

Metabolite Profiling and Chemometric Analysis of *Hypericum*

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät II
Chemie, Physik und Mathematik

der Martin-Luther-Universität
Halle-Wittenberg

vorgelegt von

Frau Dipl. Leb. Chem. Pauline Stark
geb. am 16. Januar 1992 in Erfurt

This dissertation has been written under the supervision of Prof. Dr. Ludger A. Wessjohann and Dr. Katrin Franke at the Department of Bioorganic Chemistry of the Leibniz Institute of Plant Biochemistry (IPB) in cooperation with the Martin Luther University of Halle-Wittenberg.

1st Reviewer: Prof. Dr. Ludger A. Wessjohann

2nd Reviewer: Prof. Dr. Corinna Dawid

Date of public defence: 18.10.2021



Lieben heißt, unser Glück in das Glück eines anderen zu legen.

- Gottfried Wilhelm Leibniz -

Metabolite Profiling and Chemometric Analysis of *Hypericum*

Table of Contents

Table of Contents	I
Acknowledgements	III
List of Abbreviations.....	V
Summary	1
Zusammenfassung.....	2
1 Research Objectives.....	3
2 Characterization of pharmaceutical extracts from 20 <i>H. perforatum</i> genotypes	13
3 New insights into the hypericin biosynthesis by analysis of <i>H. perforatum</i> pistil tissue	31
4 Intraspecific variance of <i>H. perforatum</i> genotypes – a metabolomics approach.....	45
5 Chemical comparison of two leaf phenotypes	83
6 Phytochemical analysis of different <i>Hypericum</i> species.....	91
7 PSYCHE – a valuable experiment in plant NMR-metabolomics.....	119
8 General discussion and conclusion	131
References	137
Appendix	148
Declaration on the author contributions	VI
Curriculum vitae.....	VIII
Publications	IX
Eidesstattliche Erklärung.....	XI

Acknowledgements

An erster Stelle danke ich Prof. Wessjohann für die Möglichkeit an diesem vielseitigen Thema am Leibniz Institut für Pflanzenbiochemie zu arbeiten. Er förderte meine Ausbildung mit vielfältigen Teilnahmen an Konferenzen, Workshops und ermöglichte mir so den Einblick in diverse spannende Forschungsthemen.

Ein besonderer Dank geht an meine Mentorin Dr. Katrin Franke. Ich habe mich zu jeder Zeit exzellent betreut gefühlt und konnte immer auf ihren Rat und die Bereitschaft für Diskussionen zählen. Sie hat mich bei allen meinen Vorhaben und Ideen unterstützt, förderte mein Selbstvertrauen und hat mich nach Verlust des roten Fadens immer wieder zurückgeführt. Nicht zuletzt ein großes Dankeschön für die Begleitung meines Schreibprozesses trotz der schwierigen Umstände. Vielen Dank für Alles liebe Katrin.

Andrea Porzel möchte ich für ihre ansteckende Begeisterung für die NMR danken. Ich habe so viel mehr gelernt als nur Spektren zu messen und auszuwerten. Vielen Dank für das Vertrauen, die Fürsorge und die Unterstützung jeder Art.

Ich möchte allen Mitgliedern des SAW-Projektes für die gute Zusammenarbeit danken. Es war spannend in einem so interdisziplinären Team zu arbeiten. Vor allem den Mitstreiterinnen der ersten Stunde, Sarah Scharfenberg und Kira Neidhardt, möchte ich für die herzliche Aufnahme in das Projekt und auch ans IPB danken. Ich habe in vielerlei Hinsicht von dieser Kooperation und Freundschaft profitiert. Kira gilt mein Dank für die Anzucht der *Hypericum*-Pflanzen am Institut und diverse erleuchtende Stunden in der Phytokammer. Sarah habe ich alles zu verdanken was mit R zu tun hat, trotz des hohen Frustrationspotenzials hat sie es geschafft mich für das Programm zu begeistern und hatte immer ein offenes Ohr für meine Statistikfragen, oder einen Code-Schnipsel zum Lösen meiner Probleme. Hinzu kommen noch ihre Dienste als Erntehelfer, Schreib- und Motivationscoach, Organisationstalent und Lieblingskaffeebegleitung – ohne sie hätte ich diese Arbeit nicht verfassen können.

Des weiteren möchte ich Paride Rizzo vom IPK Gatersleben danken. Durch das Zusammentragen und den Anbau der Vielzahl an *Hypericum* Pflanzen machte er die Studien an diesen umfassenden Datensätzen erst möglich. Luisa Möhle und Markus Krohn möchte ich für die Bemühung an den *Hypericum*-Alzheimer-Assays danken. Mariam Gaid danke ich für die umfassenden und spannenden Diskussionen über *Hypericum* und die wertvollen Ratschläge.

Annegret Laub danke ich für die Hilfen mit der LC-MS, denn trotz, dass ich selbstmessen durfte wusste ich, dass ich nicht alleine bin. Sie war die gute Seele, Laborplatznachbarin, Büronachbarin, Leidensgenossin, LC-MS-Ratgeberin, Freundin und Dauerbegleitung der zweiten Hälfte meiner Doktorandenzeit. Mit keinem sonst konnte man so gut aufgeregt sein.

Einen wertvollen Beitrag zu meiner Arbeit haben Caroline Zab und mein RISE Student Jacob Dürichen-Parfitt geleistet, vielen Dank dafür.

Clarice Noleto verdanke ich die Einführung in das Feld der praktischen Aspekte von Metabolomics-Experimenten und die fruchtbare Laboratmosphäre zu Beginn meiner Doktorandenzeit.

Helle Wangensteen von der Universität Oslo danke ich für die Durchführung der Acetylcholinesterase- und DPPH-radical-scavenger Assays. Nina Pötzsch danke ich für die Entwicklung des Drüsenzähl-Skripts während ihres Praktikums.

Ein großer Dank geht an die technischen Angestellten, die mich während meiner Arbeit begleitet haben, dabei unverzichtbare Beiträge geleistet haben und auf deren Hilfe man sich immer verlassen konnte. Danke an Anke Dettmer, Gudrun Hahn, Anja Ehrlich, Martina Lerbs, Martina Brode und Luisa Kratzmann.

Ein großer Dank geht an die Mitarbeiter der ganzen Abteilung Natur und Wirkstoffchemie, welche durch die freundschaftliche und angenehme Atmosphäre es geschafft haben, dass man immer gerne auf Arbeit gekommen ist. Für die Verkörperung des „wahren“ Forschergeistes und die ansteckende Neugier an allem Ungewöhnlichen und Neuem möchte ich Norbert Arnold danken. Vielen Dank an das ganze Technikum-Team für die tolle Zusammenarbeit und diverse Diskussionen. Lea und Jonas danke ich insbesondere für den Aufwand meine Arbeit Korrektur zu lesen. Danke an die Mitglieder der Brunchgruppe für ihre Freundschaft.

Nun möchte ich noch den wichtigsten Menschen in meinem Leben danken: meiner Familie für ihre hundertprozentige Unterstützung und bedingungslose Liebe. Meinen Freunden danke ich für jede Ablenkung von der Arbeit. Und ein besonders großer Dank geht an Robert für die Geborgenheit, Zuversicht und Lebensfreude, die ich dank ihm jeden Tag erleben kann.

List of Abbreviations

ABC	ATP binding cassette	NUS	Non uniform sampling
Aβ	Amyloid- β	nL	Narrow leaf
ACorA	Activity correlation analysis	$v_{1/2}$	Half height width
AD	Alzheimer's disease	OKS	Oktaketide synthase
bL	Broad leaf	P	Peaknumber
COSY	Correlation spectroscopy	PC	Principal component
<i>d</i>	Doublet	PCA	Principal component analysis
DAD	Diode array detector	Ph. Eur.	European pharmacopoeia
<i>dd</i>	Doublet of doublet	PKS	Polyketide synthase
DEGs	Differentially expressed genes	PPAP	Polycyclic polyprenylated acyl phloroglucinol
DG	Dark gland	PS	Pure shift
DPPH	2,2-Diphenyl-1-picrylhydrazyl	PSYCHE	Pure shift yielded by chirp excitation
FTIR	Fourier-transform infrared spectroscopy	PT	Placental tissue
GT	Genotype	QC	Quality control
G+	Glanded placental tissue	R²	Coefficient of determination
G-	Glandless placental tissue	RDB	Ring double bond equivalent
HCA	Hierarchical cluster analysis	R_f	Retention factor
HMDS	Hexamethyl disiloxane	RMSEP	Root-mean-squared error of prediction
HPLC	High performance liquid chromatography	RP	Reversed-phase
HRMS	High resolution MS	RSD	Relative standard deviation
IS	Internal standard	R_t	Retention time
J	Coupling constant	s	Singlet
L.	Linné	SNR	Signal-to-noise ratio
Lam.	Lamarck	STOCSY	Statistical total correlation spectroscopy
<i>m</i>	Multiplet	<i>t</i>	Triplet
MAO	Monoamine oxidases	TIC	Total ion current
MS	Mass spectrometry	TLC	Thin layer chromatography
MVA	Multivariate data analysis	tPS	Pure Shift tau-delay
<i>m/z</i>	Mass-to-charge-ratio	UHPLC	Ultra high performance liquid chromatography
nce	normalized collision energy	UV	ultraviolet
NMR	Nuclear magnetic resonance		

Summary

Herbal drugs are often administered as extracts, i.e. as a complex mixture of secondary metabolites. St. John's wort extracts (*Hypericum perforatum*) are applied to treat mild to moderate depression and possess wound healing properties. However, the chemical composition of the extracts depends on the plant material used. It varies with the genotype and is influenced by biotic and abiotic environmental factors. *Hypericum* species show a large genetic intraspecific variance, which complicates the production of extracts with reproducible quality. In this thesis, the metabolite variations of *H. perforatum* and within the genus *Hypericum* were investigated. Combining analytical methods (TLC, 1D NMR, 2D NMR, LC-HRMS), several hundred secondary metabolites can be detected untargeted and simultaneously and are further chemometrically evaluated. With the help of comprehensive metabolomics studies of 93 genotypes from North America and Europe, the intraspecific variance of *H. perforatum* could be characterized. In particular, rutin and hyperforin, marker compounds of the European Pharmacopoeia, were significantly reduced in more than 20% of the genotypes studied. This could have a drastic impact on efficacy. Furthermore, the content of astilbin, shikimic acid, and acetylated flavonoids also varied. Genetic characteristics such as ploidy and reproduction strategy were not reflected in the metabolite variance.

The large number of samples in the study allowed extensive correlation studies to reveal new aspects of hypericin biosynthesis. The combination of metabolome and transcriptome analysis of pistil phenotypes with and without dark hypericin-containing glands allowed the identification of the metabolites and enzymes involved in the biosynthesis. The characteristic dark and translucent glands contain the valuable ingredients hypericin and hyperforin. On the example of two leaf phenotypes, it was shown that the number of glands corresponds to the content of the gland specific metabolites. Leaf characteristics, like gland count, were determined automatically based on microscopic images.

Apart from *H. perforatum*, also other species of the genus are used in folk medicine, but only about 40% were phytochemically investigated. The analysis of 21 *Hypericum* species showed that polyprenylated acylphloroglucinols are particularly suitable for chemophenetic differentiation. An activity correlation analysis indicated that this substance class is also responsible for the activity of some species against the gram-negative bacteria *Aliivibrio fischeri*. For *H. canariense* and *H. reflexum*, 2-*O*-geranyl-methylpropanoyl-phloroglucinol was identified as the active constituent. For the multispecies dataset, a NMR Pure Shift method was optimized, and the performance concerning the metabolomics workflow was evaluated.

Breeders and researchers could use the thesis results to select genotypes or species, producing the desired metabolites for the proposed bioactivity.

Zusammenfassung

Pflanzliche Arzneimittel werden häufig als Extrakte, d.h. als ein komplexes Gemisch aus Sekundärmetaboliten verabreicht. Johanniskrautextrakte (*Hypericum perforatum*) werden zur Behandlung milder bis moderater Depressionen und zur verbesserten Wundheilung eingesetzt. Die chemische Zusammensetzung der Extrakte unterliegt jedoch immensen Schwankungen, abhängig vom Genotyp und biotischen und abiotischen Umweltfaktoren. Da *Hypericum*-Spezies auch intraspezifisch eine große genetische Varianz aufweisen, ist die Produktion von Extrakten mit reproduzierbarer Qualität anspruchsvoll. In dieser Arbeit wurden die Varianzen der Metabolitenprofile von *H. perforatum* und innerhalb der Gattung *Hypericum* untersucht. Durch die Kombination analytischer Methoden (TLC, 1D-NMR, 2D-NMR, LC-HRMS) können mehrere hundert Sekundärmetaboliten ungerichtet gleichzeitig detektiert und chemometrisch ausgewertet werden. Mit Hilfe umfassender Metabolomics-Studien von 93 Genotypen aus Nordamerika und Europa konnte die intraspezifische Varianz *H. perforatum*s bestimmt werden. Insbesondere Schwankungen von Rutin und Hyperforin, Markerverbindungen der European Pharmacopoeia, waren in mehr als 20% der untersuchten Genotypen signifikant verringert, was drastische Auswirkung auf die Wirksamkeit haben könnte. Des Weiteren variierte auch der Gehalt von Astilbin, Shikimisäure und acetylierten Flavonoiden. Genetische Charakteristika, wie Ploidie und Reproduktionsart beeinflussten die Metabolitenzusammensetzung dagegen nicht. Die große Probenzahl der Studie ermöglichte umfassende Korrelationsuntersuchungen, so dass neue Aspekte der Hypericin-Biosynthese beleuchtet werden konnten.

Die Kombination von Metabolom- und Transkriptomanalyse von Pistilphenotypen mit und ohne dunkle Hypericin-haltige Drüsen ermöglichte die Identifizierung der in die Biosynthese involvierten Metaboliten und Enzyme. Die charakteristischen dunklen und hellen Drüsen sind Speicherort für die wertgebenden Inhaltsstoffe Hypericin und Hyperforin. Anhand von zwei Blattphenotypen konnte gezeigt werden, dass die Drüsenanzahl mit dem Gehalt der entsprechenden enthaltenen Metaboliten korrespondiert. Das Zählen der Drüsen im Blatt wurde basierend auf mikroskopischen Bildern automatisiert.

Abgesehen von *H. perforatum* werden auch andere Arten des Genus ethnomedizinisch genutzt, wobei bislang nur ca. 40% phytochemisch untersucht wurden. Die Untersuchung von 21 *Hypericum*-Arten ergab, dass sich zur chemophenetischen Unterscheidung besonders die polyprenylierten Phloroglucinole eignen. Eine Aktivitäts-Korrelations-Analyse konnte zeigen, dass diese Stoffklasse auch potentiell antibakteriell ist. Für *H. canariense* und *H. reflexum* wurde als aktive Verbindung 2-*O*-Geranyl-methylpropanoyl-phloroglucinol identifiziert. Für diese Studie wurde zusätzlich eine NMR Pure Shift Methode angepasst und die Performance in einem Metabolomics Experiment bewertet. Die hier präsentierten Ergebnisse, dienen als Grundlage zur Auswahl von Genotypen mit der gewünschten Metabolitenkomposition, für die Züchtung oder weiterführende Forschung.

1 Research Objectives

The popularity of pharmaceutical products and food supplements based on natural sources progressively increases.¹ Besides, plants, fungi, and microorganisms are still a promising source to discover new bioactive compounds.² Plant-derived drugs are well known, such as the anti-cancer drug taxol (*Taxus brevifolia*), the painkiller aspirin (*Salix alba*), and the anti-malarial drug artemisinin (*Artemisia annua*). Plants have been used for thousands of years in ethnomedicine against various diseases. A frequently used plant genus is St. John's Wort (*Hypericum*). Products containing *H. perforatum*, the best-investigated species, reached the 37th place of the top-selling "herbal supplement" in the USA in 2016 (total sales 2016: US\$ 6 million).¹ The continuing high level of consumer interest in this versatile plant also explains the scientific attention in *H. perforatum* and related species. Already Theophrastus Bombast von Hohenheim (Paracelsus, 1582) described effects against "phantasmata, wüirm, wunden und balsamische tugent".³ Today, *H. perforatum* is still used to treat mild to moderate depression and is valued for its wound healing properties.

Herbal medicines, such as *H. perforatum*, are often administered as extracts, i.e. as complex mixtures of secondary metabolites. The composition of the ingredients has a decisive influence on the pharmaceutical effectiveness. Although the secondary metabolite composition is determined in the genes, it is highly variable due to biotic and abiotic environmental factors. The fluctuations make it extremely difficult to reach a reproducible quality product and identify accessions for cultivation with suitable compositions. Therefore the genetic variation within the species *H. perforatum* is investigated in this thesis (Chapters 2-4) using comparative untargeted metabolite analyses. One reason for the chemical inconsistency is the different possibilities of reproduction of *Hypericum*. By analyzing two genetically related genotypes with different leaf phenotypes, the alterability of chemical composition and leaf characteristics were shown (Chapter 5). Apart from *H. perforatum*, several other species of the genus are of great interest and are used ethnomedically against various diseases. Chapter 6 addresses the interspecific variance of 21 *Hypericum* species and their antibiotic potential.

