algorithms, thus providing a wealth of information on the current state of the organisms.

Using these methods, we wanted to answer the following questions in our research: are endophytes or soil-dwelling fungi living in the host plant's environment able to utilize the plant's metabolites? If so, what volatile metabolites might indicate this? Is there a difference between endophytes and soil fungi in the use of compounds produced by the plant? Another key question was whether the chemical composition of the host plant influences the composition of the fungal community living in it? Which compounds or compound classes play a significant role in such interactions?

2. Aims

In the course of our research for my doctoral thesis, we formulated the following objectives:

- We planned to determine the volatile organic compounds (VOCs) in the vapour space of endophytic and soil fungi incubated on plant extracts to investigate whether fungi in the horseradish and its environment are able to use and transform plant-specific metabolites.
- We also planned to explore fungal communities in the roots of horseradish cultivars grown in field conditions using metagenomic (amplicon sequencing) methods, to gain more information on the taxa that make up the endospheres of these plants.
- We planned to perform metabolomic studies on the horseradish roots to map in detail the metabolome of the plant. We also planned to identify the chemical features from metabolomic measurements in more detail, at least at the level of compound classes.
- Finally, we planned to investigate the chemical interactions between the host plant and its microbiome, which compounds are responsible for the assembly of the fungal community and which fungal taxa affect the composition of the fungal community.

3. Materials and methods

3.1. Volatile organic compound (VOC) analysis in the vapor space of horseradish-associated fungi

The horseradish extract was prepared according to the following protocol: approximately 500-1000 g of fresh, healthy horseradish roots were cut into large pieces, boiled in water for 30 min to inactivate the myrosinase enzyme, and homogenized with methanol in a solvent:molar mass ratio of 3:2. The methanolic mixture was boiled for 30 minutes under reflux. The resulting extract was filtered and evaporated to near dryness in a rotary vacuum evaporator and resuspended in approximately the same volume of water as the water content of the roots (typically 70% of the initial weight). After pre-filtration, the liquid was sterile filtered through a 0.20 μm pore diameter PES membrane and stored at -24 ˚C until further use. If necessary, the extract was supplemented with 2% sterile agar.

The fungal strains used in the present work have been previously isolated by the Department of Botany (Szűcs et al., 2018). Soil fungi were isolated from the plant-free areas of the area used for horseradish cultivation. The species-level identification of isolates E1-E7 was previously performed (Szűcs et al., 2018). For the other strains, the taxonomic identification was performed by amplification of their ITS (Internal Transcribed Spacer), α-actin or calmodulin genes and Sanger sequencing. NCBI BLAST searches were performed on the sequences and the results were used to identify fungi up to genus level.

To measure the VOCs emitted by the different fungal strains, the fungi were incubated in autoclaved headspace chromatography vials. 1 mL of warm agar-supplemented extract was pipetted into the vials, which were then spinned until the extract solidified as a thin film layer on the bottom of the vial. A liquid suspension of each fungal strain was prepared for uniform inoculation of the headspace jars. The fungi were cultured in 30 mL MEB medium, shaken at 200 rpm, at room temperature for 7-10 days. The mycelia were homogenized using a MiniMix CC homogenizer in sterile BagPage bags. The resulting suspensions were centrifuged for 2 min at 13500 rpm and the mycelia were washed with sterile water. The fungal suspensions were stored at 4 ˚C prior to inoculation. Since morphology and conidia formation capacity may differ between species, we standardized the inoculation volume on a dry weight basis, which was determined by lyophilization.

For inoculation of headspace vials, 500 μL of 500 μg dry weight equivalent fungal suspension was pipetted into the vials. Finally, the vials were sealed and incubated at room

temperature for at least 3 days. The vapor space of each culture was sampled on days 3 and 4 after inoculation. The experiment was performed with two biological replicates, for a total of 4 replicates per fungus. In order to verify that the sulfur-containing compounds are produced from specialized metabolites in the horseradish extracts, control cultures were also inoculated on MEB medium. These were prepared in one biological replicate and injected on days 3 and 4 after inoculation.

The VOC profile of the fungal vapor space was analyzed using a Bruker Scion 456 gas chromatograph, consisting of a Bruker SHS-40 Headspace sampler and a Bruker SQ mass spectrometer connected to it. A Br-5 capillary column was used for the analyses. The identification of volatile compounds was performed according to the NIST (National Institute of Standards and Technology, Gaithersburg, MD, USA; version 2.0g, build 19.05.2011) spectral library. The performance of the method was also tested with a calibration curve of standard compounds.

All the chromatographic peaks were manually evaluated on the ion chromatograms (EIC, XIC), then the most abundant characteristic ions (max. 5) from the purest spectrum (high peak sample) were manually collected for each peak. Using the list of characteristic ions, a targeted peak detection algorithm was used in mzMine 2.39 software, and the ion abundance values for the same metabolite were summed in R. To detect significant differences between fungi, raw metabolite abundance values were tested in ANOVA models (n = 4 for each fungal strain). Dunnett's post hoc tests were run on significant compounds to obtain statistical differences between individual fungi and controls. To test for statistical differences between endophytic and soil fungi, the presence of each compound in each fungal strain was averaged as a single numerical value and then tested using Fisher's exact test.

3.2. Investigation of metabolome - microbiome correlations in horseradish plants

For conservation purposes, several varieties of horseradish (Armoracia rusticana G.Gaertn., B.Mey. & Scherb.) were grown on an experimental farm in Hajdúhadház-Fényestelep (site 1, 47°39'09.8 "N 21°42'30.5 "E). A total of 13 varieties were sampled (at least four per variety) in November 2018 and 2019. In addition, 4 soil samples were collected in both years. In November 2019, additional samples of carrot (*Daucus carota* L.) and horseradish were collected as external controls from a site in Haláp (site 2). During sampling, leaves were removed from the plants and the roots were placed in sterile plastic bags and transported to the Department for further processing. As a first step in the surface

sterilization procedure, the roots were dipped in 96% ethanol for 30 s, soaked in a solution of 0.1% Tween 20 and 2.5% NaOCl containing active chlorine for 10 min, and then sterilely cut to pieces after sterile washing with water. The resulting sample was placed in sterile plastic sample containers and flash frozen in liquid nitrogen. The samples were stored at -80 ˚C until further processing.

The frozen, cut plant samples were cryogenically ground with a 20 mm steel ball in 50 mL stainless steel sterile grinding jars. Homogenized samples were lyophilized and stored in a dark desiccator on silica gel prior to genomic DNA and metabolomic extractions.

From the cryogenically ground and lyophilized plant samples, 25 mg were weighed and extracted for 5 min with 0.1% formic acid supplemented with 75% methanol at 4 ˚C. Based on our preliminary studies, this extractant gave the highest compound coverage in untargeted metabolomic studies. The extracts were centrifuged at 24000 g at 4 ˚C for 3 min, and 10× or 200× dilutions of the supernatants were prepared with the extractant for untargeted metabolomic analysis and quantification of glucosinolates, respectively. The diluted extracts were filtered through a syringe filter with a pore diameter of 0.22 μm into chromatographic vials prior to instrumental analysis.

LC-MS measurements were performed using a Dionex Ultimate 3000RS UHPLC system with a Thermo Q Exactive Orbitrap mass spectrometer. The ionization used was electrospray ionization (ESI). For untargeted metabolomic measurements, a Kinetex Polar C18 column was used. A volume of 1-1 μL of 10× diluted samples was injected. For the quantification of glucosinolates, a four-point calibration equation was formed from four glucosinolates (sinigrin, gluconasturtiin, glucobrassicin, glucoiberin) present in higher amounts in horseradish, in the concentration range 0.01-5 μg mL^{-1}. Kinetex XB-C_{18} columns were used for quantification measurements. 1 μL of 200× dilution of the horseradish samples and calibration line samples were injected.

Targeted fragmentation of all candidate compounds of interest for further analyses was performed with similar parameters, except for the mass range, which was reduced to only analyze ions that were included in the inclusion list; data from the positive and negative ion modes were recorded separately. A list of candidate compounds for fragmentation was generated based on candidates that passed QC (quality control) screening, and the list was then broken down into inclusion lists such that a maximum of 5 co-eluting candidate compounds were included in a list, thus ensuring good coverage. Based on the overlap of the lists, the most intense 2-5 candidate compounds were selected for fragmentation at 30 normalized collision energy (NCE), with an ion collection time of up to 250 ms.