Although the active ingredients responsible for the various applicabilities are not yet clearly identified, most studies are only restricted to analyzing the main constituents. To capture a bigger picture, untargeted metabolomics was used, utilizing different analytical techniques such as nuclear magnetic resonance spectroscopy (NMR) and liquid chromatography coupled with mass spectrometry (LC-MS). Since the aim of these comparative studies is the detection of as many secondary metabolites as possible, this places great demands on the analytical methods. In the last Chapter 7, Pure Shift NMR was used to improve the resolution of the often-used ¹H-NMR experiment.

1.1 The genus *Hypericum* L.

The name *Hypericum* derives from the traditional use of the herb. The ancient Greek hang the plant above their religious images to protect them from devils and demons, which why they called it υπερεικον (“upereikon”, υπερ = above, εικων = image). They thought that the plant which frees the human body from depression - in ancient times it was believed to be caused by devils - would also protect their religious figures.⁴

The genus *Hypericum* L. (St. John’s Wort) is assigned to the family Hypericaceae. Formerly, this family was considered a subfamily (Hypericoideae) of the Clusiaceae, but nowadays it is classified as an independent family due to DNA sequence analysis. The Hypericaceae is one of the five families of the clusioid clade of the order Malpighiales.^{5,6} Within the Hypericaceae, there are three tribes Cratoxyleae, Vismieae, and Hypericeae, whereby the genus *Hypericum* is assigned to the latter.^{6,7}

The genus *Hypericum* comprises about 500 species and is the largest genus of the family. It consists of flowering herbs, shrubs, and a few trees, which are distributed worldwide. In contrast to most other clusioid clade species, *Hypericum* species are not found in warm or hot climate zones but in the lowlands of moderate to cold climates and in the mountain regions of tropical climates.⁴ The center of species diversity is located in the Mediterranean Basin.⁸

Table 1.1 Taxonomy of the genus *Hypericum* L.

Taxon	Latin description
Class	Magnoliopsida
Subclass	Rosidae
Order	Malpighiales
Family	Hypericaceae
Tribe	Hypericeae
Genus	<i>Hypericum</i> L.

Most species of the genus *Hypericum* can be identified by their yellow flowers with free petals and stamens in 3-5 bundles or fascicles and their opposite simple and exstipulate leaves. Leaves and flowers often contain glandular secretions called “dark” and “pale” or “translucent” glands.^{4,9} However, the species of the genus are highly variable and are divided by morphological and phylogenetic studies in 36 sections.

1.1.1 Medicinal uses and phytochemistry

Hypericum spp. are known for their highly abundant secondary metabolites. The best-known representative, *H. perforatum* L. (Common St. John's Wort), is categorized into the core section Hypericum. It is a widespread Eurasian perennial plant, which is invasive to all other continents,

except Antarctica. The species owes its fame to its use in folk medicine for mental disorders and mild depression.

Especially in the western world, *H. perforatum* became a top selling herbal product. It is marketed as a dietary supplement, and in Europe, it is also approved as a drug (e.g., LAIF® 900 Balance, Neuroplant® Aktiv) against mild to moderate depression. In addition to its antidepressant activity, it has also been shown to have antibacterial, anti-inflammatory, antimicrobial, anti-Alzheimer, antioxidant and cytotoxic effects.¹⁰⁻¹² As a result of the divers applications, *H. perforatum* was chosen as "medicinal plant of the year 2015" and 2019 as "pharmaceutical plant of the year" in Germany.^{13,14} This shows the high interest in this species and the genus *Hypericum*. Although only *H. perforatum* has been included in the European pharmacopoeia,¹⁵ various other species are used in ethnomedicine in their countries of origin, but are only partially commercially marketed.¹⁶

The various effects are attributed to the diverse secondary metabolites. In particular, the genus-characteristic ingredients naphthodianthrones and polyprenylated acylphloroglucinols are in the focus of interest. Besides, the plants are rich in flavonoids, xanthenes, phenolic acids, and coumarins.^{17,18} The complexity of the extracts and the chemically diverse composition also explains the varying medical applications. Although there is a growing interest in other members of the genus, *H. perforatum* has been the subject of phytochemical studies for over 100 years. Main compounds of *H. perforatum* are shown in Figure 1.1. The relevant compound classes determine also the appearance of the plants, which show dark and translucent glandular structures.

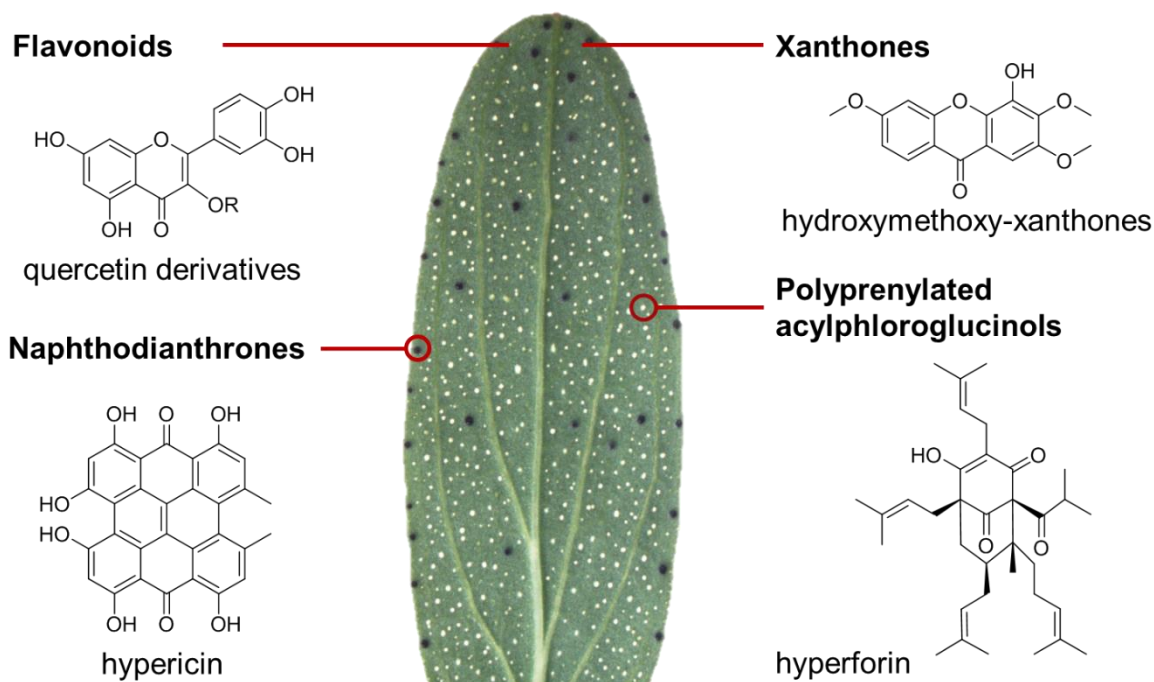


Figure 1.1 The most abundant compound classes with a typical representative and their localization marked on the image of a *H. perforatum* leaf.

1.1.2 Naphthodianthrones

Naphthodianthrones are a well-known compound class in *Hypericum* species. Their presence is visually noticeable, through the occurrence of dark glands on leaves and flower parts, like petals and stamen. These dark glands are the storage place of the naphthodianthrones hypericin and pseudohypericin, which are responsible for the dark reddish color. Around 65% of the taxonomic sections contain these dark glandular structures.^{4,9}

These compounds play an essential role in the plant defense mechanism against generalist herbivores. It was shown that the hypericin content increased about 30-100%, as a defense response to repel herbivores. However, specialist feeders, like the beetle *Chrysolina quadrigemina* are not affected by the compound and circumvent the phototoxic effect of hypericins.^{19,20} The phototoxicity is responsible for the ongoing interest in hypericin as an agent for photodynamic cancer therapy induced cell damage.²¹⁻²⁴ Although this compound is in the focus of investigations over decades, the hypericin biosynthesis is not fully understood.²⁵ Hypericin is almost exclusively found in plants but like other anthraquinones its presence was also reported from fungi.²⁶

It is believed that the hypericin core structure is formed via the polyketide pathway by type III polyketide synthase, shown in Figure 1.2. As a first step, the condensation of one unit acetyl-CoA with seven units of malonyl-CoA occurs and the formed octaketide-chain undergoes cyclization to emodin anthrone, mediated by octaketide synthase (OKS). Afterward, emodin and emodin anthrone dimerize catalyzed by the phenolic oxidative coupling like proteins (POCP) to yield emodin-dianthrone. Further oxidation steps form protohypericin and hypericin.^{25,27}

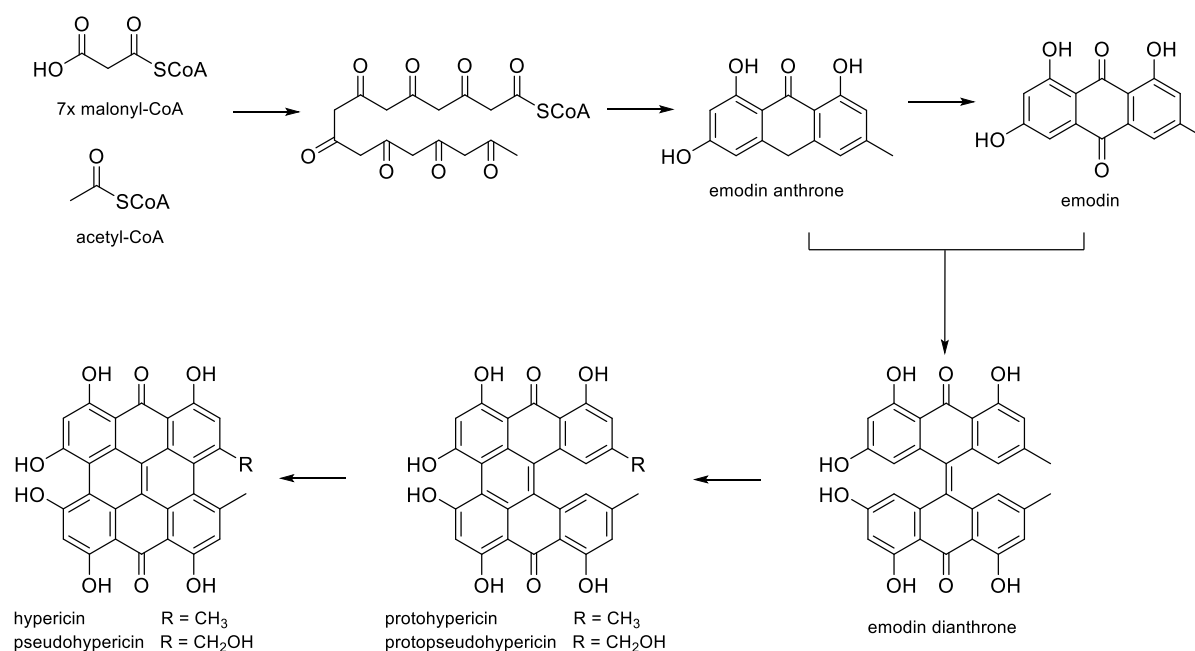


Figure 1.2 Proposed hypericin biosynthesis. Adapted from Soták et al. (2016)²⁷.

1.1.3 Acylphloroglucinols

Phloroglucinols are phenolic compounds which are derived from the phloroglucin (1,3,5-trihydroxybenzene) core. These polyketides were found in several plant families, such as *Myrtaceae*, *Cannabaceae*, *Clusiaceae* and *Hypericaceae*. They show an incredible diversity in the genus *Hypericum*. Until 2016 over 400 of these compounds were identified in 20 of the 36 taxonomic sections, extensively reviewed by Bridi et al. (2018) and Yang et al. (2018).^{28,29} The plants accumulate these compounds in secretory structures.³⁰ In *Hypericum* species, these special oil cavities appear translucent, so that the leaves look perforated.

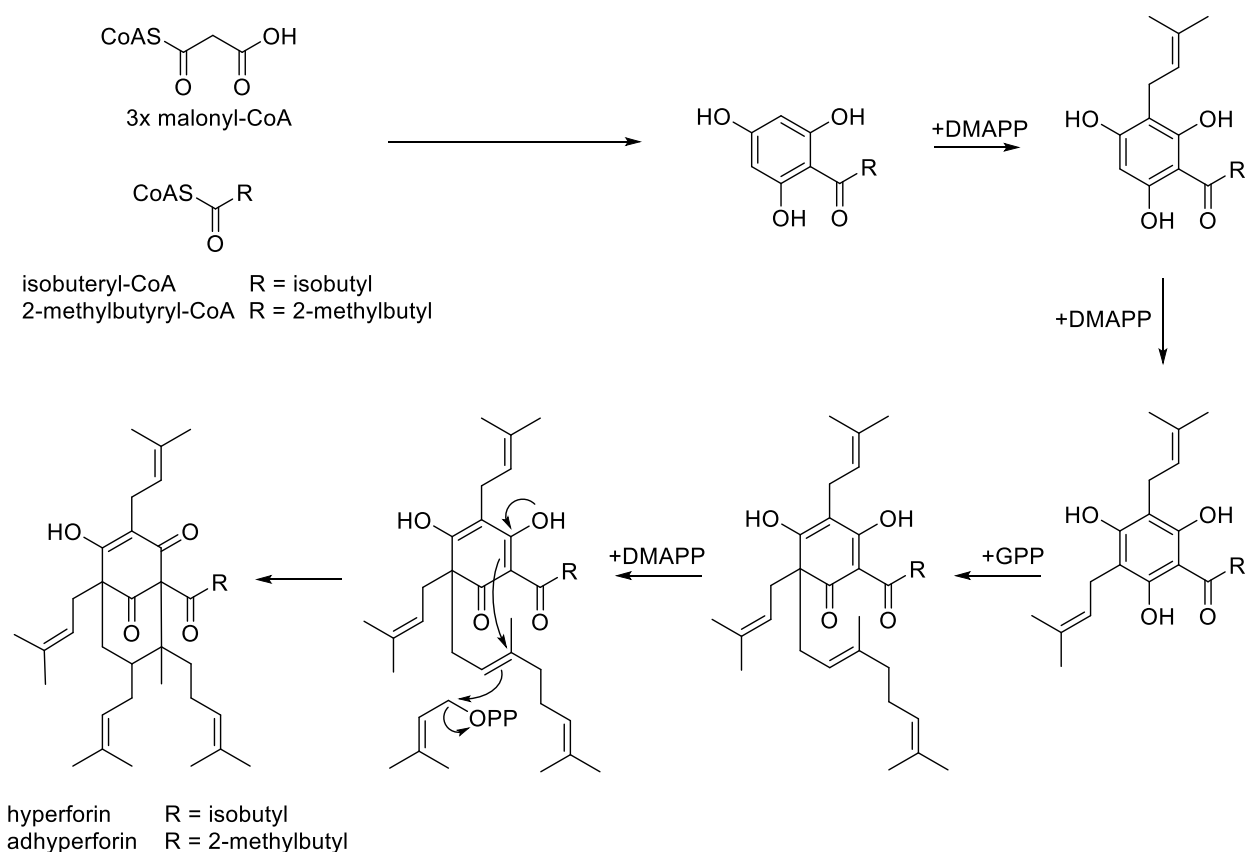


Figure 1.3 Proposed hyperforin biosynthesis adapted from Beerhues et al. (2006)³¹. (DMAPP = dimethylallyl diphosphate, GPP = geranyl diphosphate).

The biosynthesis follows the polyketide-pathway. One CoA-activated acid (isobutyryl-, 2-methylbutyryl-, or benzoyl-CoA) is condensed with three molecules of malonyl-CoA to a linear tetraketide by phlorisovalerophenone synthase.³² Further, an intramolecular Claisen-Condensation occurs, catalyzed by this polyketide synthase type III, and a cyclic acylphloroglucinol derivative is formed.³³⁻³⁶ Subsequently, prenylation reactions follow, mediated by prenyltransferases, which allows electrophilic substitutions with dimethylallyl-diphosphate (DMAPP) and/or geranyl diphosphate (GPP).^{31,37} This leads to highly substituted ring systems. A main compound of *H. perforatum* is hyperforin. The biosynthesis of this polycyclic polyprenylated acylphloroglucinol (PPAP) is shown in Figure 1.3. The geranyl sidechain undergoes an

intramolecular cyclization so that a bicyclic structure arises.³¹ Various acyl moieties, C and O prenylations, and numerous internal ring closures explain the diverse structures described.

Interest in this compound class rises because of their neuroactive potential. It is assumed that hyperforin contributes mainly to the effect of *H. perforatum* against depression.^{30,38,39} However, hyperforin limits the applicability of *Hypericum* extracts by activating cytochrome P450 enzymes in a dose dependent manner, which often affects to drug actions and interactions.⁴⁰

1.2 Plant metabolomics

The quantitative and comprehensive investigation of the metabolome, so the entire set of metabolites, in a complex biological specimen, is called metabolomics.⁴¹ The possibility to describe the chemical phenotype of a biological system is used in many different fields. Therefore the number of metabolomics experiments rises frequently since the beginning of the 20th century (2 publications in 2000 and 7607 in 2019, PubMed search “metabolomics”). Due to that, many subterms developed to distinguish between various experimental settings, but are used inconsistently. Metabolite profiling deals with a smaller number of identified metabolites, whereas the classification of samples according to their origin or biological activity is often called metabolite fingerprinting.⁴² The analysis is principally divided into two approaches: targeted, where compounds of one class or pathway are addressed, and untargeted, where all compounds are of interest. The analysis of all metabolites, however, is challenging and currently impossible. Especially plants produce a large number of metabolites with diverse structures and properties, so the combination of multiple methodologies is necessary. Among different platforms, mainly mass spectrometry and nuclear magnetic resonance spectroscopy were used.^{42,43}

In natural product research, metabolomics is applied to identify taxonomical marker compounds for authenticity analysis, to determine valuable chemotypes with a medicinal purpose, and clarify the organism's chemical adaptation to different conditions.

In order to answer a certain research question and to make a representative well-founded statement, a sufficient number of replicates must be examined. This inevitably leads to large scale data, which is processed and interpreted by multivariate data analysis (MVA).

1.2.1 Metabolomics workflow

All metabolomics experiments follow a general scheme, depicted in Figure 1.4. The first step implies the planning of the experimental design to answer a particular biological question. In plant metabolomics, this includes the decision for suitable plant material, the number of replicates, the sample amounts, the collection and storage conditions. Besides, control samples have to be considered to verify that the study of the huge sample numbers is comparable.

The sample preparation depends on the tissue, the analytical methods used, and the research question. In untargeted experiments for LC-MS and NMR analysis this comprises a grinding, extraction, and centrifugation step. An appropriate dilution has to be determined, and an internal standard should be added. In general, sample preparation should be as minimal as possible to circumvent variances during sample preparation.

Data acquisition can be performed with different analytical techniques. Most frequently used is liquid chromatography coupled to mass-spectrometry (LC-MS) due to the high sensitivity. However, the application of NMR metabolomics is increasing because it is quantitative, nondestructive, and reproducible.⁴³

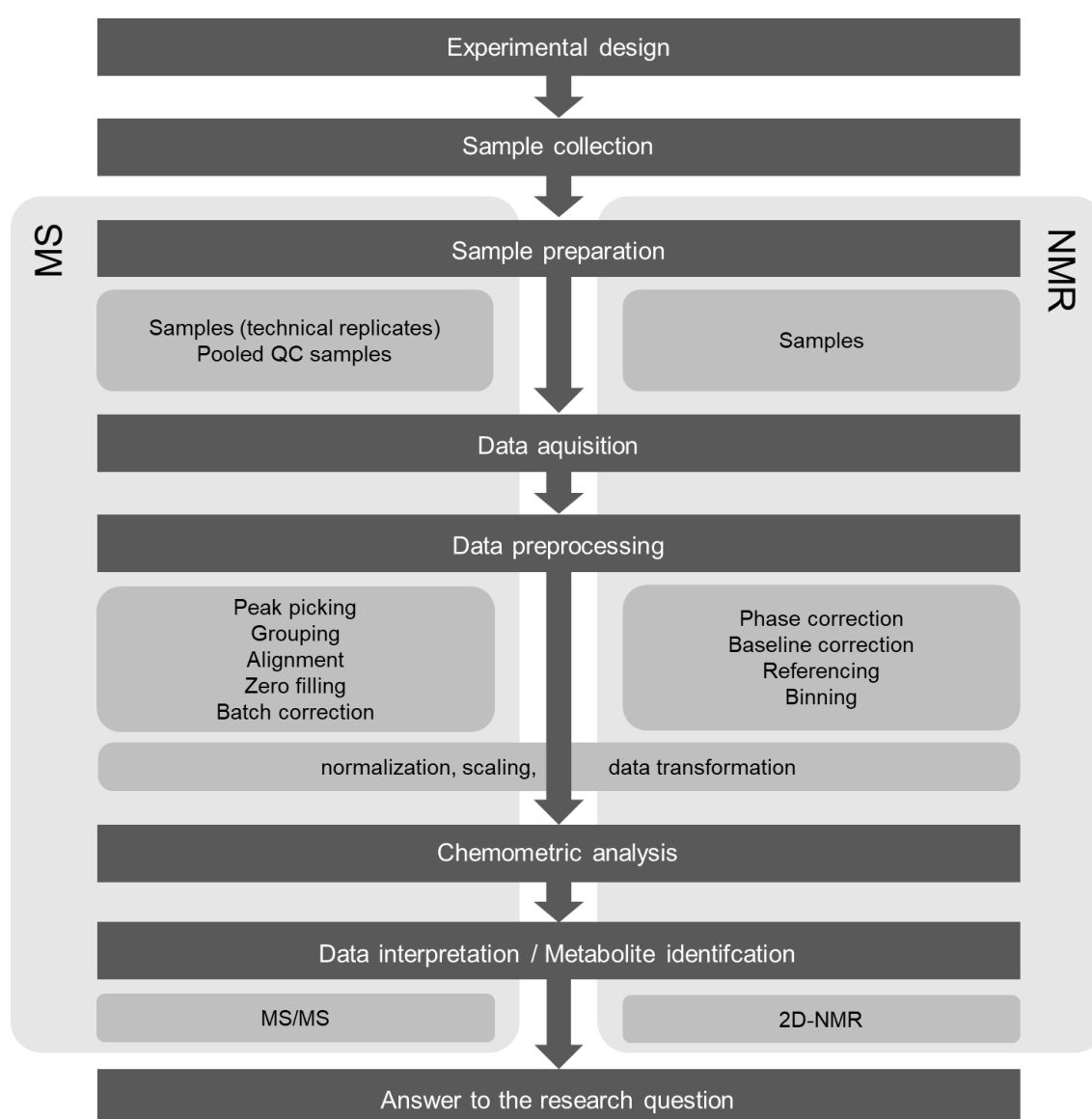


Figure 1.4 General workflow of metabolomics experiments. Special procedures for MS or NMR applications are described on the left or right side, respectively.

To turn experimental data in a data matrix evaluable with statistics, data preprocessing is a crucial step. For LC-MS, the following steps are required: peak picking (data set creation), grouping, alignment, and zero filling. Finally, the data matrix contains features defined as a combination of m/z ratio and retention time. In NMR, spectra have to be phase-corrected, baseline-corrected, and referenced. In contrast to the LC-MS, peak picking is difficult, so data sets are created by binning (dividing spectra in parts of equal size).

The received data matrices can then be statistically evaluated by chemometric methods such as univariate (e.g. t-test, ANOVA) and multivariate data analysis (see Chapter 1.2.2). The data interpretation includes mainly metabolite identification, which is still the most time-consuming part because of incomplete databases. This requires follow-up experiments, like the comparison with authentic standards, fragmentation studies with MS/MS, or additional 2D NMR experiments.

1.2.2 Chemometric methods in metabolomics

Due to the development of new technologies, more information is accessible for the analyzed samples, described by a higher number of variables. Therefore chemometric analysis is needed for the meaningful and statistical interpretation of these large-scale data sets. Chemometric analysis reduces the multidimensional data set to a lower number of dimensions, mostly based on MVA. The general workflow is graphically depicted in Figure 1.5.

The appropriate processing of the used experimental technology (e.g. MS, NMR) provides a data matrix X , which is composed of the number of samples (i) and the number of variables (k). The most widely used unsupervised method for data exploration is the Principal Component Analysis (PCA). The information of the data matrix X is reduced to uncorrelated latent variables called principal components (PC), which are linear combinations containing information from the scores matrix (T) and the loadings matrix (P). The scores plot shows the samples in the reduced space of two or three PCs and helps to identify clusters and outliers. The information about the responsible variables can be obtained from the loadings plot.

Partial Least Square (PLS) analysis is a supervised method often used in metabolomics experiments in which the distinction of two groups is acquired like the comparison between control and treated samples. The class membership or other metadata are coded in the response matrix Y . PLS determines the variables of X , which are more correlated to the response Y , by preserving maximal covariance between them.⁴⁴⁻⁴⁶ Like in the PCA, the outcome is a scores and a loadings plot.⁴⁶ A related method called Orthogonal Partial Least Square analysis (OPLS) provides the same information as PLS, but simplifies the interpretation by turning the variance correlated to the response in the first dimension.⁴⁴

Frequently used methods for similarity and dissimilarity determination are Cluster Analyses. In the most common Hierarchical Cluster Analysis (HCA) a dendrogram is built. Clusters are formed between “close” samples, which means their distance is small (such as the Euclidean distance). Although this method helps to get an overview of the sample set, the reason for clustering is not

displayed. Therefore the method is often combined with a heatmap indicating the quantities of the variables.⁴⁴

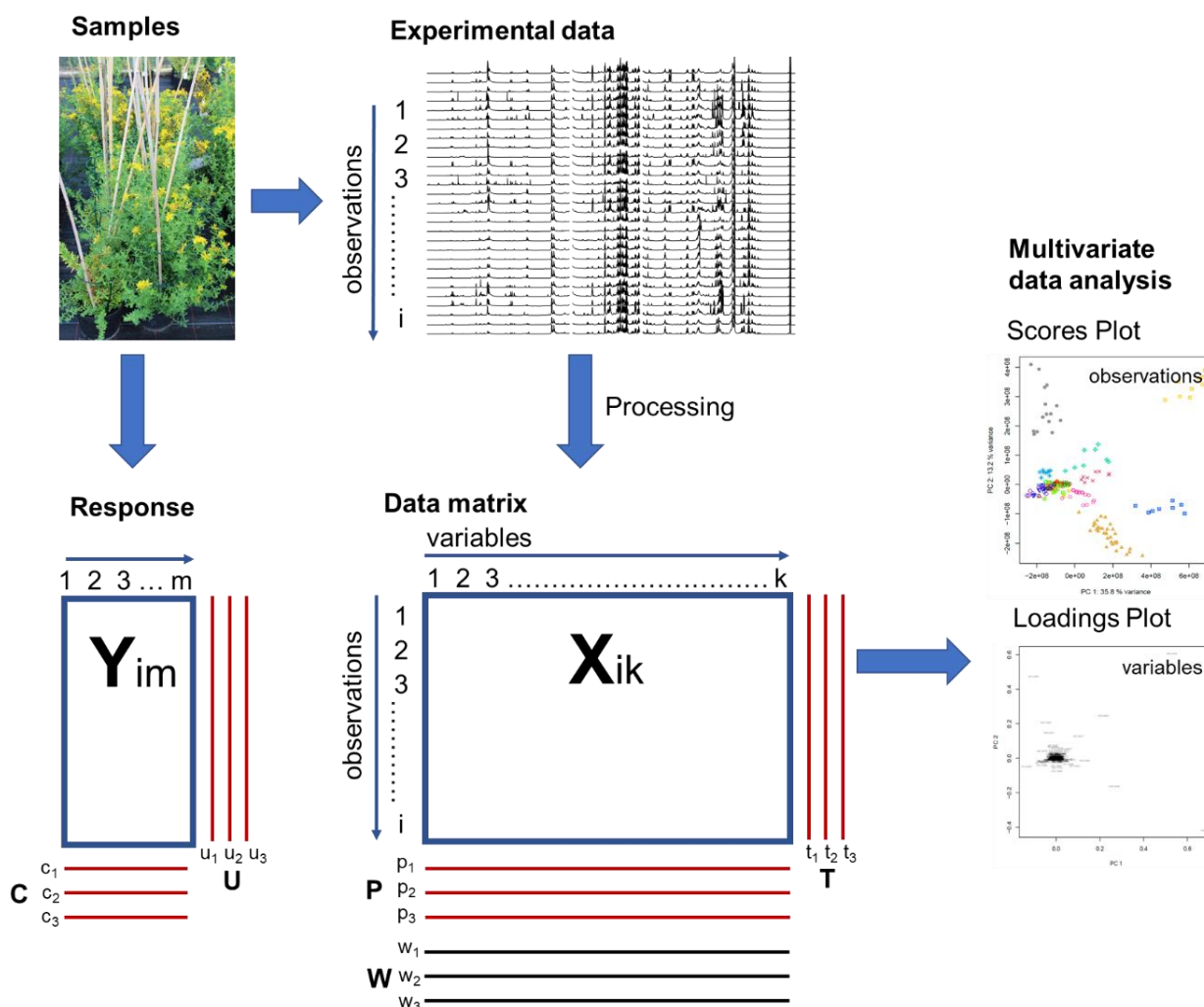


Figure 1.5 Multivariate data analysis workflow used for metabolomics experiments. The transformation of the spectral data to data matrix X consisting of k variables measured for i observations (samples). The Y matrix with m responses (e.g. metadata, activity) for i observations. Matrices were decomposed to score vectors t and loading vectors p . Calculated weight vectors w are utilized for generating scores and loadings plot of PCA. Responses are equally decomposed into scores u and loadings c . Figure adapted from Wold et al. (2001)⁴⁵ and Worley and Powers (2013)⁴⁶.

2 Characterization of pharmaceutical extracts from 20 *H. perforatum* genotypes

Abstract

Hypericum perforatum L. is an important medicinal plant in the western world. The plants are highly variable in their chemical composition, therefore the production of high-quality plant-derived medications is challenging. The metabolite composition of pharmaceuticals is both genetically determined and also highly influenced by abiotic and biotic environmental factors. The impact of the genetic background on the chemical variability of this medicinally important plant herein is explored in detail over the extractable metabolome. This study addresses the chemodiversity of *H. perforatum* L. raw material, based upon statistically-differentiated genetic backgrounds in an untargeted fashion.

To assess the intraspecific variance, aerial parts of 20 *H. perforatum* accessions, cultivated under identical growing conditions, were chemically characterized. Untargeted metabolite profiling of the genotypes was realized by a comprehensive determination of metabolite levels by TLC, UHPLC-ESI-HRMS as well as 1D-¹H-NMR spectroscopy. Principal component analysis (PCA) was used to find the variable constituents. Multivariate data analysis revealed a high variance in the phloroglucinol and flavonoid profiles. One genotype differed by lacking rutin and the phloroglucinol hyperforin, but showed an increased level of only partially prenylated precursors of hyperforin.

Untargeted metabolite profiling, based on TLC, NMR, and LC-MS, combined with multivariate data analysis, enables the determination of intraspecific chemodiversity in *H. perforatum* L. The results demonstrate variation in the composition of characteristic metabolites among different genotypes, reinforcing the necessity to select genotypes that produce the desired bioactive compounds according to the proposed pharmacological application.

2.1 Introduction

Hypericum perforatum L. (*Hypericaceae*), commonly known as St. John's Wort, is a widely distributed herbaceous plant. It has been used for centuries in traditional medicine because of its anti-inflammatory, antibacterial, and antioxidative properties.^{12,47,48} The application against mild to moderate depression led to a high degree of appreciation for the species in the Western World.⁴⁹⁻⁵¹ Recently, additional biological activities, such as anti-Alzheimer properties, have been demonstrated, whereby Hofrichter et al. (2013) showed a reduction of amyloid- β induced histopathology in APP-transgenic mice after treatment with 80% ethanolic extracts.¹¹ The complexity of secondary metabolites within *H. perforatum* extracts explains its numerous activities. The highly abundant naphthodianthrones, phloroglucinols, and flavonoids are connected to these activities (Figure 2.1).^{18,52,53} The naphthodianthrone hypericin is a photosensitizer that recently gained importance in anticancer therapy.^{22,54,55} The bicyclic prenylated acylphloroglucinol hyperforin is discussed in the context of antidepressive and neurodegenerative diseases.^{30,38,39}

Various studies have shown qualitative and quantitative variation of the bioactive compounds in commercially available St. John's Wort products.⁵⁶⁻⁵⁸ The authors concluded that the composition of the final preparations depends on the plant material and the processing steps. Scotti et al. (2019) assessed the chemical diversity of 86 globally-distributed accessions, and reported significant chemical differences in the raw material compared to the market products caused by the collection, storage, and production of the extracts.⁵⁹ As has been shown in numerous investigations, the plant ingredients are ultimately influenced by environmental and ecological effects.⁶⁰⁻⁶² A study that focused on intraspecific chemical variation, excluding environmental factors, was performed by Bagdonaite et al. (2012)⁶³ whereby 13 *H. perforatum* genotypes were cultivated and distinguished exclusively by major flavonoids measured with HPLC-DAD.

Herein, comparative metabolite profiles of 20 *H. perforatum* accessions with identical cultivation conditions were investigated with the goal of estimating intraspecific variance based solely on the genetic background of the plants. Three different techniques were used for the chemical analysis to cover a broad range of compounds, including: thin-layer chromatography (TLC) as recommended by the European Pharmacopoeia¹⁵, 1D-¹H-nuclear magnetic resonance (NMR) experiments, and ultra-high-performance liquid chromatography coupled with mass spectrometry (UHPLC-MS).