The raw measurement files were converted to mzXML format and analyzed using the XCMSOnline 2.7.2 (XCMS 1.47.3) platform. For the quantification of glucosinolates, the previously mentioned four-point calibration equation was used for all glucosinolates whose quantification fell within the linear range of the calibration. In such cases, targeted peak detection was performed in mzMine 2.53 software.

The QC (quality control) samples for metabolomics measurements were prepared according to the principle of "long-term reference samples" (Dudzik et al., 2018; Evans et al., 2020), which states that if we are investigating similar samples over a longer period of time (up to several years), we need to prepare a sample mixture that contains a similar matrix as the samples to be measured and can be kept in a chemically stable state throughout the duration of our studies. Accordingly, we mixed concentrated extracts of one sample from each of the 2018 Fényestelep varieties (23 in total) and the 2019 Haláp sampling points (13 in total), then pipetted 50 mL of the mixture into 2 mL cryotubes and stored in liquid nitrogen until further use. For each measurement sequence, 2 mL of QC sample was allowed to thaw and then diluted 10× to form QC samples prior to measurement.

After removal of isotopic and adduct features, the values obtained from the integration of the QC samples were then used to generate candidates that could be measured with reliable linearity and accuracy. We then removed all features with a relative standard deviation (RSD) greater than 30% and used a less frequently applied filter to ensure a strong linear relationship between the signals from the candidate compounds and their concentrations. This was done for all features passing the RSD filter by evaluating the linearity between abundance and concentration values using Pearson correlation. To do this, we used a QC linearity sample set (dilution series, the most concentrated QC sample was twice as concentrated as the real samples) and a blank sample. This way, only features that reacted linearly within the dilution range were retained. Only features with an $R2 > 0.8$ were retained for further analyses, so features present in significant amounts in the blank samples were also discarded.

Finally, a LOESS (locally estimated smoothing) function was fitted to the data for each features separately, based on the QC samples. The assumed theoretical sensitivities between the known values (QCs) were calculated for each metabolite separately using the fitted curve, and the candidate intensities of the real samples were corrected by these values. In practice, this expresses each abundance value as a "fold-change" value, where the reference (1.00) is the abundance of the features in the mean sample, the QC sample. This procedure also allows sequences measured months apart to be pooled and analyzed together without the use of

standards. After this we performed the correction by the dry weight of the extracted plant material.

MS^2 spectra were collected from the raw measurement files using the CluMSID package. For each candidate compound, the 10 purest MS^2 spectra were used to generate a consensus spectrum, which was exported from R and imported into SIRIUS 4.9.9 for annotation. We then used the SIRIUS CSI:FingerID and CANOPUS algorithms to determine the putative structure of organic compounds and the Classyfire hierarchical classes for each feature separately. We also manually evaluated the SIRIUS suggestions against literature data, aiming to achieve Metabolomic Standards Initiative (MSI) identification level 2.

To effectively identify the fungi in horseradish, we had to ensure that the primers commonly used in the literature to amplify ITS only amplified horseradish sequences to a non-intrusive extent. To test this *in vitro*, we amplified the full ITS of horseradish clones, as well as ITS1 and ITS2 sequences, using primer pairs ITS1-F_KYO2 & ITS4, ITS1-F_KYO2 & ITS2_KYO2, and ITS3_KYO2 & ITS4. In order to have a specific primer pair that does not bind in the ITS2 region of the horseradish but is specific for all Ascomycota, Basidiomycota taxa, a new forward primer was designed for the existing ITS4_KYO3 reverse primer. Using the DECIPHER R package, we performed alignments from ITS sequences of previously described fungi of horseradish origin and relevant plants (e.g. *Armoracia*, *Daucus*) and searched for new candidate primers based on consensus sequence details. Primary candidates with practically useful melting temperatures were subjected to further *in silico* analyses. Our designed forward primers in combination with the ITS4_KYO3 reverse primer were tested by PCR *in vitro* on horseradish clones and various fungal gDNAs. According to our tests, the best performing candidate primer was 5' - TTT CAA CAA CGG ATC TCT T - 3', referred to as "ITS3_NOHR". The performance of this primer was also evaluated *in silico* using the UNITE 8.3 fungal ITS database, followed by a pilot Illumina sequencing.

For both plant, soil and "QC" samples, the ITS2 region was amplified using ITS4_KYO3 and our ITS3_NOHR primer. An equimolar mixture of the libraries was sequenced on Illumina MiSeq platform using MiSeq Reagent Kit V3 (2×300 bp paired-end reads, 600 cycles). Pilot sequencing of primer performance was performed using MiSeq Reagent Kit Nano V2 (150 bp paired-end reads, 150 cycles). Library construction and sequencing was performed by the staff of the Genomic Medicine and Bioinformatics Core Facility, University of Debrecen.

After sequencing, the data were demultiplexed and FASTQ files were generated. All further sequence analysis was performed in R using the DADA2 package. At the suggestion

of the authors of DADA2, we used ASVs (amplicon sequence variants) instead of OTUs (Operational taxonomic units), which are commonly used in metagenomics, as they are more advantageous when combining data from different sequencing runs and easier to distinguish individual contaminant sequences identified in control samples. Taxonomic assignment was performed using a "Naive Bayesian" classification algorithm with the DADA2 function "assignTaxonomy". The classification algorithm performed the assignment based on the UNITE 8.3 Fungal ITS database. We then filtered all ASVs that did not reach a minimum bootstrap confidence level of 80% or if they could be identified by the algorithm only at a taxonomic level of "Kingdom" or less. In order to verify the reliability of the method, we also performed NCBI BLASTn searches on the pre-filtered ASVs in the more rigorously controlled but significantly smaller RefSeq database, and then manually checked the taxonomic assignments by DADA2 based on the taxonomic information of the results.

To characterize and compare the alpha and beta diversity of the samples, we considered the type of samples (site 1 or site 2 horseradish, carrot, soil) and the species (site 1 horseradish). The taxonomic richness of the samples was estimated using the ACE index. The diversity of the samples was determined using the Shannon, Dominance (Simpson) and Buzas & Gibson indices. Kruskal-Wallis tests were used to statistically compare the taxonomic richness and diversity of the above-mentioned groups. Beta diversity was estimated using Bray-Curtis similarity, Whittaker diversity and unweighted UniFrac distance. Beta diversity was plotted using principal coordinate analysis (PCoA) based on similarity or distance values. We also performed one-way similarity analysis (ANOSIM) to determine differences within and between groups. For significantly different groups, we used SIMPER (Clarke, 1993) to determine the taxonomic units responsible for the differences. All analyses were performed in Past v4.09 (Hammer et al., 2001) software, except for the unweighted UniFrac analysis, which was performed in R.

Prior to downstream analyses, ASVs with the same taxonomic assignment were pooled and the number of individual ASVs within each pool was not further considered. The data were assessed using abundances pooled at the phylum, order and genus level, and diversity metrics (previously calculated from raw, unpooled ASVs). Where the bootstrap confidence value of the taxonomic assignment was less than or equal to 80%, i.e. the assignment was not reliable at the genus level, the lowest reliable level was used. The latter resulted in mixed identification levels for many ASVs. Correlation analyses of metagenomic and metabolomic candidate compounds were only performed within the Fényestelep cultivar dataset ($n = 8 \times 2 \times 4$), while community differences between sample groups were investigated using both

cultivars, carrot and soil samples, and Haláp horseradish. The metagenomic data are compositional in nature and therefore a clr transformation was performed on the data set.

Significant differences between sample groups and species were determined using principal component regression. For this purpose, the normalized chemical and fungal data were subjected to sparse principal component analysis (sPCA). For chemical and fungal abundance data, the number of principal components was 12 and 6, respectively. Statistical tests were only performed on principal components that covered at least 2.5% of the variance of the total variance of the data set, such principal components were tested using ANOVA models in R. Spearman's correlations between sPCA scores or appropriately scaled data were calculated using the cor.test function in R. All p-values from statistical tests were subjected to Benjamini-Hochberg correction (n = 18004) for false discovery rate (FDR) correction.