2.2 Experimental

2.2.1 Plant material

Hypericum perforatum L. seeds originating from accessions of six different countries in Europe and North America were cultivated at the IPK Gatersleben in the greenhouse at 21 °C by day hours and 18 °C during darkness in long day conditions (16 hours of light) and with light intensity of

300 $\mu\text{mol sec}^{-1} \text{ m}^{-2}$ while keeping the humidity levels between 60 and 70%. In total twenty genotypes (GTs) were used (for detailed information see Appendix 2.1), 15 of which were genetically characterized by Molins et al. (2014)⁶⁴. Furthermore, three crossed accessions and two lines with unknown genetic background were included. The aerial parts of the plants were harvested in 2013 and air dried in the dark.

2.2.2 Reference standards

Reference standards of caffeic acid, chlorogenic acid, ferulic acid, epicatechin, isoquercetin, quercetin, and sucrose were obtained from Sigma-Aldrich. Shikimic acid, rutin, and quercitrin were purchased from Roth. The authentic reference compounds hyperforin, hypericin, and hyperoside were available from the in-house compound library of the department Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry (IPB), Halle (Saale), Germany.

3-Geranyl-methylpropanoyl-phloroglucinol (**17**) was isolated from field-grown *H. punctatum* Lam. Plants (accession number: NMN69). Fresh plant material (43.6 g) was extracted by threefold maceration with 5% aqueous methanol (3.0 g extract). In total, 1 g of the crude extract was submitted to a centrifugal partition chromatography column (CPC-250, Gilson) coupled to a preparative liquid chromatography purification system (PLC 2050, Gilson) using a solvent mixture of *n*-hexane, ethyl acetate, ethanol, and water (2:6:2:5 v/v), flow 5 ml min⁻¹, and a rotor speed of 2200 rpm. Fractions containing *m/z* 331.1 were combined (50.60 mg) and further purified with HPLC-DAD (Agilent Infinity 1260/1290) water:acetonitrile gradient solvent system (each containing 0.1 % formic acid) connected to an RP-C18 column (5 μm , 150 x 10 mm, ODS-A YMC). The structure of the received white compound (3.7 mg) was elucidated using LC-MS/MS and ¹H, HSQC, HMBC, and COSY NMR experiments (400 MHz) and confirmed by comparison with Sarkisian et al. (2012).⁶⁵

2.2.3 Extraction procedure and sample preparation

The air-dried plant material was crushed to a fine powder in a macerator (La moulinette D56). The extraction procedure was adapted to the production of pharmaceutical *H. perforatum* products. Therefore, 25 g of the powdered aerial parts of plants were used for extraction. 250 ml 80% aqueous ethanol was added to each sample and stirred for 1.5 h at 45 °C. After removal of the supernatant, the extraction procedure was repeated with 250 ml solvent and stirring for 1 h. The combined supernatant was filtered, and the solvent was removed by rotary evaporation (40 °C) and lyophilization. The extracts were ground with a mortar and stored at -80 °C.

For the production of a representative quality control sample (QC), plant material from all GTs was mixed in equal parts. The mixed plant powder was extracted as described for the single GTs.

2.2.4 TLC analysis

For each GT, 4 μl extract (5 mg/ml in methanol) was applied on SilicaGel plates (60 G F254, 20 x 10 cm, Merck). The analysis was performed with a mobile phase containing: ethyl acetate,

formic acid (Roth), acetic acid (Roth), distilled water and dichlormethane (100:10:10:11:25 V/V/V/V/V).⁶⁶ After migration, the plate was air-dried and sprayed with natural product reagent (0.5 % methanolic 2-aminodiphenylborinate). The signals were visualized with UV light at 366 nm (Camac photovisualizer).

2.2.5 NMR analysis

The NMR based metabolomics approach was adapted from the investigation of Porzel et al. (2014).⁵² Samples for the NMR measurements were prepared in triplicates. Each consisted of 25 mg extract per 0.8 ml deuterated methanol (Deutero) containing 0.935 mM hexamethyl disiloxane (HMDS). The ¹H-NMR spectra were recorded with an Agilent (Varian) VNMRs 600 NMR spectrometer operating at a frequency of 599.83 MHz at +25 °C. ¹H-NMR spectra were measured with: pulse angle = 90°, relaxation delay = 21.0941 s, acquisition time = 3.9 s, number of transients = 160, zero filling 128K. The standard CHEMPACK 7.1 pulse sequences (gHMBCAD and gHSQCAD) implemented in Varian VNMRJ 4.2C spectrometer software was used to record the HMBC and HSQC spectra of the QC sample. The spectra were referenced to HMDS at 0.062 ppm for ¹H-NMR and 1.96 ppm for ¹³C NMR.

2.2.6 NMR data processing

¹H-NMR spectra were baseline corrected and reduced to bins with a spectral width of 0.02 ppm using MestreNova version 11.0. The resulting integral list was further processed with R version 3.5.1. The solvent regions at 4.7–5.0 ppm (water), 3.25–3.4 ppm (methanol), and 2.13–2.17 ppm (acetone) were removed, and the total sum normalization was applied.

2.2.7 UHPLC-MS/PDA analysis

For UHPLC-MS analysis per sample, 2 mg plant extract was dissolved in 1 ml LC-MS grade methanol (Honeywell) containing 8 µg/ml umbelliferone (HPLC grade Sigma) as an internal standard by brief mixing on a vortex mixer followed by ultrasonication for 5 min. Following centrifugation (10 min, 14.000 rpm), the supernatant was diluted to the final concentration of 0.5 mg extract per ml methanol and applied to UHPLC-MS/PDA. The preparation of the extracts was performed in triplicates per GT, and finally 2 µl were injected. The MS system was coupled to an ultra-high-performance liquid chromatography (UHPLC) system (Dionex UltiMate 3000, Thermo Scientific), fitted with a RP-C18 column (1,9 µm; 50 x 2.1 mm; Hypersil-GOLD; Thermo Scientific; column temperature: 40 °C), and a photodiode array detector (PDA, Thermo Scientific; 220–650 nm). For UHPLC separation a water:methanol gradient solvent system (each containing 0.1% formic acid) at a flow rate of 400 µl/min was applied (95:5 for 1 min, 10 min gradient to a ratio of 0:100, hold for 5 min, returning to 95:5 in 1 min, isocratic hold for 4 min).⁶⁷ Negative ion high-resolution ESI mass spectra were obtained from an Orbitrap Elite mass spectrometer (ThermoFisher Scientific, Bremen, Germany) equipped with a HESI electrospray ion source (spray voltage 4.0 kV; source heater temperature: 325 °C; capillary temperature 300 °C; FTMS

resolution 15.000). Nitrogen was used as sheath and auxiliary gas. The CID mass spectra (buffer gas: helium) were recorded in data dependent acquisition mode (dda) using normalized collision energies (NCE) of 35%. The QC sample was injected every sixth sample, as recommended by Dunn et al. (2011).⁶⁸ The data were evaluated with the Xcalibur software 2.2 (Thermo Fisher).

2.2.8 UHPLC-MS data processing

The raw spectra of samples and QCs were converted to mzML format using ProteoWizard (3.0.11110) to get centroided data. Further processing was done in R (3.5.1, <https://cran.r-project.org>) using xcms (3.2.0).⁶⁹ Peak picking based on the CentWave Algorithm was performed with `findChromPeaks(ppm=5, peakwidth=c(7,25), sntresh=10, noise=1000, prefilter=c(3,1000))`. After that, the peaks were grouped over all samples with `groupChromPeaks(minFraction=1, bw=5)`. The retention times were stable with exceptions of strong retention time shifts of hypericin peaks. Thus, no overall retention time correction was performed. Peakfilling was performed with `fillChromPeaks()` to avoid missing values that would affect later analysis. Feature intensities were logarithmized before statistical calculations. The XCMS processing resulted in 765 features characterized by retention time, m/z value, and intensity. Because of the retention time shift of hypericin, the compound appeared in two features. To overcome the wrong picking, features belonging to hypericin were condensed to feature “hyp503” and the +1 and +2 isotope to “hyp504” and “hyp505”, respectively.

2.2.9 Multivariate data analysis and data record

Feature and bin tables were analyzed with principal component analysis (PCA) performed with the R package `pcaMethods` (1.72.0)⁷⁰.

Raw data is available in MetaboLights (MTBLS969). It includes raw data of LC-MS and NMR measurements as well as the processed feature tables.

2.2.10 Biological activity assays

Acetylcholinesterase inhibition was determined by using Ellman's method according to the procedure by Passos et al. (2013)⁷¹. The samples were tested in triplicates in five concentrations (12.5–200 $\mu\text{g/ml}$), and the assay was repeated three times. As positive control tacrine (Sigma-Aldrich) ($\text{IC}_{50} = 18.2 \pm 0.3 \mu\text{g/ml} / 77.4 \pm 1.3 \text{ nM}$) was utilized.

DPPH-radical scavenger activity assay was performed as described by Wangenstein et al. (2004).⁷² The samples were investigated in triplicates, and as positive control quercetin ($\text{IC}_{50} = 2.9 \pm 0.1 \mu\text{g/ml}$) was used.

The cytotoxicity against a colon cancer (HT29) and a human prostate cancer (PC3) cell line were evaluated at two concentrations (0.05, 50 $\mu\text{g/ml}$) applying the method described by Dos Santos et al. (2019)⁷³.

2.3 Results and discussion

The aim of this study was the investigation of the intraspecific variance of *H. perforatum* metabolites. To assess the influence of the genetic background, 20 field-grown accessions, cultivated at the same time and under identical conditions, were compared based on their chemical composition. The genotypes (GTs) comprise European native and North American invasive accessions belonging to three genetic clusters which differed in ploidy, apomixis expression and reproduction strategies (Appendix 2.1).⁶⁴ Extracts, comparable to commercial preparation, were produced with 80% aqueous ethanol and the untargeted metabolite profiling of the plant extracts was performed with TLC, UHPLC-ESI-HRMS in negative ion mode, and ¹H-NMR. The obtained spectral data were analyzed by PCA to reveal differences in the metabolite levels.

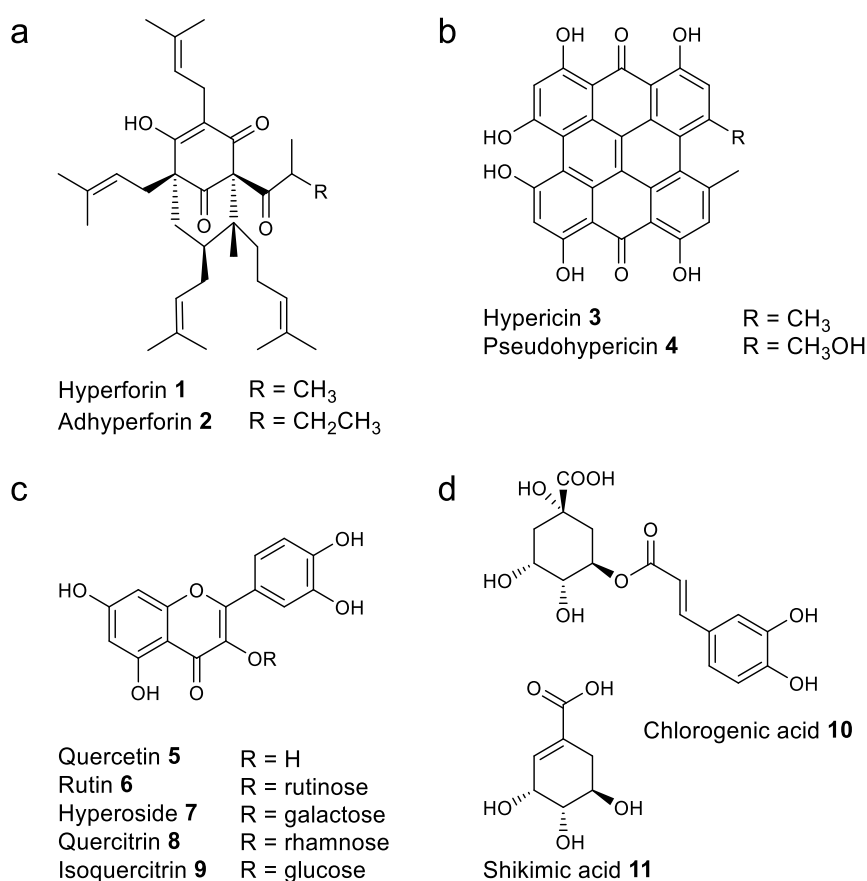


Figure 2.1 Selected major compounds detected in *H. perforatum* belonging to the compound classes: a) phloroglucinols, b) naphthodianthrones, c) flavonoids, and d) organic acids.

2.3.1 TLC analysis

TLC screening is an easy and fast tool to detect the main constituents of *H. perforatum*. It is used for the identification of plant material as *Hyperici herba*, however, the most relevant constituents are hyperforins (**1**, **2**) and are not considered by this method. In the European Pharmacopoeia (Ph. Eur. 10.0) hypericin (**3**), pseudohypericin (**4**), the quercetin derivatives rutin (**6**), and hyperoside (**7**), as well as chlorogenic acid (**10**), are defined as marker compounds for *H. perforatum* extracts.¹⁵

In Figure 2.2, the TLC fingerprints of the 20 GTs were compared and molecular differences are observable. Although the signal intensities vary, all GTs contain the red elution bands of hypericin (**3**) ($R_f = 0.85$) and pseudohypericin (**4**) ($R_f = 0.8$), as well as hyperoside (**7**) ($R_f = 0.38$). The blue band of chlorogenic acid (**10**) ($R_f = 0.36$) was also detected throughout the samples, which supports the latest decision of the Ph. Eur. 10.0 to define chlorogenic acid (**10**) as a characteristic constituent. The yellow band of rutin (**6**) ($R_f = 0.18$) is less abundant in GTs 15 and 17, and absent in GTs 3 and 20. All tested GTs fulfill the mentioned criteria of the European pharmacopeia, except GT 3 and 20 which do not contain rutin in detectable amounts. Similarly to our findings, other studies also reported lines that are lacking this compound.⁷⁴⁻⁷⁶ Based on these results, the question arises if rutin is suitable as a marker compound.

The quercetin monoglycoside at $R_f = 0.61$ was determined as quercitrin (**8**) and is only present in GTs 6, 15, 17, and 20. As a result for GTs 15, 17, and 20, notably, quercitrin (**8**) is enriched, and the rutin (**6**) content is decreased. A negative correlation between these two quercetin-derivatives in *Hypericum* was already described by previous studies.^{63,76,77} Quercetin (**5**) is glycosylated by quercetin-3-rhamnosyltransferase to yield quercitrin (**8**); whereas for the biosynthesis of rutin (**6**), quercetin is initially transformed by quercetin-3-*O*-glucosyltransferase to produce isoquercitrin (**9**) followed by a rhamnosyl transfer catalyzed by flavonol-3-*O*-glucoside *L*-rhamnosyltransferase.⁷⁸ Due to their shared biosynthetic precursor, the formations of both compounds compete.

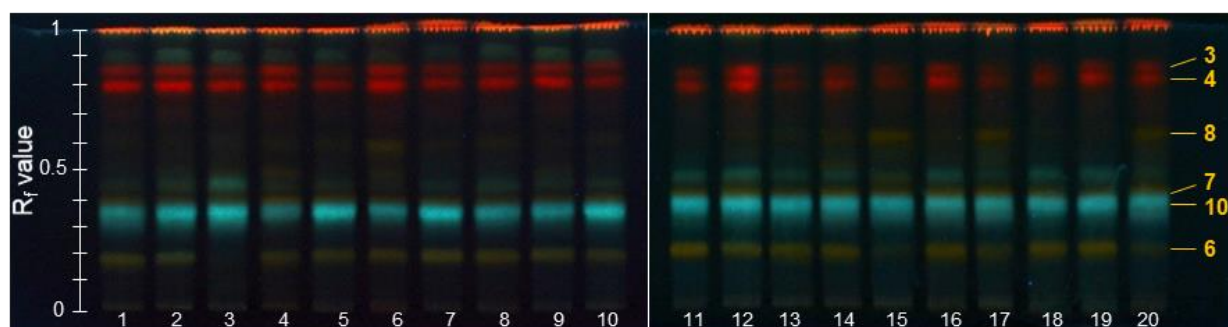


Figure 2.2 TLC fingerprints of *Hypericum perforatum* lines 1-20. Developed with the mobile phase: ethyl acetate, formic acid, acetic acid, water, dichlormethane (100:10:10:11:25 v/v/v/v/v), sprayed with natural product reagent and visualized at 366 nm. Identified compounds: hypericin (**3**), pseudohypericin (**4**), rutin (**6**), hyperoside (**7**), quercitrin (**8**), chlorogenic acid (**10**).