4. Results

4.1. Volatile organic compound (VOC) analysis in the vapor space of horseradish-associated fungi

4.1.1. Identification of endophytes and soil fungi

Prior to performing our incubation experiments, we needed to classify fungal isolates taxonomically. Sequencing of marker genes, most commonly ITS for fungi, is now a cheap and easy-to-perform procedure that allows the identification of a large number of environmental fungal isolates at least at genus level using bioinformatics databases. In our work, a total of 43 horseradish endophyte and soil fungi were analyzed and identified at least to genus level.

4.1.2. Volatile organic compounds (VOCs) identified from the vapor space of fungi

Because of the large number of fungi to be tested, an agar film procedure was developed that allowed the parallel incubation and chemical screening of tens of separate cultures. VOCs of several compounds classes were identified from the vapor space of each fungal culture, including esters, short-chain alcohols, a short-chain acid, an aromatic compound, a monoterpene, as well as several sulfur-containing VOCs and nitriles, the latter presumably derived from glucosinolates, found in horseradish extract. We were also able to compare the spectra and retention times of several compounds with those authentic standards, other compounds were identified putatively by NIST spectral library searches.

Of the compounds identified, 2-phenylethyl isothiocyanate, phenyl propionitrile, allyl isothiocyanate and allyl cyanide are clearly degradation products of glucosinolate origin. In addition, 2 other sulfur-containing compounds have been identified, which are also assumed to be of glucosinolate origin. In addition to these, the presence of six ester compounds, ethyl acetate, ethyl propionate, methyl-1-butanol acetate, methyl acetate, methyl formate and propyl acetate, was also detected. Other classes of compounds included two aromatic compounds (benzaldehyde, styrene), an organic acid (acetic acid), a ketone (acetone) and a monoterpene (β-phellandrene). 18 peaks were not identified, mainly due to a low signal-to-noise ratio, resulting in a low Similarity Score.

In our previous studies (Szűcs et al., 2018), we have already attempted to detect, unsuccessfully, the presence of allyl cyanide from fungal vapor using an activated carbon desorption method. However, our inoculation method described in the present work was

excellent for the detection of this compound. Furthermore, it is important to mention that no sulfur-containing degradation products were detected in the vapor space of fungi growing on malt extract medium, thus confirming their glucosinolate origin.

4.1.3. VOC patterns of fungi

As we have seen, we identified alcohols, esters and glucosinolate degradation products from the vapor space of several fungi, but it is important to note that these compounds were also present in trace amounts in the vapor of control, uninoculated horseradish extracts, which may mean that they can also be formed as a result of spontaneous degradation. However, a significant number of endophytic and soil fungi produced significantly higher amounts of these compounds compared to the control.

In the case of acetone, the most prominent strains were E1 - *Fusarium oxysporu*m, E15 - *Phomopsis* sp. and S18 - *Penicillium* sp., which produced significant amounts of acetone, a commonly occurring VOC in fungi such as the endophytic *Muscodor albus* (Strobel, 2011).

Significantly more allyl cyanide was detectable in the vapor space of several *Fusarium* sp. strains (E8, S7 and S19) compared to the control, and in one *Curvularia* soil fungus (S5).

Allyl isothiocyanate, a product of the degradation of sinigrin, the major glucosinolate of horseradish, was significantly more detectable in the vapor space of 3 fungi compared to the control, E12 - *Plectosphaerella* sp., S1 - *Notophoma* sp. and S21 - *Paraphoma* sp. but was detectable in several cases with relatively high variance.

Several compounds were absent in the vapor space of control samples (uninoculated horseradish extract), including benzaldehyde, β-phellandrene, dimethyl sulfide, acetic acid, ethyl acetate, ethyl propionate, methyl 1-butanol, methyl-1-butanol acetate, methyl 1-propanol, methyl acetate, methyl formate, propyl acetate, carbon disulfide and styrene. This makes it certain that these were produced by the fungi during incubation.

Among the sulfur-containing VOCs, carbon disulfide was present in the vapor space of all fungi, but four fungal strains produced it in significant amounts, E17 - *Colletotrichum* sp., E21 - *Plectosphaerella* sp., S16 - *Fusarium* sp. és S18 - *Penicillium* sp. In addition to carbon disulfide, dimethyl sulfide was also detected, but only 4 fungal strains produced it in detectable quantities: E7 - *Oidiodendron cerealis*, E16 - *Cadophora* sp., S3 - *Curvularia* sp. and S26 - *Penicillium* sp.

An additional VOC, probably also containing sulfur based on its fragmentation pattern, was detected in several samples (retention time 3.33 min, m/z = 72), but was not identified in detail. This potentially sulfur-containing but unidentified compound was produced in

significantly high amounts by a large number of strains of the genus *Fusarium* (E1, E3, E8, E9, S10, S11, S12, S15, S19, S25) and by *Macrophomina phaseolina* (E2) compared to the control.

A group of fungi belonging to the genus *Fusarium* (E1, E8, E9, S10, S11, S12, S15, S19, S25) showed a distinct pattern of production of esters such as ethyl acetate. Other esters, including methyl acetate, ethyl propionate, propyl acetate, were also prominent in the production of the strains mentioned above, with significant amounts of these compounds being detectable in all cases. These strains were also notable for the production of alcohols (e.g. methyl-1-butanol).

Among the more abundant compounds, methyl formate, which was produced by strains S13 - *Penicillium* sp., S16 - *Fusarium* sp. and S18 - *Penicillium* sp. in significant amounts compared to the control, and methyl-1-butanol acetate, which was detected in significant amounts in the vapor space of strains E12 - *Plectosphaerella* sp., E17 - *Colletotrichum* sp. and S21 - *Paraphoma* sp. Some less abundant compounds were acetic acid and styrene detected in one *Fusarium* sp. isolate (S11), benzaldehyde in E2 - *Macrophomina phaseolina*, S19 - *Fusarium* sp. strains and β-phellandrene produced by E15 - *Phomopsis* sp. in notable amounts.

4.2. Investigation of metabolome - microbiome correlations in horseradish plants

4.2.1. Glucosinolate content of horseradish varieties

Headspace-GC-MS analysis has shown that several members of fungi isolated from horseradish and its environment are able to utilize and transform compounds found in horseradish. Based on this and our previous results (Szűcs et al., 2018), we hypothesized that the natural, in vivo endosphere of horseradish may contain an even higher proportion of fungi that have similar chemical interactions with the plant. To investigate this, we collected a large number of horseradish roots for 2 years and processed and stored them in a metabolome-sparing manner. Logically, we hypothesized that glucosinolates are the most dominant in horseradish-fungal chemical interactions, as they are precursors of highly diverse antimicrobial compounds (Plaszkó et al., 2022, 2021). As these compounds also constitute the largest part of the specialized metabolite pool of horseradish, we considered it of utmost importance to determine the concentrations of the major glucosinolates in a targeted, quantitative manner. In all the varieties studied, the main glucosinolates were sinigrin and gluconasturtiin, while glucobrassicin and glucoiberin were also quantifiable in smaller

amounts. The dry weight concentration of sinigrin in each variety ranged from 1.14 ± 0.66% to 3.43 ± 0.65%, and that of gluconasturtiin ranged from 0.36 ± 0.18% to 0.67 ± 0.31%. The concentrations of glucoiberin and glucobrassicin ranged from 0.03% to 0.07%. The relative variation in glucosinolate concentrations within varieties was relatively high, 26.9% for sinigrin and 19.8% for gluconasturtiin, but the variation between varieties was smaller than expected. For sinigrin, gluconasturtiin, glucobrassicin, glucoiberin, the differences between the highest and lowest mean values were 3.33, 1.99, 20.5 and 2.81 times, respectively. The concentration of aliphatic sinigrin and glucoiberin was much more affected by the cultivar itself ($p_{uncorr.}$ = 0.0008 and 0.0049) than by glucobrassicin and gluconasturcin ($p_{uncorr.}$ = 0.6510 and 0.0227). In 2019, significantly less sinigrin, glucoiberin and glucobrassicin were synthesized in the varieties ($p_{uncorr.}$ = 5.21E-7, 0.0019 and 0.0024, respectively), while there was no significant year effect on the amount of gluconasturtiin ($p_{uncorr.}$ = 0.2465). We were also able to investigate additional glucosinolates by non-targeted metabolomics on a relative abundance basis.