2.3.2 ¹H-NMR analysis

¹H-NMR measurements provide a quantitative overview of the major compounds present in plant crude extracts, the so-called metabolic fingerprint. The complexity of the ¹H-NMR spectrum of *H. perforatum* extracts is shown exemplarily in Figure 2.3a. The spectrum can be roughly classified into three parts, which contain characteristic signals of typical compound classes. The range of 2.5–0.25 ppm shows the methyl groups of the prenylated side chains of phloroglucinols and typical signals of fatty acids. In the domain from 4.7–3.0 ppm, signals corresponding to sugars are present. Signals shifted to the low field from 8.0–5.0 ppm belong to olefinic and aromatic

compounds like flavonoids. Out of the crude extract, 11 compounds could be assigned by comparison with NMR spectra of standards and confirmed by the evaluation of 2D NMR analysis using HSQC, and HMBC (Figure 2.3 b–d, Table 2.1).

The PCA of the $^1\text{H-NMR}$ experiments is shown in Figure 2.4. First, the whole spectra (excluding the solvent regions) were used to investigate the variation of the GTs (Figure 2.4a). The corresponding loadings plot explains which bins cause the clustering of the GTs (Figure 2.4b). The signals of primary metabolites like sugars (5.39–3.40 ppm), mainly sucrose (**12**) and fatty acids (**13**) (1.35–1.27 ppm), are responsible for clustering the samples. Additionally, signals which correspond to the phloroglucinol hyperforin (**1**) discriminate the GTs. In PC1 (58.1% variance), the GTs 6 and 12 are separated because of low sugar content and high hyperforin (**1**) content. Interestingly, GTs 6 and 12 are characterized as hybrids of two different statistically supported genetic backgrounds⁶⁴ (GT 12: 60% red and 40% green; GT 6: 20% red and 80% green) (Appendix 2.1). The difference to the GTs with just one genetic background may result from a heterotic effect.

PC2 (17.6%) separates GT 3 from the other lines. The low intensities of typical hyperforin (**1**) signals in combination with high fatty acid (**13**) content cause this deviation. Hyperforin (**1**) is discussed as the primary neuroactive compound in *H. perforatum* extracts. The absence or significantly lower content could lead to the loss of activity and the depreciation of the medicinal product.³⁸ Besides the positive effects, hyperforin (**1**) induces expression of cytochrome P450 enzymes which results in accelerated drug degradation, and the induction of transporter efflux proteins.^{79–81} Consequently, the application in some patients can lead to altered activity and negative drug interactions. Therefore extracts with a low hyperforin (**1**) content are of particular interest to decrease the adverse effects of St. John's Wort extracts.^{79,80}

To exclude the primary metabolites from the variance, the spectrum part from 5.5–10 ppm was chosen for further PCA analysis. These chemical shifts are characteristic of proton signals of aromatic and olefinic systems, which typically occur in secondary metabolites such as flavonoids. The variation in the aromatic region is mainly caused by flavonoids and shikimic acid (**11**) (Figure 2.4d). The loadings plot reveals a negative correlation between quercitrin (**8**) and rutin (**6**). GTs 15, 17, and 20 are characterized by a high content of quercitrin (**8**) and a low rutin (**6**) content which confirms our findings of the TLC analysis. The TLC also showed the absence of rutin (**6**) in GT 3 which contradicts the location of this GT in the PCA of the NMR. However, a detailed evaluation of the spectrum indicated that GT 3 does not contain rutin (**6**) in significant amounts but shows other signals in the characteristic regions of rutin causing the misleading placement. These signals may belong to another closely related quercetin-glycoside.

GT 6 has an outstanding position that is characterized by a high quercitrin (**8**) and rutin (**6**) content, and which is in contrast to the other three separated GTs showing a negative correlation between both metabolites. The biggest group of GTs is shifted in the positive direction of PC1 (36.5%). These GTs have increased signals of rutin (**6**), catechin (**14**), and shikimic acid (**11**). The region

between 6.73 ppm and 6.80 ppm presents overlapping signals of catechin and shikimic acid, so that the influence of each compound is not determinable.

Table 2.1 Resonance assignments of compounds in *H. perforatum* extracts from ^1H - and ^1H - ^{13}C -NMR spectra (600 MHz, methanol- d_4)

No	Compound	Assignments	^1H , δ [ppm] multiplicity (J)	^{13}C δ [ppm] (HSQC)
1	Hyperforin	CH ₃ , 12	1.07 d (6.5 Hz)	22.5
		CH ₃ , 13	1.05 d (6.5 Hz)	21.3
		CH ₃ , 14	0.93 s	15.2
		CH ₃ , 19, 20, 24, 25, 29, 30, 34, 35	1.55-1.69 s	18.1; 17.9; 18.3; 18.3; 26.1; 26.1
		CH ₂ , 26b	3.03 m	23.8
		CH, 27	5.21 m	126.8
		5	Quercetin	CH, 6
CH, 8	6.40 [*]			94.9
CH, 2'	7.72 d (2.2 Hz)			116.1
6	Rutin	CH, 6	6.21 [†]	99.9
		CH, 8	6.39 [†]	94.9
		CH, 2'	7.66 d (2.1 Hz)	117.7
		CH, 5'	6.86 d (8.3 Hz)	116.0
		CH, 1'''	4.52 [†]	102.4
		CH, 6'''	1.12 d (6.2 Hz)	17.8
7	Hyperoside	CH, 6	6.21 [*]	99.9
		CH, 8	6.40 [*]	94.9
		CH, 2'	7.83 d (2.2 Hz)	122.9
		CH, 5'	6.87 d (8.6 Hz)	116.7
8	Quercitrin	CH, 6	6.21 [*]	99.9
		CH, 8	6.37 d (2.1 Hz)	94.9
		CH, 2'	7.34 d (2.1 Hz)	117.0
		CH, 5'	6.91 d (8.3 Hz)	116.0
		CH, 6'	7.30 dd (2.1, 8.3 Hz)	122.8
		CH, 2''	4.21 dd (1.6, 3.4 Hz)	71.67
9	Isoquercitrin	CH, 6	6.21 [*]	99.9
		CH, 8	6.40 [*]	94.9
		CH, 2'	7.70 d (2.2 Hz)	117.5
		CH, 2'	7.05 d (2.1 Hz)	115.1
10	Chlorogenic acid	CH, 7'	7.59 d (15.9 Hz)	147.1
		CH, 8'	6.28 d (15.9 Hz)	127.7
		CH, 4	4.36 [†]	67.3
11	Shikimic acid	CH _{2a} , 7a	2.19 ddt (18.0, 5.1, 1.7 Hz)	31.7
		CH _{2b} , 7b	2.73 ddt (18.0, 5.1, 1.7 Hz)	31.7
		CH 1	5.39 d (3.8 Hz)	93.6
12	Sucrose	CH 2	3.42 dd (3.8, 9.8 Hz)	73.1
		CH _{2a} 6a'	3.60 d (12.4 Hz)	64.2
		CH _{2b} 6b'	3.64 d (12.4 Hz)	64.2
		CH 3'	4.09 d (8.2 Hz)	79.3
		CH ₃ terminal	0.90 [*]	14.5
13	Fatty acid	(CH ₂) _n	1.28	30.7
		(CH ₂) _n	1.32	30.3
		CH _{2a} , 4a	2.86 dd (4.7, 16.6 Hz)	29.2
14	Catechin/ Epicatechin	CH _{2b} , 4b	2.73 [*]	29.2
		CH, 6	5.94 d (2.4 Hz)	96.3
		CH, 8	5.91 d (2.4 Hz)	95.8
		CH, 2'	6.97 d (1.9 Hz)	115.3

* overlapping signals † unresolved signal

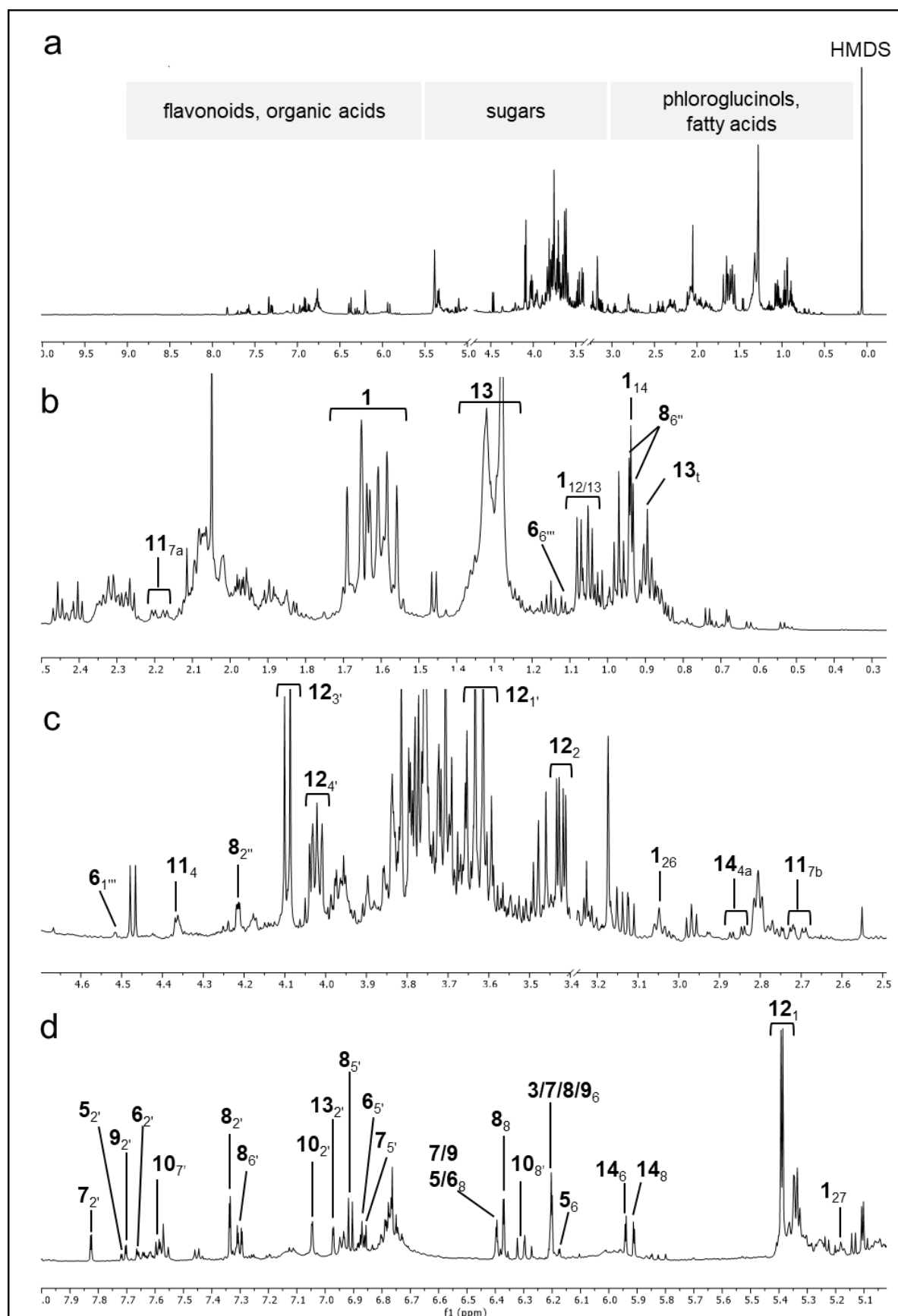


Figure 2.3 Full ^1H -NMR spectra of *H. perforatum* extract of GT 15 with characteristic regions (a). Selected sections of the spectrum that contain signals of specific compound classes: b) 0.25-2.5 ppm hyperforin (**1**) and fatty acids, c) 2.5-4.7 ppm sugars, d) 5.0-8.0 ppm phenolics. For a list of constituents assigned to peaks, see Table 2.1. Peaks were assigned using NMR spectra of standards.

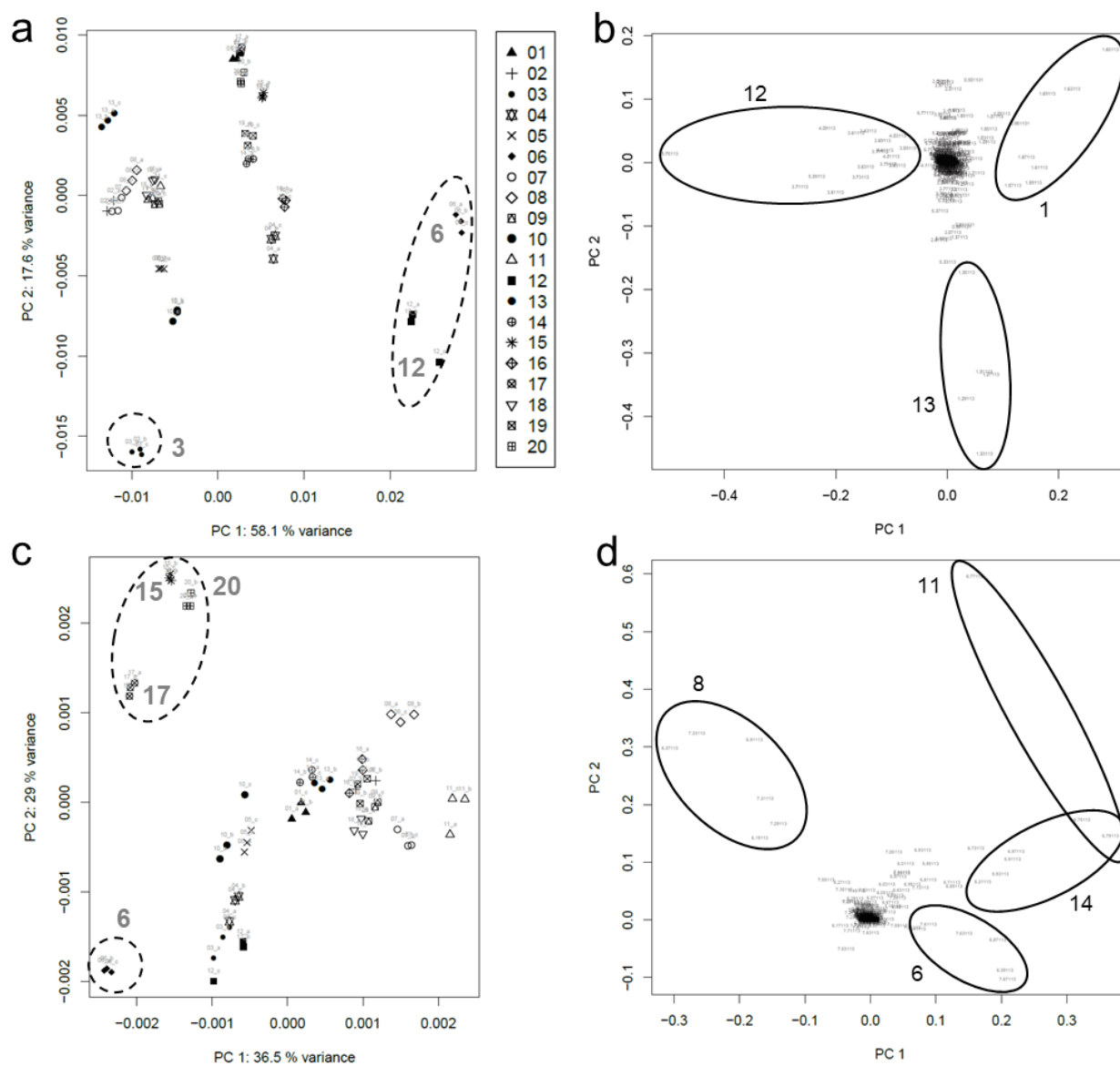


Figure 2.4 PCA of $^1\text{H-NMR}$ of *H. perforatum* extracts 1-20: a) scores plot and b) loadings plot of the full spectrum; c) scores plot and d) loadings plot of the aromatic region (5.5-10.0 ppm). Discriminating compounds are hyperforin (1), rutin (6), quercitrin (8), shikimic acid (11), fatty acid (13) and catechin/epicatechin (14).

2.3.3 UHPLC-MS analysis

The combination of ultra high performance liquid chromatography (UHPLC) with high-resolution mass spectrometry (MS) reveals insights into the chemical composition of plant crude extracts. The high sensitivity enables the detection of low concentrated compounds with respect to their ionization properties. Figure 2.5a shows the chromatogram of the pooled *H. perforatum* QC sample. Peaks (P) were tentatively assigned by their accurate mass and MS/MS fragmentation pattern. Main compounds were identified by comparison with authentic standards. The results are listed in Table 2.2.

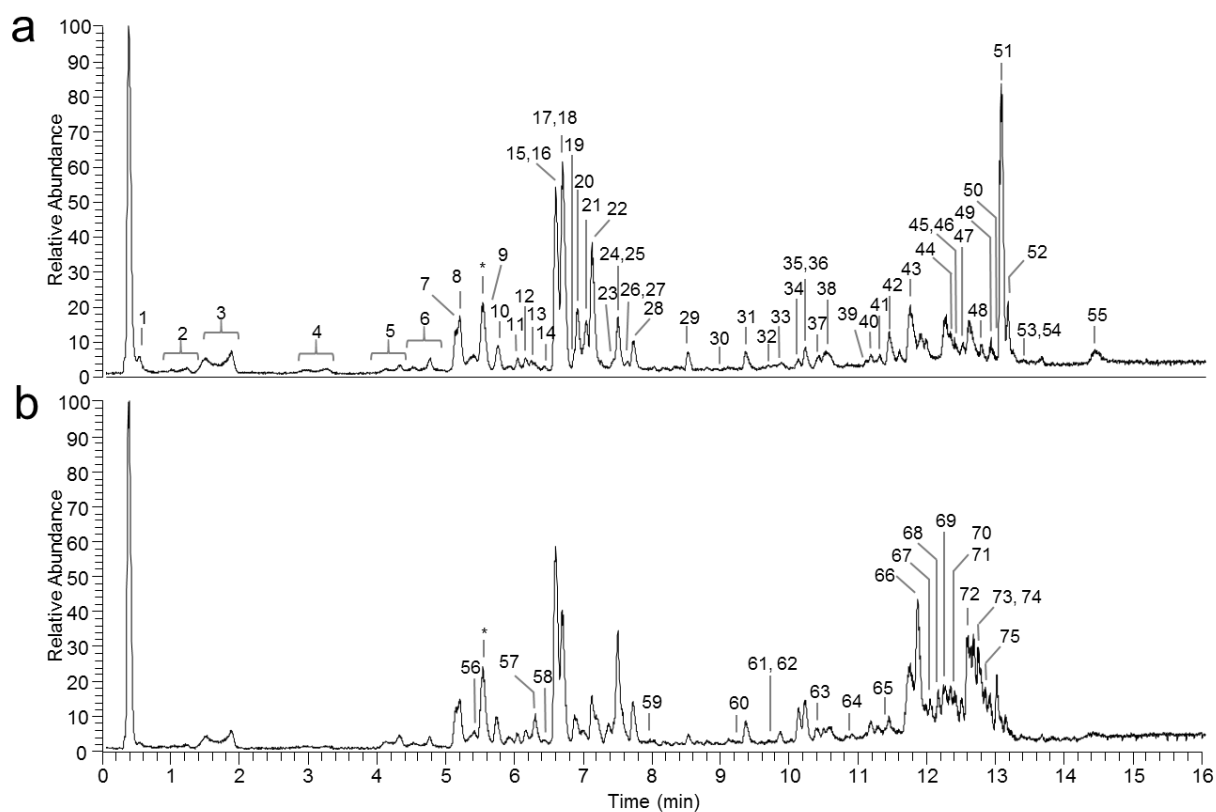


Figure 2.5 Total Ion Chromatogram of a) QC sample (pooled sample of *H. perforatum* GTs 1-20) b) GT 3. Numbers correspond to peak numbers listed in Table 2.2. The asterisk marks the internal standard umbelliferone.

The chemical variance of the 20 *H. perforatum* GTs is described in the PCA (Figure 2.6). The PCA is based on the feature table automatically generated by XCMS and displays the chemical variance of the 20 *H. perforatum* GTs. However, as a consequence of the MS technique, each compound is usually represented by more than one feature due to the appearance of isotopes and adduct ions. The 22.4% variance of PC1 is explainable by the outstanding position of GT 3. The corresponding chromatogram (Figure 2.5b) shows apparent differences in comparison to the pooled sample chromatogram. In particular, nonpolar compounds that elute between 11 and 14 min deviate significantly from the average profile.

These results were further validated by evaluating the loadings plot of PC1. In the negative direction of PC1, the features which are increased in GT 3 correspond to compounds in the phloroglucinol family. Interestingly, only compounds with less than 4 prenylated moieties are upregulated (P59, P61-P75). In contrast, the usually most abundant bicyclic polyprenylated acylphloroglucinols, hyperforin (**1**, P51), and adhyperforin (**2**, P52), are almost absent in GT 3; the corresponding features are to be found in the positive direction of PC1. Furthermore, a lower content of degradation products of these two compounds, such as P45, P46, P47, and P54 is also influencing the separation. Two features, which are not yet annotated also have a positive impact on PC1. Their molecular formulas $C_{36}H_{54}O_7$ (P48) and $C_{35}H_{52}O_7$ (P53) suggest polyprenylated phloroglucinols like those identified by Guo et al. (2018)⁸² in *H. perforatum* and Zhang et al.

(2015)⁸³ in *H. uralum*. Additionally, the content of acetylated flavonoids like P20-P23, P25, and P26 is decreased in GT 3.

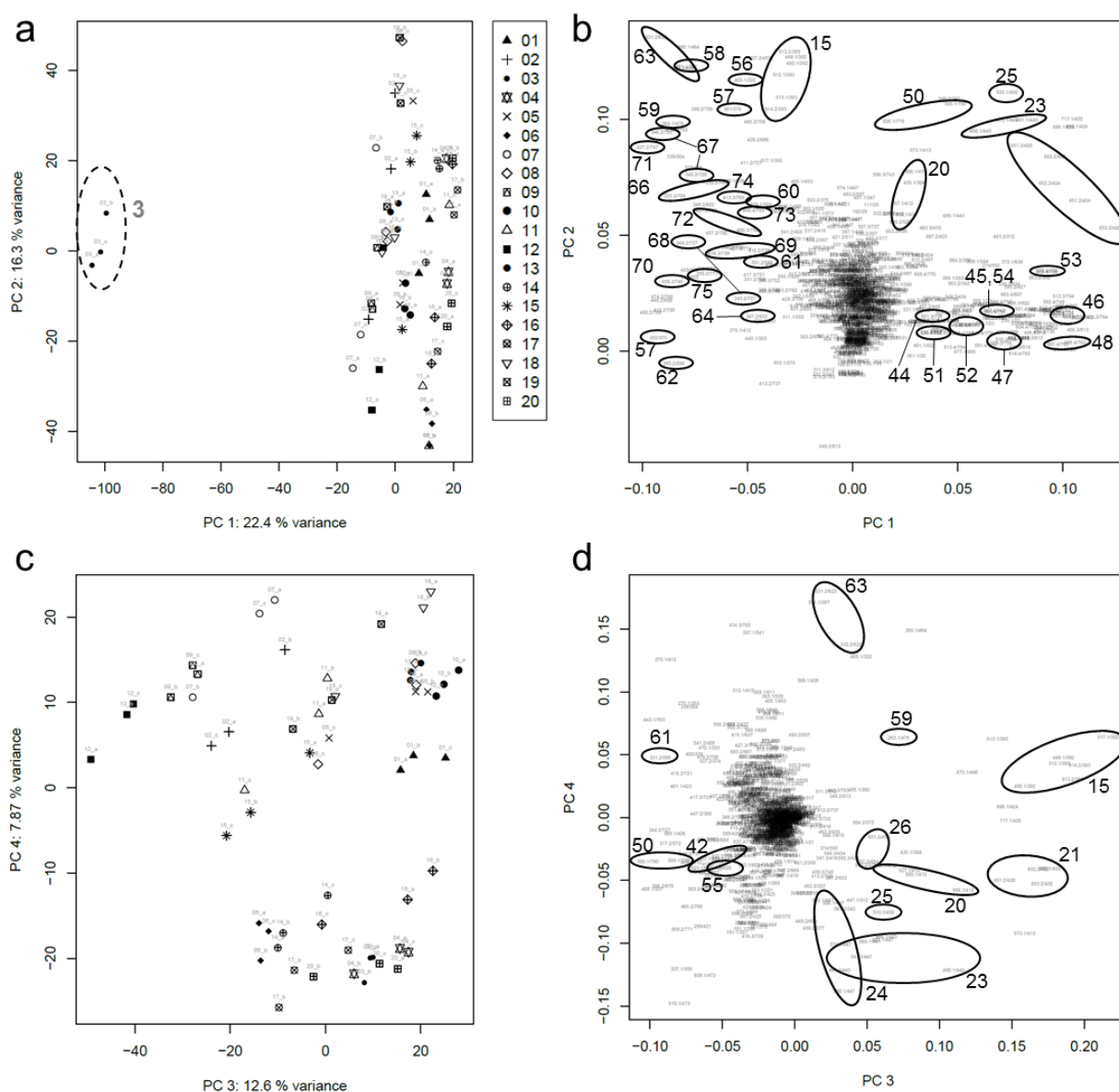


Figure 2.6 PCA of UHPLC-MS data of *H. perforatum* extracts 1-20: PC1 and PC2: a) scores plot, b) loadings plot with assigned peak numbers of Table 2.2; PC3 and PC4 c) scores plot, d) loadings plot with assigned peak numbers.

These substantial changes in the phloroglucinol profile of GT 3 could be explained by an incomplete hyperforin biosynthesis so that only precursors such as dimethylallylphlorisobutyrophenone (**15**, P59) and deoxycohumulone (**16**, P61) accumulate in the plant. The biosynthesis of hyperforin comprises the stepwise electrophilic substitutions with two DMAPP units and one GPP onto isobutyrophenone (Figure 1.3).^{33,35-37} In GT 3, the last step, the transfer of the GPP unit, seems to be missing so that higher prenylated phloroglucinols are not present. 3-Geranyl-1-(2-methylpropanoyl)-phloroglucinol (**17**, P66) was identified as the main phloroglucinol in GT 3. It is known as an antibacterial compound from *H. punctatum*.⁶⁵ Thus the application of GT 3 in antimicrobial formulations could be suitable, and our results suggest that

GT 3 is an excellent model for a comparative transcriptome analysis of the enzymes involved in hyperforin biosynthesis.

PC3 (12.6%) and PC4 (7.9%) still separate the GTs (Figure 2.6c). GTs 12 and 9 are located in the scores plot at negative values of PC3. This component is mainly influenced by increased naphthodianthrone features such as hypericin (**3**, P55), pseudohypericin (**4**, P42) and protohypericin (**18**, P50) in negative directions. In contrast to the increased content of naphthodianthrone, the amount of acetylated flavonoids P20-P23, P25, P26, and astilbin (**19**, P15) is decreased.

The PCA plots were colored to analyze whether the clusters could be explained by metadata such as the statistical genetic background, ploidy level, reproduction strategy or origin of the samples (Appendix 2.1). None of the investigated metadata was able to clearly explain the clusters. However, in PC3 and PC4 a slight clustering of GTs with blue genetic background⁶⁴ can be observed in the upper right of the scores plot. Consequently, no reliable statement can be made based on this metadata. In order to determine whether this is a real correlation, a larger number of samples with completely measured genetic information will be necessary.

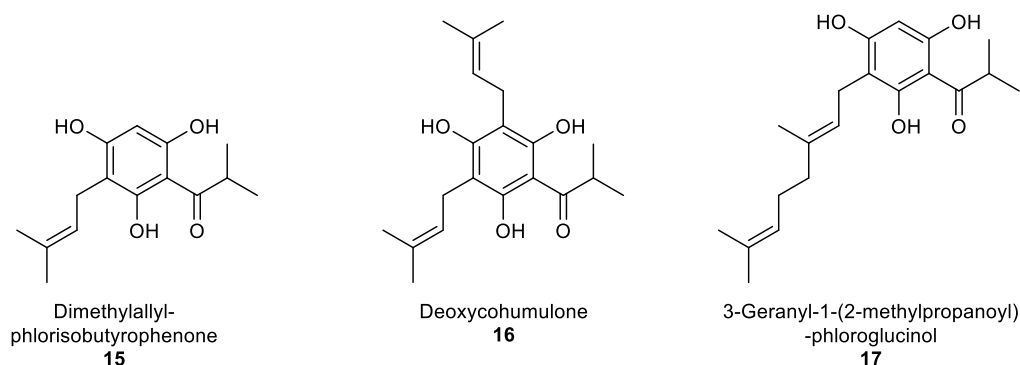


Figure 2.7 Chemical structures of polyprenylated acylphloroglucinols enhanced in GT 3.

Table 2.2 Detected LC-MS signals for *H. perforatum* extracts and their putative annotation. IL indicates the Level of identification suggested by Sumner et al. (2007).⁸⁴

No	feature	[M-H] ⁻ [m/z]	Rt [min]	Molecular Formula [M-H] ⁻	Δ [ppm]	MS ² (basepeak marked bold)	UV [nm]	Compound annotation	IL
P1	191/22	191.0197	0.48	C ₆ H ₇ O ₇	-1.205	173, 111	273, 309	Citric acid	2
P2	315.1/58 315.1/71	315.0716	0.97, 1.17	C ₁₃ H ₁₅ O ₉	-1.635	165, 153 , 152, 109	-	O-Glucosyl-hydroxy- benzoic acid	2
P3	353.1/87 353.1/109	353.0872	1.43, 1.82	C ₁₆ H ₁₇ O ₉	-1.629	191 , 179, 135	241, 301, 325	Neochlorogenic acid	2
P4	337.1/175 337.1/192	337.0920	2.90, 3.19	C ₁₆ H ₁₇ O ₈	-2.672	191, 173, 163 , 119	229, 313	Coumaroylquinic acid	2
P5	353.1/256	353.0873	4.27	C ₁₆ H ₁₇ O ₉	-1,205	191 , 179	291, 322	Chlorogenic acid (10)	2
P6	577.1/282	577.1329	4.71	C ₃₀ H ₂₅ O ₁₂	-3.863	559, 451, 425 , 407, 289, 287	281	Procyanidin I	2
P7	289.1/307	289.0712	5.09	C ₁₅ H ₁₃ O ₆	-2.046	245 , 205, 179	281	Catechin/ Epicatechin (14)	2
P8	289.1/307	289.0712	5.18	C ₁₅ H ₁₃ O ₆	-2.046	245 , 205, 179	281	Catechin/ Epicatechin (14)	2
P9	421.1/337	421.0771	5.50	C ₁₉ H ₁₇ O ₁₁	-1,436	301 , 331, 403	-	Mangiferin	2
P10	431.2/342	431.1915	5.69	C ₂₀ H ₃₁ O ₁₀	-1,995	385 , 285, 284	275	Unidentified I	4
P11	479.1/361	479.0815	5.99	C ₂₁ H ₁₉ O ₁₃	-3.431	317, 316	272	Myrecetin-glycosid	2
P12	421.1/367	421.0766	6.10	C ₁₉ H ₁₇ O ₁₁	-2.362	331, 301 , 258	285, 325	Isomangiferin	2
P13	577.1/372	577.1334	6.18	C ₃₀ H ₂₅ O ₁₂	-3.083	559, 451, 425 , 407, 289, 287	-	Procyanidin II	2
P14	451.1/383	451.0870	6.37	C ₂₀ H ₁₉ O ₁₂	-2.769	361, 341, 331	280, 327	Unidentified II	4
P15	449.1/392	449.1084	6.52	C ₂₁ H ₂₁ O ₁₁	-0,968	323, 303 , 285, 151	-	Astilbin (18)	2
P16	463.1/393	463.0876	6.54	C ₂₁ H ₁₉ O ₁₂	-1,359	301 , 300	262, 350	Hyperoside (7)	1
P17		463.0867	6.61	C ₂₁ H ₁₉ O ₁₂	-3.194	-	-	Isoquercitrin (9)	2
P18	609.1/399	609.1453	6.65	C ₂₇ H ₂₉ O ₁₆	-1.310	343, 301 , 300, 271, 255,	262, 357	Rutin (6)	1
P19	433.1/409	433.0767	6.81	C ₂₀ H ₁₇ O ₁₁	-2.089	343, 302, 301, 300	-	Avicularin (Quercetin- arabinoside)	2
P20	505.1/412	505.0976	6.87	C ₂₃ H ₂₁ O ₁₃	-2.383	463, 445, 301, 300	265, 358	Quercetin-acetylglycoside I	2
P21	651.2/420	651.1547	6.99	C ₂₉ H ₃₁ O ₁₇	-2.983	609, 591 , 301, 300, 299, 271	265, 358	Acetylrutin I	2
P22	447.1/425	447.0924	7.07	C ₂₁ H ₁₉ O ₁₁	-1.934	301 , 300	265, 358	Quercitrin (8)	1
P23	489.1/443	489.1024	7.38	C ₂₃ H ₂₁ O ₁₂	-3.024	285 , 284, 300, 429, 443	-	Kampferol-acetylglycoside	2
P24	505.1/447	505.0976	7.44	C ₂₃ H ₂₁ O ₁₃	-2.383	300 , 301, 445, 463	265, 358	Quercetin-acetylglycoside II	2
P25	533.1/450	533.0932	7.5	C ₂₄ H ₂₁ O ₁₄	-0.804	489 , 371	-	Unidentified III	4
P26	651.2/454	651.1553	7.56	C ₂₉ H ₃₁ O ₁₇	-2.046	609, 591 , 301, 300, 271	-	Acetylrutin II	2
P27	337/456	337.0019	7.58	C ₁₄ H ₉ O ₈ S	-1.962	337, 321 , 257, 242	-	1,3-Dihydroxy-5- methoxyxanthone-4- sulfonic acid	2
P28	301/461	301.0347	7.67	C ₁₅ H ₉ O ₇	-2.312	273, 257, 178 , 151	370	Quercetin (5)	1
P29	537.1/509	537.0812	8.46	C ₃₀ H ₁₇ O ₁₀	-2.774	493, 444, 443 , 417, 386, 385,	270, 338	I3,I18-Biapigenin	2
P30	259/541	259.0242	9.02	C ₁₃ H ₇ O ₆	-2.128	-	-	Tetrahydroxyxanthone	2
P31	327.1/559	327.0864	9.30	C ₁₈ H ₁₅ O ₆	-3.153	327, 291, 284, 283, 271 , 272, 258, 229, 171	-	Isoprenyltetrahydroxy- xanthone I	2
P32	257/583	257.0454	9.71	C ₁₄ H ₉ O ₅	-0.726	243, 242	-	Dihydroxymethoxy- xanthone	2
P33		327.0867	9.81	C ₁₈ H ₁₅ O ₆	-2.083	327 , 326, 298, 297, 258	-	Isoprenyltetrahydroxy- xanthone II	2
P34	327.1/605	327.0869	10.08	C ₁₈ H ₁₅ O ₆	-1.564	327 , 326, 308, 283, 272, 258	-	Isoprenyltetrahydroxy- xanthone III	2
P35	341.1/610	341.1025	10.16	C ₁₉ H ₁₇ O ₆	-1.734	326 , 283	-	Trihydroxy-methoxy- isoprenyl-xanthone	2
P36	357.1/611	357.0975	10.18	C ₁₉ H ₁₇ O ₇	-1.249	342, 329, 326, 325 , 313, 298	-	Unidentified IV	4