4.2.2. Untargeted metabolomics

As no significant differences in glucosinolate content were found between horseradish varieties, the role of untargeted LC-MS/MS metabolomics in understanding plant-fungal interactions became even more important. The XCMS online peak-detection algorithm identified a total of 2576 chemical features in positive and negative ion modes, of which 1310 remained after removal of isotope peaks and adducts. The QC samples used in the instrumental analysis were also used to determine the linearity, relative standard deviation and reproducibility of the measurement. These were used to filter out the candidate compounds that showed a maximum relative standard deviation of 30% and a linearity of at least 0.8, allowing a total of 355 features to be further investigated.

MS/MS fragmentation was also performed on 233 features with high within-sample variability. Based on the fragmentation patterns, potential candidate compounds were identified by Sirius software and verified against literature data (MSI - Metabolomics Standards Initiative Level 2 identification). The features include specific metabolites such as: flavonoid glycosides (kaempferol aglycone), polyphenolic compounds (phenylpropanoid and coumarin glycoside), indole derivatives and primary metabolites (phospholipids, amino acid derivatives, peptides).

Candidate compounds that could not be validated from the literature are referred to either by their putative compound class (Classyfire hierarchy, MSI level 3 identification)

(Djoumbou Feunang et al., 2016) or by their m/z - retention time (not identified). At MSI level 3, we were able to rank several features into the following classes of compounds: 2 cyanogenic glycosides, 1 flavonoid glycoside, 1 glucosinolate, 3 lipids (or lipid-like structures), 9 amino acid derivatives, 14 glycosides and 19 other putative aromatic compounds.

The metabolomic results revealed that several varieties had very similar chemical patterns, and a subset was selected to be the chemical side of the data set for determining microbiome-metabolome correlations. Based on the PCA (principal component analysis) plots of the chemical data, we manually selected 8 species that covered the majority of the candidate compounds with high relative variance that were approximately identified. The principal component values were significantly different between the 8 selected varieties (p = 0.0332), indicating that there was a significant chemical difference between the varieties. The distribution of effect size values calculated when comparing different sampling years or varieties can also be used to illustrate metabolomic differences.

By calculating the difference between the values of the varieties with the smallest and the largest amount of a given candidate compound and then dividing the difference by the standard deviation, the effect size values (Cohen's d) per candidate compound were obtained. The median of the effect size values with standard deviation of 1.289 indicated that there is a significant degree of chemical variability between varieties. Compared to the effect size values for the year factor, the median was relatively high at 0.4678, indicating that the chemical composition of horseradish roots may vary to a non-negligible extent from year to year. Of the 359 candidate compounds tested, 59 had an effect size of 1 or more for the year factor, presumably due to differences in weather between 2018 and 2019. As we did not aim to investigate chemical differences between varieties, data were entered into the further data analysis models without grouping by variety or year. The resulting chemical diversity allowed us to find direct correlations between candidate compounds and endophytic fungal abundance values.

4.2.3. Amplicon sequencing

Another layer of our efforts to understand plant-fungus interactions has been to explore the endophytic fungal community in plants. Currently, one of the most suitable methods for this is metagenomics, including high-throughput amplicon sequencing, which can identify tens or hundreds of thousands of fungal sequences per sample. As already shown, ITS can be a good marker gene for fungal identification and is frequently sequenced in metagenomic

studies. Within the ITS, the use of the ITS2 subregion for fungal amplicon sequencing can have important advantages, e.g., less variability in the length of the amplicons compared to the ITS1 subregion, and more universal primary binding sites. Due to its advantageous properties, its use on second-generation sequencing platforms is accepted (Ihrmark et al., 2012; Nilsson et al., 2019; Toju et al., 2012). When amplification of fungal sequences is performed from living plant tissue, it is advisable to use at least one primer that specifically amplifies fungal sequences (Toju et al., 2012).

During our in silico pilot experiment, the universal fungal primers commonly used in the literature were not effective in the horseradish-endophyte system, as the host plant DNA was also highly amplified. As a consequence, bioinformatics methods were used to design a primer that does not have a binding site in the ITS2 region of horseradish, but does in as many fungal regions as possible. We also performed a preliminary sequencing experiment with the ITS3_KYO2/ITS4_KYO3 and fITS7/ITS4_KYO3 primer pairs, where at least 40% of the amplicons were identified as plant-derived ITS sequences according to the UNITE 8.3 eukaryotic database (Abarenkov et al., 2021b). However, using our ITS3_NOHR forward primer and the ITS4_KYO3 reverse primer used in the literature, the percentage of ITS reads with a clear plant origin was 0.2%. Based on our preliminary sequencing results, we used our developed forward primer in further experiments.

In silico, the ITS3_NOHR primer was able to bind to 95.3% of unique Ascomycota ITS sequences and 85.8% of unique Basidiomycota ITS sequences according to the UNITE 8.3 fungal database (Abarenkov et al., 2021a), provided that we allowed up to 1 mismatch error for the alignment algorithm (the most frequent mismatch was in base 6 at the 5' end of the primer). The primer efficiency approximates that of ITS3_KYO2 and fITS7 primers in these fungal strains (Ihrmark et al., 2012).

The screening step removed 0.7% of the total unique ASVs by removing ASVs also found in negative control samples. The ASVs removed presumably included dermatophytic genera of human origin such as *Malassezia*. A much larger impact on the overall data set was the removal of ASVs that did not give meaningful hits in either the UNITE or NCBI RefSeq (fungal ITS) databases, as 78.5% of the unique ASVs (26.35% of the total reads) were removed. Examination of the 50 most common such sequences using NCBI BLAST revealed that most of them are of some plant origin but not ITS sequences. The screening steps resulted in a total of 2673 ASVs, with a median read count of 26116 per sample.

The use of sequencing QC samples proved to be very useful in optimizing the filtering parameters, as all 4 replicates showed almost identical ASV composition with very low

variance. Due to this property, QC samples may be useful in correcting sequencing errors, but this requires further investigation.

4.2.4. Analysis of microbiome diversity

When processing the raw data from amplicon sequencing, we observed that the fungal community data (e.g. individual reads) of the soil samples differed significantly from the plant samples. Furthermore, based on literature data, it could be assumed that the fungal community in soil samples differs significantly from the endosphere of horseradish, and this is supported by the different diversity indices. Diversity metrics were constructed based on 2673 ASVs that passed the preliminary screening steps. For the horseradish samples from site 1, the average ASV richness was 45.2 (range 27-94), for site 2 45.4 (range 24-109), and for the carrot control samples we found an average value of 39.5 (range 29-48). In contrast, the soil samples showed a much higher richness with an average value of 383.6 (range 109-609). ACE and Shannon indices also showed similar dynamics. The Simpson dominance index indicated that some fungal taxa were significantly more abundant in the horseradish samples. Although the overall dominance was relatively low in the carrot and soil samples, the abundance values were much more uniform in these samples compared to the horseradish samples. Irrespective of dominance values, abundance values were less uniform as measured by the Buzas and Gibson indices. Overall, there was significantly more variability in the diversity index values within the group of horseradish samples.

Beta diversity between samples was assessed using Bray-Curtis similarity and Whittaker's diversity values. Both beta diversity metrics showed that soil samples were extremely different from other sample types. In addition, beta-diversity values were higher within the horseradish sample groups than in the other sample sets, with low R values in the ANOSIM tests also supporting this observation. Although the differences between the different sample groups (soil, carrot, horseradish) were much larger than the differences within groups (due to the large distance between soil samples), the horseradish samples showed a high variability for both beta-diversity metrics.

A more sophisticated analysis, an unweighted UniFrac analysis, was also performed to examine diversity. A robust phylogenetic tree is a prerequisite for the UniFrac method, so we constructed several trees to test their reliability. Based on the metrics of the multiple fits, the UniFrac analysis was reliable. UniFrac dyadicity values showed a similar arrangement to previous diversity metrics. The ANOSIM tests of UniFrac distancing values were also comparable to previous results, although showing greater variability within soil samples.

Overall, both alpha and beta diversity metrics indicated significant variability between and within horseradish sample groups, which may be related to variation in the metabolome of horseradish samples.

4.2.5. Endophytic fungal community composition of horseradish varieties

ANOVA analysis revealed significant differences in fungal communities between sample groups at the strain, family and genus level ($p = 10^{-18}$), similar to the beta diversity metrics. In the horseradish samples from site 1, the most abundant ASVs were classified in the fungal orders Cantharellales, Glomerellales, Hypocreales, Pleosporales, Saccharomycetales, Sordariales, which also include several plant-related genera (typical endophyton, epiphyton, pathogen), e.g. *Claviceps*, *Colletotrichum*, *Epichloë*, *Fusarium*, *Rhizoctonia*. The taxonomically less resolvable ASVs were mainly included in the Ceratobasidiaceae, Nectriaceae, Pezizaceae, Sordariomycetes, Ascomycota groups. Carrot control samples from the same area showed a similar taxonomic composition.

In contrast to the plant samples, a significantly different fungal community was observed in the soil samples from site 1, dominated by the orders Filobasidiales, Mortierellales, Hypocreales, Sordariales, Thelebolales, Umbelopsidales. The most common genera were *Humicola*, *Metarhizium*, *Mortierella*, *Penicillium*, *Pseudogymnoascus*, *Solicoccozyma*, *Trichoderma*, and *Umbelopsis*. In addition, unidentified members of the Chaetomiaceae family and other Ascomycota ASVs were abundant. It is important to note that the soil communities did not change significantly between 2018 and 2019, which may indicate that the composition of the communities observed in the plant samples is not primarily influenced by the soil.

In several cases, one or two endophytic taxa were highly dominant in a sample. This phenomenon was mainly observed in horseradish samples but not in carrot, which may be due to the difference in sampling depth (64 + 12 horseradish, 4 carrot), but was also clearly visible when examining the dominance indices. In most cases, the most dominant taxa accounted for 20-40% of all reads (median 33.57%), but in one sample it reached 95.9%. This phenomenon was not so prevalent in soil samples (median 21.61%).

Since the plant-soil differences were much larger than the variability between horseradish samples, we also examined the data from horseradish samples from site 1 in more detail. Several fungal taxa showed high variability between varieties, suggesting to us that they could be used to build good correlation models in the future. Examples of such taxa include Agaricomycetes (0.01% - 6.5%), Morosphaeriaceae and *Melanoleuca* (0 - 10%),

Monosporascus (0.04% - 17.1%) and *Setophoma* (<0.01% - 6.3%). In many cases these differences were correlated with some compound candidate, as will be shown later.

Regarding the composition of the fungal community, significant within-species variability was observed in some cases. For cultivars C, G, I, K and M, the average pattern was stable between the two sampling years, whereas for cultivars A, U and W, considerable variability in the 2-year fungal community pattern was observed. A significant proportion of taxa showing variability showed correlations with chemical features (e.g. Xylariales, Sordariales, Pleosporales, Eurotiales, Capnodiales, etc.).

4.2.6. Microbiome - metabolome correlations

After the individual analyses of the metabolomics and metagenomics data, these data were analyzed together finally. The integration of the different omics data sets, the most complex and difficult to interpret elements of the study of plant-fungus interactions. Fortunately, increasingly sophisticated statistical methods, models and dimensionality reduction techniques can greatly aid the biological interpretation of such complex data sets.

The principal components of the fungal data sets aggregated at the family or genus level showed significant Spearman correlations with different candidate compounds in 37 and 41 cases, respectively. Due to the large number of cases, rather than interpreting PCA loading values, we attempted to examine direct Spearman correlations between clr-transformed fungal and chemical abundance data, an increasingly widely used approach (Quinn and Erb, 2021). One important advantage of this approach is that it does not omit compounds from the analysis that are not correlated with any other (and therefore are in an unstudied principal component dimension), and it also makes interpretation of the data much simpler. The main disadvantage is that a very large number of statistical tests have to be performed, and therefore statistical correction has to be applied to eliminate false positives, which significantly reduces the statistical power. Due to the previously described transformations, the fungal dataset was no longer compositional (Gloor et al., 2017). The metabolomics data were expressed either as true concentrations (glucosinolate content) or as fold-change values relative to the QC of the candidate compounds (values from non-targeted metabolomics), and therefore the metabolomics dataset is not compositional. As a consequence, no correlation artifacts were assumed and the method was found to be usable. Although there was a clear overlap between the family and genus level data sets, where one genus was very dominant within a family, we still managed to find significant correlations of 99 (family level) and 72 (genus level) using this method.

At order level, the most affected fungal strains were observed to belong to the orders Xylariales, Capnodiales, Sordariales and Saccharomycetales. It is important to note that in some samples, the proportion of these reads reached 73.4%, 15.7%, 74.2% and 38.4% of the total reads, respectively. Some other taxa from other groups also showed correlations with some chemical features, e.g. Laboulbeniomycetes, Agaricomycetes and Eurotiales, but these accounted for less than 1% of the reads on average.

When data aggregated at least at genus level were examined, fungi belonging to the taxa *Monosporascus*, *Setophoma*, *Tetracladium*, and Morosphaeriaceae, Agaricomycetes had the greatest impact on plant chemistry. These taxa accounted for 5.6%, 1.3%, 0.3%, 1.5% and 0.8% of the total readings, respectively, with maximum percentages of up to 73.4%, 32.1%, 28.6%, 59.2% and 21.7%. Other fungal taxa also varied, but with smaller effect sizes, e.g. *Fusarium*, *Melanoleuca*, *Brachyphoris*, *Thanatephorus*, and fungi belonging to the Pezizaceae and Pleosporales groups, which accounted for 4.0%, 5.4%, 0.4%, 4.9%, 2.4% and 3.2% of the total readings, respectively. Overall, ASVs, accounting for 35.23% of all reads, were correlated with one or more chemical features.

The list of the most influential candidate compounds includes many natural compounds from different biosynthetic classes. Surprisingly, the main glucosinolates did not result in a crude correlation with any of the fungal groups, despite being found in very high concentrations in horseradish, compared to other cabbage species. When comparing the effect size values of the significant correlations between classes of compounds and fungal orders, it is not possible to clearly select the class of compounds with the greatest effect, although small differences are apparent.

Looking at the genus level data, the hypothesis that glucosinolates play a key role in fungal community formation cannot be supported, at least as much as the effect of flavonoid glycosides or other glycosides that have been identified. In addition, lipids and lipid-like molecules and peptides also appear to play an important role, although the relative standard deviation values for peptides were high. The highest effect size values (absolute effect size > 2) were found for peptide-like compounds, which according to the Canopus algorithm may belong to gamma-glutamyl amino acid derivatives or additional peptides. The other larger group of correlations (abs. effect size > 1.5) also includes primary and specialized metabolite classes such as: flavonoids and other putative glycosides, and a glucosinolate. These show a strong correlation with the abundance of several fungal taxa. In addition to these, some peptides, phospholipids and N,N-(dimethyl)-thiobenzamide are also found in this group. At the order level, the most strongly affected fungal taxa were Xylariales, Capnodiales and

Saccharomycetales, while at the genus level, *Setophoma*, *Monosporascus* and *Tetracladium* were the most affected.

Plotted on a heatmap, we observed that several classes of primary and specialized compounds were involved in these interactions, including lipids, indole derivatives, glycosides and peptides. Compounds with high correlations with one or more fungal taxa may be scattered in multiple clusters on the heat map, implying that the results cannot be explained simply by multi-correlation of compounds.

Our results presented in this thesis suggest that plant-fungal interactions are indeed a web of complex networks, and that the 2 dimensions (metabolomics-metagenomics) we have presented already explain many biological phenomena. However, the exploration of such networks is not a trivial task, and the set of problems can be extended even further by adding new omics layers. Several solutions to this problem may exist, e.g. the construction of correlation networks, which is an important cornerstone of our future research.