2. Characterization of pharmaceutical extracts from 20 *H. perforatum* genotypes

No	feature	[M-H] ⁺ [m/z]	Rt [min]	Molecular Formula [M-H] ⁺	Δ [ppm]	MS ² (basepeak marked bold)	UV [nm]	Compound annotation	IL
P37	327.1/622	327.0869	10.35	C ₁₈ H ₁₅ O ₆	-1.564	327 , 326, 312, 297, 284, 283, 271, 272, 258	-	Isoprenyltetrahydroxanthone IV	2
P38	325.1/628	325.0710	10.49	C ₁₈ H ₁₃ O ₆	-2.435	325, 311, 310 , 309, 307	-	Unidentified V	4
P39	669.1/633	669.1233	10.53	C ₃₅ H ₂₅ O ₁₄	-2.562	519	-	S-Skyrin-6-O-β-arabinofuranoside	1
P40	699.1/664	699.1346	11.06	C ₃₆ H ₂₇ O ₁₅	-1.321	519	-	S-Skyrin-6-O-β-glucopyranoside	1
P41	339.1/668	339.0867	11.13	C ₁₉ H ₁₅ O ₆	-2.039	324 , 339, 325, 183	-	Paxanthone	2
P42	521.1/682	521.0864	11.32	C ₃₀ H ₁₇ O ₉	-2.678	521 , 477	-	Protopseudohypericin	2
P43	519.1/702	519.0709	11.62	C ₃₀ H ₁₅ O ₉	-2.380	519 , 503, 487	285, 334, 580	Pseudohypericin (4)	2
P44	501.3/736	501.3000	12.26	C ₃₃ H ₄₁ O ₄	-2.001	-	-	Unidentified VI	4
P45	567.4/742	567.3684	12.36	C ₃₅ H ₅₁ O ₆	-1.185	549, 535 , 523, 498, 399, 329	-	33-Deoxy-33-hydroperoxyfurohyperforin	2
P46	551.4/743	551.3728	12.38	C ₃₅ H ₅₁ O ₅	-2.572	482 , 411, 413, 383, 397, 329, 315, 343, 399	-	Furohyperforin	2
P47	509.3/756	509.3283	12.60	C ₃₂ H ₄₅ O ₅	2.046	481 , 465, 439, 398, 371, 327, 313	-	Furohyperforin a	2
P48	597.4/765	597.3778	12.74	C ₃₆ H ₅₃ O ₇	-3.193	551 , 528, 483, 445	-	Unidentified VII	4
P49	591.3/773	591.2604	12.86	C ₃₄ H ₃₉ O ₉	0.863	515 , 559	285, 409	Unidentified VIII	4
P50	505.1/780	505.0923	13.00	C ₃₀ H ₁₇ O ₈	-1.209	505 , 461	-	Protohypericin (18)	2
P51	535.4/782	535.3790	13.03	C ₃₅ H ₅₁ O ₄	-0.529	466 , 397, 383, 315, 313	275	Hyperforin (1)	2
P52	549.4/788	549.3937	13.12	C ₃₆ H ₅₃ O ₄	-2.300	480 , 411, 397, 329, 313	-	Adhyperforin (2)	2
P53	583.4/788	583.3631	13.12	C ₃₅ H ₅₁ O ₇	-1.521	525, 495, 455, 441, 385 , 329, 275	-	Hyperforatone I/J/K	2
P54	565.4/787	565.3761	13.12	C ₃₆ H ₅₃ O ₅	-2.066	496, 426, 425	-	Furoadhyperforin	2
P55	hyp503	503.0769	13.71	C ₃₀ H ₁₅ O ₈	1,649	503 , 461, 459	285, 435	Hypericin (3)	1
P56	405.1/322	405.0814	5.37	C ₁₉ H ₁₇ O ₁₀	-3.184	243, 271, 229	-	Dihydroxyxanthone-glucoside	2
P57	381/375	380.9911	6.24	C ₁₅ H ₉ O ₁₀ S	-2.757	301 , 261	-	Quercetin-sulfate	2
P58	371.1/387	371.1339	6.44	C ₁₇ H ₂₃ O ₉	-2.386	283, 209 , 139	-	Methylbuteryl-phloroglucinol-glucoside	2
P59	263.1/476	263.1285	7.93	C ₁₅ H ₁₉ O ₄	-1.453	194 , 166	-	Dimethylallyl-phlorisobutyrophenone (15)	1
P60	325.1/553	325.0712	9.21	C ₁₈ H ₁₃ O ₆	-1.727	-	-	Xanthone	2
P61	331.2/595	331.1908	9.91	C ₂₀ H ₂₇ O ₄	-2.091	262 , 234, 194, 166	-	Deoxycoumestrolone (16)	2
P62	365.2/599	365.1961	9.98	C ₂₀ H ₂₉ O ₆	-2.442	347, 321, 295	-	Unidentified Phloroglucinol	3
P63	331.1/623	331.1907	10.38	C ₂₀ H ₂₇ O ₄	-2.242	262, 194 , 166, 151	-	Unidentified Phloroglucinol	3
P64	347.2/652	347.1855	10.86	C ₂₀ H ₂₇ O ₅	-2.642	329, 311, 303, 277, 195 , 151	-	Unidentified Phloroglucinol	3
P65	429.2/686	429.2271	11.4	C ₂₅ H ₃₃ O ₆	-2.637	429 , 401, 385, 359	-	Unidentified Phloroglucinol	3
P66	331.2/709	331.1907	11.82	C ₂₀ H ₂₇ O ₄	-2.303	313, 287 , 261, 207, 194	-	3-Geranyl-1-(2-methylpropanoyl)-phloroglucinol (17)	1
P67	345.2/720	345.2065	12.00	C ₂₁ H ₂₉ O ₄	-1.891	345, 327, 301 , 276, 261, 221	-	Unidentified Phloroglucinol	3
P68	343.2/727	343.1906	12.11	C ₂₁ H ₂₇ O ₄	-2.455	325, 315 , 274, 221	-	Unidentified Phloroglucinol	3
P69	416.3/739	415.2482	12.30	C ₂₅ H ₃₅ O ₅	-1.944	397, 357 , 343, 289, 275	-	Unidentified Phloroglucinol	3
P70	459.2/745	459.2742	12.41	C ₂₇ H ₃₉ O ₆	-2.204	429, 427 , 412, 399, 357, 342, 329, 314	-	Unidentified Phloroglucinol	3
P71	427.2/747	427.2478	12.45	C ₂₆ H ₃₅ O ₅	-2.803	412, 343	-	Unidentified Phloroglucinol	3
P72	400.3/752	399.2531	12.51	C ₂₅ H ₃₅ O ₄	-2.562	399, 355 , 330, 287, 275, 261, 219	-	Unidentified Phloroglucinol	3
P73	496.4/769	495.3474	12.74	C ₃₂ H ₄₇ O ₄	-1.278	426, 411 , 357, 289	-	Unidentified Phloroglucinol	3
P74	413.3/765	413.2695	12.75	C ₂₆ H ₃₇ O ₄	-0.539	369, 343, 276, 275, 233 , 221, 208	-	Unidentified Phloroglucinol	3
P75	359.2/771	359.1855	12.85	C ₂₁ H ₂₇ O ₅	-2.498	359, 341, 331, 290 , 275, 273, 261, 247, 235	-	Unidentified Phloroglucinol	3

2.3.4 Evaluation of biological activities

The number of studies that report positive effects of *H. perforatum* extracts on neurodegenerative diseases continues to increase.^{11,85,86} Current medications against Alzheimer's disease inhibit the enzyme acetylcholinesterase (AChE), therefore the extracts were tested in an AChE inhibition assay. None of the extracts showed an inhibition higher than 25% at the highest test concentration of 200 µg/ml, and thus no such effect can be claimed (Figure A2.1b). The results are consistent with those of Hernandez et al. (2010), where IC₅₀ values greater than 1000 µg dry extract/ml were reported.⁸⁵

Neuronal health is promoted by compounds with antioxidative effects,⁸⁷ and the DPPH-radical scavenger assay assesses the chemical suitability of providing such an effect. At the highest concentration tested (166.7 µg/ml), the extracts showed a scavenging rate of 40 to 60% (Appendix 2.1a). Although the effect is lower than that reported by other studies, phenolic compounds like the flavonoids, which are well known for antioxidative effects, are most likely responsible.⁸⁸⁻⁹² The chemical differences in the flavonoid profiles of the 20 GTs could be responsible for the fluctuating values in the competition assay. Most importantly, the results suggest that the positive effect of *H. perforatum* on neurodegenerative diseases are unlikely to depend on general antioxidative properties, but may be connected to another target like the ABCC1 transporter activation described by Hofrichter et al. (2013).¹¹

Cytotoxicity of the *H. perforatum* extracts was tested against two cancer cell lines (PC3, HT29). Cell viability was not affected by the application of the extracts at different concentrations (0.05 µg/ml up to 50 µg/ml; Figure Appendix 2.1c), and thus low cytotoxicity underlines the suitability of using *H. perforatum* medicinally.

Even though the 20 GTs vary in their metabolite composition, the biological activities tested in this study were comparable. Due to the complex structure of the raw extracts, the effects of the different constituents are overlapped and averaged. This could be a reason why GTs, which were not in line with the Ph. Eur., showed similar results.

2.3.5 Comparison of the different analytical methods

TLC and NMR offer a simple and fast screening of major compounds. With the obtained fingerprints, extracts from 18 *H. perforatum* GTs could be determined as medicinally acceptable preparations as described by the Ph. Eur.¹⁵ In these cases, TLC showed all species characteristic marker compounds, although quantitative differences in the flavonoids were detectable. The GTs 3 and 20 did not contain rutin (**6**) in detectable amounts and GTs 15 and 17 just low amounts. These results were in line with the NMR experiments.

The NMR technique also provides a comparable overview of the flavonoid profiles, it is not restricted to UV detectable compounds, and independent of the elution behavior depending on the mobile phase. Thus, changes in primary metabolites and hyperforin were also detectable, an important discovery considering their influence on the bioactivity of the final medicinal product.

As already described in other NMR studies of *H. perforatum*,^{52,59} naphthodianthrones were not detectable in ¹H-NMR spectra using deuterated methanol as solvent due to the occurrence of keto-enol tautomers and low concentrations. In general, the NMR analysis is a powerful tool to determine the quality of *H. perforatum* extracts, it gives insights into the chemical variation of primary and secondary constituents, and high reproducibility enables the comparison of a huge number of measurements over time.

The LC-MS analysis covers all essential classes of secondary metabolites in *H. perforatum* ethanolic extracts. Thus, it was possible to detect changes in the phloroglucinol profile in GT 3. Also, minor compounds are detectable due to the high sensitivity, exemplified by the detection of low abundant acetylated flavonoids (P20-P24). The biological effects of these compounds are not described yet, but their occurrence is highly variable and should be considered in future investigations. LC-MS is the most common technique used in metabolomics experiments; however, one disadvantage in contrast to NMR is low reproducibility, with results depending strongly on the specific equipment and conditions used. In particular, the elution of naphthodianthrones is dramatically influenced by pH changes. The buffering of the mobile phase solvents is therefore strongly recommended to decrease pH changes during long measuring periods.

2.4 Conclusion

Altogether, the applied untargeted metabolomics concept based on TLC, NMR and LC-MS combined with multivariate data analysis enables the investigation of the chemodiversity within the species. All analytical techniques used in this comparative study could discern differences in the metabolite composition of pharmaceutical extracts of 20 *H. perforatum* GTs. However, unexpectedly this variance could not be correlated to ploidy level, reproduction strategy or origin of the samples. We could show that research on the chemical composition is necessary to yield high-quality raw material for the production of medicinal products. The data set included one genotype without the valuable compounds rutin (**6**) and hyperforin (**1**), but showed an increased level of only partially prenylated precursors of hyperforin. The results reveal that comparative metabolite profiling can help breeders to select genotypes producing the desired bioactive compounds according to the proposed pharmacological application of the plant.

3 New insights into the hypericin biosynthesis by analysis of *H. perforatum* pistil tissue

This chapter includes the metabolomic analysis of pistil tissue for the elucidation of the hypericin biosynthesis. These investigations were part of a cooperative work, which was published as:

Discovery of key regulators of dark glands development and hypericin biosynthesis in St. John's wort (*Hypericum perforatum*).

Paride Rizzo, Lothar Altschmied, Pauline Stark, Twan Rutten, André Gündel, Sarah Scharfenberg, Katrin Franke, Helmut Bäumlein, Ludger Wessjohann, Marcus Koch, Ljudmilla Borisjuk and Timothy F. Sharbel.

Plant Biotechnology Journal (2019) 17, pp. 2299–2312, doi: 10.1111/pbi.13141.

Abstract

Hypericin is a molecule of high pharmaceutical importance that is synthesized and stored in dark glands (DGs) of St. John's Wort (*Hypericum perforatum*). Understanding which genes are involved in dark gland development and hypericin biosynthesis is important for the development of new *Hypericum* extracts that are highly demanded for medical applications. We identified two transcription factors whose expression is strictly synchronized with the differentiation of DGs. We correlated the content of hypericin, pseudohypericin, endocrocin, skyrin glycosides, and several flavonoids with gene expression and DG development to obtain a revised model for hypericin biosynthesis. Here, we report for the first time genotypes which are polymorphic for the presence/total absence (G+/G-) of DGs in their placental tissues (PTs). DG development was characterized in PTs using several microscopy techniques. Fourier transform infrared microscopy was established as a novel method to precisely locate polyaromatic compounds, such as hypericin, in plant tissues. In addition, we obtained transcriptome and metabolome profiles of unprecedented resolution in *Hypericum*. This study addresses for the first time the development of dark glands and identifies genes that constitute strong building blocks for the further elucidation of hypericin synthesis, its manipulation in plants, its engineering in microbial systems, and its applications in medical research.

3.1 Introduction

The genus *Hypericum* has a worldwide distribution and includes more than 460 species occupying very diverse habitats.⁹ The most known representative, *Hypericum perforatum*, better known as St. John's Wort, is not only a model organism for the study of apomixis (asexual reproduction) but also an ancient medicinal plant mainly used for the treatment of depressions.⁹³⁻⁹⁵ Its secondary metabolites, hypericin (**3**) and hyperforin (**1**), are among the most intensively studied bioactive compounds.^{55,96} Whereas hyperforin (**1**) is contained in translucent (or pale) glands^{30,97} which give the leaf a perforated appearance and hence the species its scientific name, hypericin (**3**) on the other hand accumulates in so-called dark glands (DGs) which differentiate in leaves and other tissues.⁵⁵ Hypericin (**3**) is considered a promising agent for cancer photodynamic therapy²¹⁻²³ being classified as a type II immunogenic cell death (ICD) inducer.²⁴ By targeting the endoplasmic reticulum, hypericin (**3**) induces cellular damage and apoptosis signalling.²⁴ Additionally, hypericin (**3**) may also function as an inhibitor of β -amyloid fibril formation which has opened applications in the treatment of Alzheimer's disease.^{98,99} It has been shown that crude extracts of *H. perforatum* can reduce the memory impairment in amyloid precursor protein (APP)-transgenic mice.¹¹

Despite its importance, knowledge about the biosynthesis of hypericin is still incomplete (Figure 1.2).^{27,35,100,101} An octaketide synthase (OKS;¹⁰²) catalyzing the condensation of acetyl-CoA with seven malonyl-CoAs yields octaketide products, but not the expected cyclic forms.³⁵ For the Hyp-1 protein¹⁰³, initially thought to catalyze the dimerization of emodin (**20**) and emodin anthrone (**21**), as well as further C-C bond formation between the naphthodianthrone halves, it has been shown that the mRNA is expressed in *H. perforatum* tissues unrelated to dark glands.¹⁰⁰ Therefore, based on mRNA expression data Soták et al. (2016a) suggested that these functions are encoded by the POCP genes, but did not provide any functional data for the respective proteins.²⁷ Very recently, Kimáková et al. (2018) raised doubts on emodin anthrone (**21**) and emodin (**20**) as intermediates and suggested an important role of skyrin (**22**) in hypericin biosynthesis.¹⁰¹ The ultimate accumulation of hypericin (**3**) within dark glands is surrounded by even more uncertainty, except some structural hints on vesicle transport by Onelli et al. (2002).¹⁰⁴

Though the presence of DGs correlates with hypericin (**3**) content,⁵⁵ the triggers towards the initiation and ultimate differentiation of these organs are unknown. Leaves are the model organ for most studies on hypericin (**3**) and DGs.^{27,104,105} Highest concentrations of hypericin (**3**), however, are found in the flowers,¹⁰⁶ and pistils may be far more suited to study both hypericin biosynthesis and the formation of DGs which differentiate from the placental tissue.