5. Discussion

5.1. Volatile organic compound (VOC) analysis in the vapor space of horseradish-associated fungi

5.1.1. Identified fungi

The endophytic fungi we identified are also commonly found in the endospheres of plants of Brassicaceae, and also some less common strains, such as *Volutella* sp. We have found several genera that were described in *Arabidopsis thaliana*, also a member of Brassicaceae, e.g. *Phoma*, *Phomopsis* and *Plectosphaerella* (Junker et al., 2012).

Although we were able to isolate many more soil fungi compared to the endophytes, it appeared to be a more homogeneous group in terms of species diversity, dominated mainly by *Fusarium* and *Penicillium* species. This is not surprising, as only a few Ascomycota taxa are dominant in soil communities worldwide, including the genera mentioned (Egidi et al., 2019).

5.1.2. Performance of the proposed headspace GC-MS inoculation method

With our method, the sealed headspace glass is assumed to provide a microaerophilic environment for the fungi, which stimulates the production of VOCs. This phenomenon has been observed previously in endophytic fungi (Schoen et al., 2017). Using our method, we

have described, for the first time to our knowledge, the VOC patterns of fungi belonging to the genera *Cadophora*, *Curvularia*, *Notophoma*, *Paraphoma*, *Plectosphaerella*, *Pyrenochaeta*, *Setophoma* and *Volutella*. The VOC profiles of representatives of some other genera we have studied have been described previously, e.g. *Colletotrichum* (Rojas-Flores et al., 2019), *Penicillium* (Ndagijimana et al., 2008), *Phomopsis* (Singh et al., 2011), in which studies several compounds were found, similar to those we have identified, such as methyl-1-butanol, acetone (Singh et al., 2011), methyl-1-butanol acetate, benzaldehyde (Rojas-Flores et al., 2019) and styrene (Ndagijimana et al., 2008). The VOCs produced by fungi can provide them with various competitive advantages, especially if nutrients produced by the host plants are also available to them. This phenomenon can also be exploited by plants for their own benefit, for example by releasing various exudates into their environment (Zeng et al., 2003), thus attracting and feeding fungi that can help their development. By altering the composition of exudates, e.g. the proportion of glucosinolates, the plant can change the composition of the rhizosphere community to its own benefit (Bressan et al., 2009; DeWolf et al., 2023).

5.1.3. Glucosinolate decomposition by fungi

It has long been known that fungi are also capable of degrading glucosinolates, with thioglucosidase activity first described from *Aspergillus sydowi* nearly 80 years ago (Reese et al., 1958). Since then, the list of glucosinolate-degrading fungi has expanded considerably, as reviewed by us (Plaszkó et al., 2022), e.g. *Macrophomina phaseolina*, *Phoma radicina*, *Setophoma terrestris* (Szűcs et al., 2018), *Aspergillus clavatus*, *Fusarium oxysporum* (Smits et al., 1993).

Since no sulfur-containing compounds were detected in the vapor space of fungi growing on malt extract medium, it is likely that sulfur-containing compounds produced by fungi incubated on horseradish extract could be glucosinolate degradation products. Based on our preliminary calculations, glucosinolates in horseradish account for about 30-35% of the total sulfur content of the plant. Calibration samples containing a mixture of different standards showed good relative standard deviations (RSD) for allyl cyanide, allyl isothiocyanate, dimethyl sulfide and carbon disulphide. The instrumental analysis of 2-phenylethyl isothiocyanate and phenyl propionitrile could not be reliably reproduced due to their low volatility and are not discussed further. Allyl cyanide from sinigrin was very well detectable by our method, being detectable in the vapor space of virtually all fungal cultures, and 4 strains produced significantly high amounts compared to controls. However, it is

important to note that the formation of allyl cyanide may not only be the product of enzymatic reactions, pH or ionic changes may contribute to its formation by spontaneous decomposition. Allyl isothiocyanate is also undeniably a compound of sinigrin origin, and was produced in significant quantities by 3 strains of fungi. Isothiocyanates can form conjugates with glutathione or thiol side chains of proteins, which are virtually impossible to detect by GC-MS (Hanschen et al., 2014; Plaszkó et al., 2021). This may also contribute to the low number of fungal strains detected, as it is assumed that fungi release glutathione into the environment as a kind of chemical adaptation mechanism to eliminate isothiocyanates that are toxic to them. A similar conclusion was reached in our previous studies (Szűcs et al., 2018). We also detected carbon disulfide in several fungal strains, in 4 cases in significant amounts compared to controls, which could be derived from non-enzymatic degradation of allyl isothiocyanate (Pecháček et al., 1997) or even from detoxification pathways by fungi. This may be supported by the relatively high correlation between the levels of allyl isothiocyanate and carbon disulfide in the headspace vapor spaces ($R^2 = 0.608$).

5.1.4. Separation of endophytes and soil fungi by VOC patterns

The vast majority of the fungal isolates could be uniquely characterized by their VOC patterns, as the amount of several compounds differed very significantly between isolates, such as allyl cyanide ($p = 5.55 \times 10^{-48}$), allyl isothiocyanate ($p = 3.48 \times 10^{-4}$) and carbon disulfide ($p = 5.40 \times 10^{-13}$). Nevertheless, the endophytic and soil fungal sets we investigated were not significantly distinguishable, either in terms of VOC patterns or in terms of the amount of individual compounds. This observation may also be due to the fact that while the endophytic group showed a high taxonomic diversity, this was not the case for the soil fungi group. Although it is important to note that there was a very significant difference between *Fusarium* and other genera of fungi in terms of the compounds detected in their vapor space, no significant difference was found between the two groups studied.

5.2. Investigation of metabolome - microbiome correlations in horseradish plants

5.2.1. Compounds identified by untargeted metabolomics

In addition to glucosinolates, kaempferol glycosides and phospholipids have also been detected from horseradish roots before (Herz et al., 2017), supporting our observations. The isomers of the putatively identified glucosinolates have been described from horseradish as isothiocyanate degradation products (Blažević et al., 2020), although pentyl glucosinolate has

not been previously found in horseradish samples (Agneta et al., 2012). The presence of the m/z 259,0126 fragment (Rochfort et al., 2008), which is characteristic of glucosinolates, and the Canopus hits suggest that the candidate compound is most likely a glucosinolate. The other compounds include a coumarin glycoside, which to our present knowledge has not been previously described from horseradish roots, but is known to be an important constituent of exudates from some members of the Brassicaceae plant family (Sarashgi et al., 2021), so its presence is not surprising. Indole-3-carbaldehyde and other tryptophan-derived compounds are also frequently found in the Brassicaceae literature and have been detected in various species of the plant family (Bednarek et al., 2011; Li et al., 2023). Overall, the suggestions generated by the CSI:FingerID and Canopus algorithms have significantly shortened the time required to identify candidate compounds, although not all suggestions have proven to be completely accurate when examined in detail.

5.2.2. Variability of the glucosinolate content and metabolome

Glucosinolates are essentially precursors of antifungal compounds (Plaszkó et al., 2022, 2021) and are of particular importance in the prevention of plant pathogenesis (Agee et al., 2010; Frerigmann et al., 2016; Kuhn et al., 2017). In a comparative study on six Italian horseradish cultivars, similar sinigrin, gluconasturtiin, glucobrassicin and glucoiberin glucosinolate ratios were reported as those we found in our dataset (Agneta et al., 2014). The variability between sampling years is presumably due to differences in the amount of rain that falls during the growing seasons, as temperature and water play an important role in plant production (Nguyen et al., 2013). It is also important to note that the 2-20-fold differences in the concentrations of the main glucosinolates between cultivars were found to be sufficient to test the concentration data using Spearman correlation tests after autoscaling.

5.2.3. Endophytic fungal community composition of horseradish varieties

In the plant samples, we found several genera that have been previously described as endophytes, including *Paraphoma* (Kang et al., 2021), *Plectosphaerella* (Feng et al., 2021; Wei et al., 2021), *Podospora* (Penner and Sapir, 2021). Some of these have been identified in Brassicaceae plants, e.g. *Exophiala* (Maciá-Vicente et al., 2016), *Plectosphaerella* (Ważny et al., 2021), *Setophoma* (Poveda et al., 2020; Szűcs et al., 2018).

All alpha and beta diversity metrics showed clear differences between different sample types. As expected, soil samples showed the highest richness and diversity compared to other samples. This is not surprising, as the plant microbiome is generally less diverse (Sasse et al.,

2018), as a very strong filter may be the inability of many soil fungi to colonize plant tissues. It has also been previously described that root endophytic communities of phylogenetically distant plant families show surprisingly high similarity (Toju et al., 2019).

In contrast, dominance and equality indices showed that horseradish samples have higher variability within their own category than soil samples. A similar phenomenon has been described previously (Seabloom et al., 2019), where the authors concluded that fungal endophytic communities vary within a field but are not consistently affected by host plant nutrient supply. Nevertheless, as we will see later, if a sufficiently large number of replicates of a single variety are available, it is possible to successfully build models that can explore correlations between plant chemistry and fungal colonization. In most of the horseradish samples, a few fungal strains had relatively high dominance, which has been observed previously in several plant families (Toju et al., 2019). However, the degree of dominance is still striking, which may suggest that there are pioneer or fast-growing opportunistic species that may occupy a significant part of the plant niche. This also raises interesting questions about how many replicates would need to be examined in detail in such studies.

5.2.4. Correlations between fungal and chemical features

Our results suggest that complex chemical methods, such as metabolomics, can be highly advantageous in the study of plant-microbe interactions. Although this method cannot replace studies using knock-out mutant plants to demonstrate the effect of a compound on the assembly of the plant microbiome, metabolomics has shown that about one third of the plant fungal community correlates with changes in the plant metabolome.

Although the designed experimental setup allowed us to search for correlations, our results should not be interpreted directly as causal relationships, as there are many other explanations that may underlie the phenomena. For example, a positive correlation between candidate compounds and fungal abundance values could be the result of a positive feedback loop between "recruitment" and elicitation. In such cases, the fungus follows the chemical signals emitted by the plant and then attempts to colonize it, which the plant tries to control by a response reaction (Sasse et al., 2018). The biosynthesis of the compound is stabilized at a level that the fungus can tolerate, but no longer results in a further plant response, creating a balanced antagonistic relationship between the plant and the fungus (Schulz et al., 2015). If this biosynthetic rate is higher than the baseline, then presumably a positive correlation between fungal abundance values and metabolite concentrations can be seen. Since metabolomics is a "snapshot" of dynamic processes (Shen et al., 2023), it is also possible that

an increase in the concentration of a candidate compound is the result of a fungal invasion, an event that may cause accumulation of, for example, amino acids, short peptides or cell wall-forming monomers. It is worth mentioning that fungi may use specific cell wall degrading enzymes when colonizing living tissues (Sun et al., 2023; Zuccaro et al., 2011). Since the invasion strategy of different fungal taxa may differ, the plant defends itself with a combination of different biosynthetic compounds (Narayani and Srivastava, 2017; You et al., 2021).

Presumably, negative correlations between chemical and fungal data are seen in cases where the fungus attempts to invade the host plant but biosynthesis of a compound that successfully reduces fungal colonization is triggered. Plants that are not capable of biosynthesis of such compounds are much more likely to be colonized by such a fungal strain, although this phenomenon can only be narrowed down to compounds with proven antifungal activity (Bednarek et al, 2009; Lipka et al., 2010). In other cases, the negative correlation may also indicate that during cell and apoplast colonization, fungi take up certain compounds, so that plant tissues may be partially depleted of them.

Of the classes of compounds identified, flavonoid glycosides were shown to be one of the most influential groups in the fungal abundance correlations. Many flavonoids, including kempferol derivatives, have been shown to be antifungal compounds under in vitro conditions (Al Aboody and Mickymaray, 2020; Sá et al., 2023), and there is also in vivo evidence that these compounds may play a role in fungal defense. There are publications in the literature describing that in fungal infection, biosynthesis of flavonoids or their downstream products may be increased (Förster et al., 2022) or accumulate in higher amounts, e.g. in *Cucumis sativus* plants (McNally et al., 2003). Since flavonoids have shown positive correlations with most fungi, it is conceivable that they may play an important role in the colonization-elicitation feedback loop mentioned above or in attracting fungi capable of forming arbuscular mycorrhizal connections (Wu et al., 2023).

A considerable literature on the antifungal aspect of glucosinolate degradation products is already available, but since glucosinolates per se have no antifungal activity, it is likely that degradation products have a real role in the regulation of fungal colonization, as summarized previously by us (Plaszkó et al., 2022). The effect of native glucosinolates on fungi investigated in the present research does not seem clear even in the present data set, as the abundance of Saccharomycetales species is negatively correlated with the approximate glucosinolates identified, but members of the Xylariales and Morosphaeriaceae are positively correlated. The major glucosinolates do not seem to show correlations, but this may be due to

the fact that horseradish has a particularly high concentration of glucosinolates compared to other Brassicaceae vegetables, which may imply that only the formation of downstream products is a prerequisite for stopping fungal colonization and that it is not necessary (or even possible) to enhance glucosinolate biosynthesis.

Not surprisingly, phytoalexin-like compounds showed negative correlations with fungal abundance values, several Ascomycota ASVs that could not be identified in more detail were negatively correlated with an indole-3-carboxylic acid derivative, and an indole-3-methyl amino acid derivative and indole-3-methyl cysteine were also found to be active. Close relatives of these compounds are frequently reported in publications on *Arabidopsis*-fungal interactions, where a significant role is attributed to these compounds in inhibiting fungal colonization (Bednarek et al., 2011; Fukunaga et al., 2017; Gamir et al., 2014; Kuhn et al., 2017; Sanchez-Vallet et al., 2010), hence we believe that high levels of these compounds contribute to the inhibition of fungal colonization.

Peptides in general do not have significant antifungal activity, but specific peptides with antifungal activity have been described (Fan et al., 2023), although they are much larger than some amino acids. Nevertheless, it is interesting to note that peptides 423,1387@12.58 and 471,1058@11.06 showed a very strong negative correlation, with higher concentrations of these peptides leading to a reduction in the amount of fungi belonging to the Xylariales, Capnodiales, Saccharomycetales, Sordariales, Agaricomycetes groups and the genera *Setophoma*, *Monosporascus*, *Melanoleuca*. More remarkably, one of these was identified as a glutathione isothiocyanate adduct, which is formed during the detoxification of isothiocyanates in fungi (Szűcs et al., 2018). In a targeted search of the data (using mzMine software), we also found the features 407,1071@11.86 and 471,1388@13.35, which are most likely glutathione adducts with isothiocyanates of sinigrin and gluconasturtiin, and a trace of a glucobrassicin origin was also found. It is worth mentioning that while the amount of glucoiberin-derived isothiocyanate adduct showed a correlation of 0.709 with the actual amount of glucosinolate of glucoiberin, this value was below 0.35 for sinigrin and gluconasturtiin. This is encouraging from the point of view that we are probably not seeing sample preparation artifacts, i.e. downstream products from the myrosinase reaction due to insufficient cryogenic homogenization. However, the presence of these adducts raises a number of additional questions that could form the basis for further research. As isothiocyanate derivatives or adducts are part of the glucosinolate defense system (Hiruma et al., 2013; Piślewska-Bednarek et al., 2018), it is hypothesized that glucosinolate degradation can occur without visible concentration changes. As shown above, this may also be a result of

the particularly high concentration of glucosinolates in horseradish compared to other cruciferous plants (Gonda et al., 2016). Instead of directly investigating glucosinolates, it may be more relevant to study downstream degradation products when studying fungus-plant interactions, which may shed light on additional, hitherto unknown phenomena. This approach has already been applied to *Arabidopsis thaliana* plants, where a wide range of glucosinolate downstream degradation products have been investigated (Bednarek et al., 2011). Glutathione also plays an important role in the generation of glucosinolate-derived isothiocyanate downstream products, presumably also in the protection against autotoxicity (Hématy et al., 2020; Hiruma et al., 2013; Piślewska-Bednarek et al., 2018), which is supported by our results. The ability to biosynthesize various glucosinolate downstream products, such as indole compounds, is an important prerequisite for the regulation of colonization by fungi (Bednarek et al., 2009). However, it is important to note that the list of key compounds involved in these processes is not yet complete (Frerigmann et al., 2016; Kuhn et al., 2017). Based on our results, we hypothesize that glutathione-conjugated isothiocyanate degradation products of glucosinolate origin may be a good starting point for future plant-fungal interaction studies.

The positive correlations shown by the amino acid derivatives, however, primarily support the "fungal recruitment" hypothesis, since the primary metabolites, not surprisingly, may also serve as nutrients for fungi. It is also possible that the increased fungal diversity could lead to an increase in the abundance of protein-degrading enzymes in the apoplast, which could result in more amino acids and amino acid derivatives that could be taken up by fungi. Other primary metabolites (phospholipids and lipid-like molecules, nucleotide derivatives) showed mixed correlations with individual fungal taxa. Although specific phospholipids with antifungal activity have been described (Cho et al., 1999), they do not have direct antifungal activity and it can be assumed that their reduced abundance is mainly due to their use by the fungi as a source of carbon and phosphorus.

6. Summary

The research underlying the present thesis focused on the fungal endophytes of the horseradish microbiome. We used GC-MS and LC-MS/MS instrumental analysis and next-generation sequencing methods to investigate the chemical interactions between endophytes and the host plant.

In the first phase of our study, we investigated the volatile organic compounds (VOCs) produced during the incubation of endophytic/soil fungi isolated from/near horseradish roots on a medium prepared from horseradish roots. Our proposed incubation method proved to be suitable for efficient analysis of the VOCs produced by the fungi. Among the volatile compounds detected in the headspace of the growing fungi, we found, for example, alcohols, esters and ketones, as well as sulfur-containing compounds that are most likely degradation products of glucosinolates in horseradish. Not only allyl cyanide and allyl isothiocyanate of sinigrin origin were detected, but also carbon disulphide and dimethyl sulfide. The presence of these compounds suggests the existence of alternative degradation or detoxification pathways. Furthermore, we firstly described the VOC patterns of representatives of several fungal genera (e.g. *Curvularia*, *Notophoma*, *Paraphoma*, *Setophoma*). Although the VOC patterns of the individual fungi were significantly different, the endophytic and soil fungi groups could not be significantly distinguished on the basis of these patterns.

In the second phase of our research, we investigated the *in vivo* chemical interactions between the metabolome of different horseradish varieties and their endophytic fungi, integrating untargeted metabolomic and metagenomic data. Changes in the metabolome of horseradish roots showed several correlations with the relative abundance of different endophytic fungi. At least one third of the identified fungal taxa were significantly correlated with one or more chemical features. Using untargeted metabolomics, we were also able to putatively annotate several features, including amino acid derivatives, flavonoid glycosides, glucosinolates, indole derivatives, lipids, aromatic/polyphenolic compounds, peptides, etc. Our results suggest that untargeted metabolomics can be a very useful complementary technique to other omics (metagenomics, proteomics, transcriptomics) approaches. Our metabolomics - metagenomics combination has also revealed a number of well-understood phenomena in the complex world of plant - fungus interactions. We hope that our research can provide a good starting point for future multi-omics studies.

7. List of publications related to the dissertation

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/351/2023.PL
Subject: PhD Publication List

Candidate: Tamás Plaszkó
Doctoral School: Doctoral School of Pharmacy
MTMT ID: 10076234

List of publications related to the dissertation

1. **Plaszkó, T.**, Szűcs, Z., Cziáky, Z., Ács-Szabó, L., Csoma, H., Géczi, L., Vasas, G., Gonda, S.: Correlations between the metabolome and the endophytic fungal metagenome suggests importance of various metabolite classes in community assembly in horseradish (Armoracia rusticana L., Brassicaceae) roots.
   Front. Plant Sci. 13, 1-34, 2022.
   DOI: https://doi.org/10.3389/fpls.2022.921008
   IF: 5.6
2. **Plaszkó, T.**, Szűcs, Z., Kállai, Z., Csoma, H., Vasas, G., Gonda, S.: Volatile Organic Compounds (VOCs) of Endophytic Fungi Growing on Extracts of the Host, Horseradish (Armoracia rusticana).
   Metabolites. 10 (11), 1-15, 2020.
   DOI: http://dx.doi.org/10.3390/metabo10110451
   IF: 4.932

List of other publications

3. Gonda, S., Szűcs, Z., **Plaszkó, T.**, Cziáky, Z., Kiss-Szikszai, A., Sinka, D. Z., Bácskay, I., Vasas, G.: Quality-controlled LC-ESI-MS food metabolomics of fenugreek (Trigonella foenum-graecum) sprouts: Insights into changes in primary and specialized metabolites.
   Food Res. Int. 164, 112347, 2023.
   DOI: http://dx.doi.org/10.1016/j.foodres.2022.112347
   IF: 8.1 (2022)
4. **Plaszkó, T.**, Szűcs, Z., Vasas, G., Gonda, S.: Interactions of fungi with non-isothiocyanate products of the plant glucosinolate pathway: A review on product formation, antifungal activity, mode of action and biotransformation.
   Phytochemistry. 200, 1-33, 2022.
   DOI: http://dx.doi.org/10.1016/j.phytochem.2022.113245
   IF: 3.8

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. **Plaszkó, T.**, Szűcs, Z., Vasas, G., Gonda, S.: Effects of Glucosinolate-Derived Isothiocyanates on Fungi: A Comprehensive Review on Direct Effects, Mechanisms, Structure-Activity Relationship Data and Possible Agricultural Applications.
   J. Fungi. 7 (7), 1-38, 2021.
   DOI: http://dx.doi.org/10.3390/jof7070539
   IF: 5.724

6. Gonda, S., Szűcs, Z., **Plaszkó, T.**, Cziáky, Z., Kiss-Szikszai, A., Vasas, G., Mikóné Hamvas, M.: A Simple Method for On-Gel Detection of Myrosinase Activity.
   Molecules. 23 (9), 1-11, 2018.
   DOI: http://dx.doi.org/10.3390/molecules23092204
   IF: 3.06

7. Szűcs, Z., **Plaszkó, T.**, Cziáky, Z., Kiss-Szikszai, A., Emri, T., Bertóti, R., Sinka, L. T., Vasas, G., Gonda, S.: Endophytic fungi from the roots of horseradish (Armoracia rusticana) and their interactions with the defensive metabolites of the glucosinolate - myrosinase - isothiocyanate system.
   BMC Plant Biol. 18 (1), 1-15, 2018.
   DOI: http://dx.doi.org/10.1186/s12870-018-1295-4
   IF: 3.67

8. Papp, N., Gonda, S., Kiss-Szikszai, A., **Plaszkó, T.**, Lőrincz, P., Vasas, G.: Ethnobotanical and ethnopharmacological data of Armoracia rusticana P. Gaertner, B. Meyer et Scherb. in Hungary and Romania: a case study.
   Genet. Resour. Crop Evol. 65 (7), 1893-1905, 2018.
   DOI: http://dx.doi.org/10.1007/s10722-018-0663-0
   IF: 1.296

Total IF of journals (all publications): 36,182
Total IF of journals (publications related to the dissertation): 10,532

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

19 July, 2023

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár

8. Acknowledgements

I would like to thank my supervisor Dr. Sándor Gonda for his selfless help in coordinating my research and writing this thesis, for facilitating my professional development, and for our 8 years of professional collaboration.

I am also grateful to my co-authors, Dr. Gábor Vasas, Dr. Zsolt Szűcs, Dr. Zoltán Cziáky, Dr. Lajos Ács-Szabó, Dr. Hajnalka Csoma, Zoltán Kállai and Dr. László Géczi for their help in preparing the publications on which my thesis is based.

I would like to thank Dr. Szilárd Póliska and the staff of the Genomic Medicine and Bioinformatics Core Facility for sequencing.

I would like to thank all the staff of the Department of Botany and the current and former members of our research group for their help, work and professional comments in the preparation of my thesis.

I would like to thank Dr. Eszter Csoma and Dr. Árpád Tósaki for their tireless work in the administration of our doctoral school.

Special thanks go to my family for all their selfless help during my doctoral research.

Finally, I would like to thank my girlfriend for everything, but most of all for her mental health support during my doctoral research.