By phenotyping 93 accessions, we discovered a polymorphism consisting of glanded (G+) and glandless (G-) placental tissues (PTs). The glands in the placenta can reach a very high density and, in contrast to their foliar counterpart, differentiate much later in development. The correlation between dark glands and hypericin (**3**) accumulation was validated by fluorescence and FTIR

microscopy (data not shown). The late development of placental DGs enabled developmental studies in which pre- and post-DG differentiation stages were compared. This is the first time that the development of dark glands is addressed at the molecular level. Metabolomics and transcriptomics data identified novel compounds associated with DGs as well as regulatory genes associated with the development of these organs and provided novel candidate genes associated with hypericin biosynthesis. Our study demonstrates that the placenta of *H. perforatum* is a novel highly sensitive model tissue for the study of DGs and associated biosynthetic pathways and provides novel insights into these processes.

3.2 Experimental

3.2.1 Plant material and growth conditions

The 93 wild accessions of *H. perforatum* L. used in this study cover the complete range of ploidy, mode of reproduction, and genetic backgrounds as determined by Koch et al. (2013)¹⁰⁷ and Molins et al. (2014)⁶⁴ (Appendix 3.1). Fifty seeds per genotype were sown 1 cm deep in 12 x 10 cm pots and kept in long-day conditions: 16 h light, 250 $\mu\text{mol/s/m}^2$ at 21°C; and 8 h darkness at 18°C. High humidity levels were kept by using germination capsules with translucent caps. Four weeks after sowing, seedlings were transferred into 16 x 15 cm pots and moved to a greenhouse with 16 h of light, at 300–400 $\mu\text{mol/s/m}^2$, at 21°C, and 8 h of darkness at 17°C, at 70% average humidity. Three months after germination, plants destined for pistil phenotyping were transferred to the field. Plants used for RNAseq experiments were grown in a phytotron with 16 h of light at 400–450 $\mu\text{mol/s/m}^2$ at 23°C, and 8 h of darkness at 18°C at 70% humidity.

3.2.2 Phenotyping and histological analysis

All field-grown accessions were analyzed under a Stemi 2000 Zeiss stereomicroscope for the presence or absence of dark glands within the pistil. Three open flowers per plant and five plants per accession for a total of 1395 flowers were examined. Genotypes with up to 40 glands per PT were classified as G+ PT, and those with >40 dark glands per PT were classified as G++ PT. Genotypes without placental dark glands were classified as glandless (G- PT). For histological studies, isolated pistils were fixed and embedded in Spurr resin according to Rutten et al. (2003)¹⁰⁸. Semi-thin sections (2 μm) were cut on a Reichert-Jung Ultracut S (Leica, Vienna, Austria), stained with crystal violet and examined in a Zeiss Axio Imager light microscope (Carl Zeiss, Jena, Germany).

3.2.3 Metabolite analyses

3.2.3.1 Extraction of plant material

For a comparative chemical analysis, pistils from open flowers of three glanded (G++ PT: H06-1988, HyPR-3, HyPR-1) and three glandless PT genotypes (G-PT: H06-1498, H06-1369, H06-

3220) were selected. For each genotype, three biological replicates comprising ten pistils each were used. After lyophilization, samples were ground in a ball mill (MM 400; Retsch, Haan, Germany) for 30 s at 30 Hz. The resulting powder was mixed with LC-MS-grade methanol (10 mg/mL) containing 8 $\mu\text{g/mL}$ umbelliferone (HPLC-grade; Sigma, Darmstadt, Germany) as an internal standard. After brief mixing on a vortex, extraction was continued in an ultrasonic bath for 15 min. Following centrifugation (15 min, $14,000 \text{ min}^{-1}$) the supernatant was used for UHPLC-MS/PDA and TLC analysis.

3.2.3.2 TLC analysis

Pistil extracts (see above, replicates combined) were applied to SilicaGel plates (60 G F254, 9 x 7 cm; Merck, Darmstadt, Germany). The TLC procedure, adapted from Ernst (2003)⁶⁶, is described in detail in chapter 2.2.4.

3.2.3.3 High-resolution UHPLC-MS/PDA analysis

Negative ion high-resolution ESI mass spectra were obtained as described in the above chapter 2.2.7.

3.2.3.4 MS raw data processing and multivariate data analysis

MS data processing was performed in R with the XCMS package (version 1.52.0, bioconductor.org). Xcalibur raw output files (*.raw) were converted into standard format mz-Data files (*.mzML) utilizing proteowizard (proteowizard.sourceforge.net). Peak picking was performed in XCMS with centWave parameters: ppm = 10, peakwidth = c(5,12), snthr = 5 and prefilter = c(3,1500). After peak grouping (minfrac = 1, bw = 5, mzwid = 0.002), retention time correction was performed using LOESS correction and peak grouping was repeated. Missing values were filled with fillPeaks function. The final data matrix contained features (mz/RT) in rows and samples in columns. For data evaluation, an output table was used for partial least squares discriminant analysis (PLS-DA) performed with the R package pls (version 2.6-0). Gland presence was coded as the Y variable (G++ PT = 1, G- PT = 0). Results were further statistically analyzed to determine significant differences in feature intensity. Homogeneity of variance was checked with an F-test to decide the usage of the t-test or the variance independent Welch two-sample t-test. Effect size and power of analysis were calculated feature-wise. The identification of known compounds was based on exact mass of detected ions and fragmentation patterns compared to massbank.eu.

3.2.4 RNA extraction

Total RNA was extracted from placental tissues at different stages of flower development as measured by flower bud (FB) length. The FB classes selected were 2.5 to 3.5 mm (FB25; dark gland predifferentiation), 4.5 to 5.5 mm (FB45; dark gland differentiation) and 7.5 to 8.5 mm (FB75; dark gland postdifferentiation). Placental tissues were isolated from the pistils and ovules removed. After initial storage on ice in a 70% ethanol solution containing 0.1% Tween-20, samples

were frozen in liquid nitrogen and stored at -80°C . For each developmental stage, PTs from at least 20 individuals were collected. The large number of PTs enabled analyses to be performed without RNA amplification steps. Samples were ground using six 3-mm metal beads in a Retsch grinder running at 30 Hz for three minutes. RNA was extracted by InviTrap Spin Plant RNA Mini Kit following the manufacturer's instructions. Finally, total RNA was screened by an Agilent 2100 Bioanalyzer securing that only samples with $\text{RIN} > 9$ were used for subsequent sequencing applications.

3.2.4.1 RNA sequencing and data processing

cDNA libraries were prepared from 400 to 1000 ng of total RNA using a Lexogen SENSE RNA-Seq Kit following instructions of the manufacturer. The libraries were sequenced using a HiSeq 2500 high-throughput flow cell. A sequencing output of $15 \text{ to } 20 \times 10^6$ 100-nt-long single reads was obtained. The complete RNAseq data set was deposited to the European Nucleotide Archive (ENA) with accession number PRJEB30287. Data quality was assessed using FastQC software. Adapter trimming was performed using the standard settings of the Cutadapt software¹⁰⁹. Quality trimming was performed using the command `CLC_quality_trim` from the CLC Assembly Cell software (version 5.0.1; Qiagen, Hilden, Germany). *De novo* transcriptome assembly was carried out with Trinity¹¹⁰. The resulting contigs were compared with known *H. perforatum* proteins and with *Arabidopsis thaliana* (ARAPORT11) as well as *Ricinus communis* (v0.1 Phytozome) proteins using BLASTX (E-value $\leq 10^{-10}$). Read mapping to the assembly was carried out with Kallisto¹¹¹ for each library separately. Read counts for all isoforms of a gene fragment were summed and used for calculation of differential expression using the R package DESeq2¹¹². P-values were adjusted for multiple testing using the Benjamini–Hochberg method. The adopted thresholds for calling differential expression were as follows: $\text{FDR} \leq 0.01$ and absolute \log_2 fold change ≥ 1 .

3.3 Results and discussion

3.3.1 Polymorphism in placental tissue used for the study of dark gland related biosynthetic processes

Under natural conditions, dark glands occur in most organs of *H. perforatum*, except roots, although hormone induced dark glands in lateral root cultures were reported by Murthy et al. (2014)¹¹³. Although flowers contain the highest concentrations of hypericin (**3**),¹¹⁴ research on hypericin biosynthesis nearly exclusively focused on the leaves.^{27,104,105,115} Since leaves without dark glands are unknown in *H. perforatum*, it has been common practice to separate leaf lamina from leaf rim to compare between glanded and glandless tissues.^{27,105}

The presence of hypericin (**3**) in pistils is usually overlooked, with few exceptions.¹¹⁶ *H. perforatum* has a pistil composed of three carpels with parietal-to-axial placentation. Carpels

with parietal placentation show that ovules arise submarginally whereas dark glands arise marginally (Figure 3.1) as they do in related organs like leaves, petals, and sepals.⁶⁷

The 93 analyzed accessions (Appendix 3.1) show a large variation in the occurrence of placental DGs. Twenty-one genotypes with on average >40 of DGs per PT were classified as carrying heavily glanded placentas (G++ PT), while 40 genotypes lacking DGs completely were classified as G- PT. The glanded and glandless phenotypes were confirmed by histological studies (Figure 3.1c,d), which showed that the G- phenotype was not due to retarded or aborted gland development.⁶⁷ The remaining 32 accessions displaying an average of 1 to 40 DGs per PT were classified as G+ PT. The most heavily glanded G++ PTs packed over 130 DGs per pistil. G-accessions always display dark glands in other organs like leaves, sepals, petals, and anthers. This indicates that the mechanism for DG formation is present in these genotypes, but inactive in placental tissues.

During flower development, also placental DGs evolve (Figure 3.1e–i). Based on morphology changes of the glands during this process, three developmental phases can be defined: predifferentiation (FB shorter than 4.5 mm; Figure 3.1e,f), differentiation (FB 4.5–5.5 mm; Figure 3.1g), and postdifferentiation (FB > 5.5 mm; Figure 3.1h,i). This classification was used for the subsequent multistage transcriptomic study performed here.

3.3.2 Metabolic validation and characterization of the dark gland's presence-absence phenotypes

3.3.2.1 TLC analysis

TLC analysis confirmed the presence of hypericin (**3**) in extracts from G++ accessions and its absence from G- accessions (Figure 3.2a). In the G++ accessions, the hypericin band (**3**, $R_f = 0.87$) was accompanied by a further prominent signal with red fluorescence, which was assigned to the closely related pseudohypericin (**4**, $R_f = 0.81$). This additional signal was missing from the G- accessions with the exception of a weak band in genotype 4 (H06-1369, Figure 3.2a). This result is in accordance with the assumption that naphthodianthrone are restricted to dark glands and should therefore be absent in G- pistils.

3.3.2.2 UHPLC-HRMS analysis

To gain additional information, metabolite profiles of these extracts were obtained by UHPLC-ESI-HRMS/PDA. Supervised PLS-DA was used to distinguish between the differentially glanded lines. The evaluation of the RMSEP plot (root-mean-squared error of prediction; Figure 3.3a) indicates that the first two components sufficiently describe the model. The score plot, shown in Figure 3.3b, confirms a good separation of the two phenotypes based on component 1. According to the correlation loadings plot (Figure 3.3c), the most promising features that contribute to the

separation are described in Table 3.1. Statistical tests conducted on the selected correlated features from Table 3.1 are reported in Appendix 3.2.

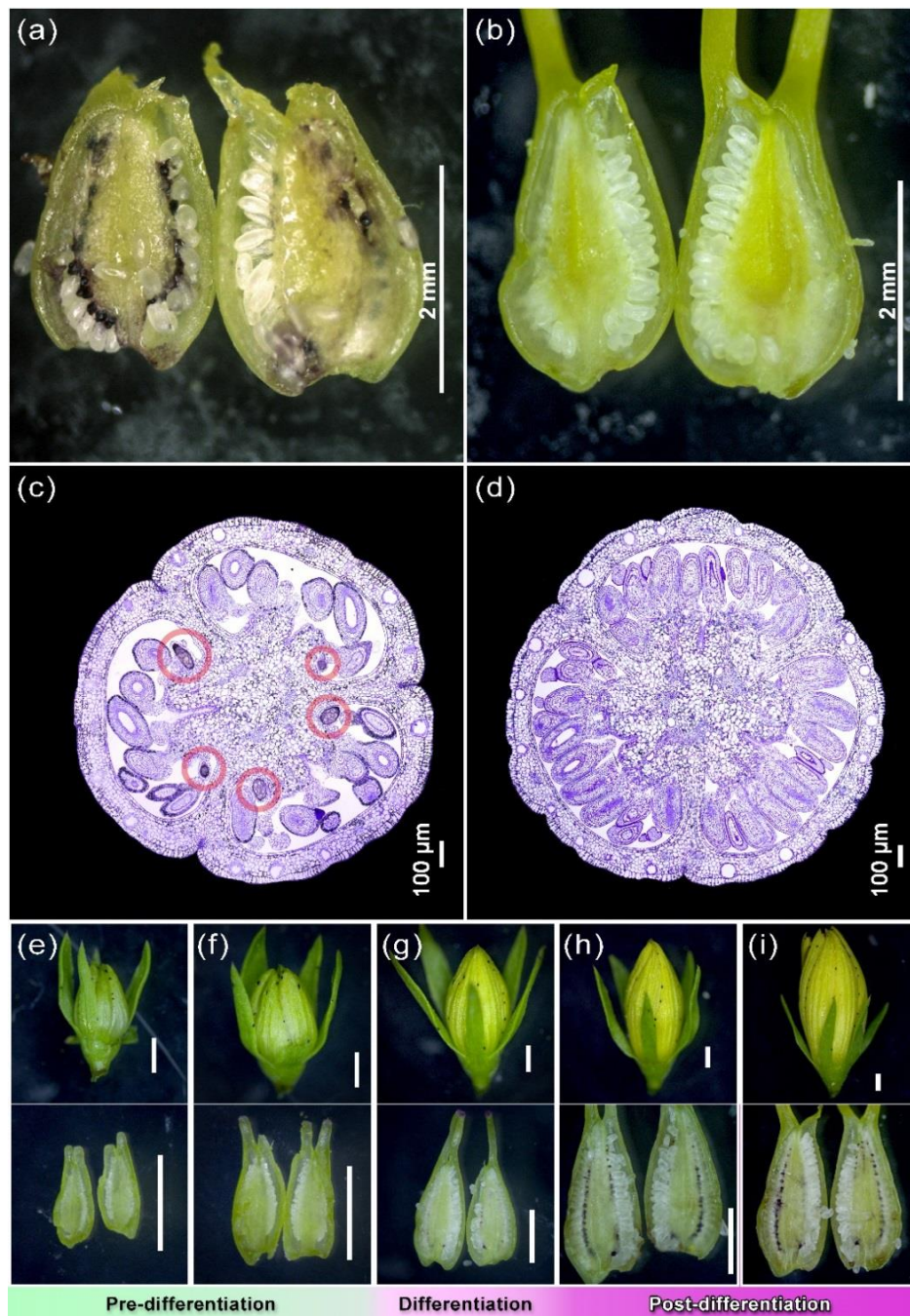


Figure 3.1 Sections of *Hypericum perforatum* pistils (a to d) and dissections of pistils from multiple stages of a G++ PT genotype (e to i). (a) Glanded placental tissue phenotype (G++ PT) from genotype HyPR-05; (b) glandless placental tissue phenotype (G- PT) from genotype H06-3251. Scale bars = 2 mm. (c,d) High-resolution comparison of two pistil transverse sections from open flower stage. G- PT (right): glandless placental tissue phenotype where no dark glands are present in any part of the placental tissue (or of the entire pistil section). G++ PT (left): glanded placental tissue phenotype; here 5 dark glands (highlighted in red) are visible on the surface of the placental tissue. Scale bars = 100 μm; upper panel (e–i): flower buds; lower panel (e–i): corresponding dissected pistils; (e–f) no dark glands in pistils, flower buds 2.82 mm and 3.23 mm long, respectively; (g) differentiation of dark glands in pistils, flower bud 5.17 mm; (h–i) developed dark glands in pistils, flower buds 6.76 mm and 9.22 mm, respectively. Scale bars in (e–i) = 1 mm.

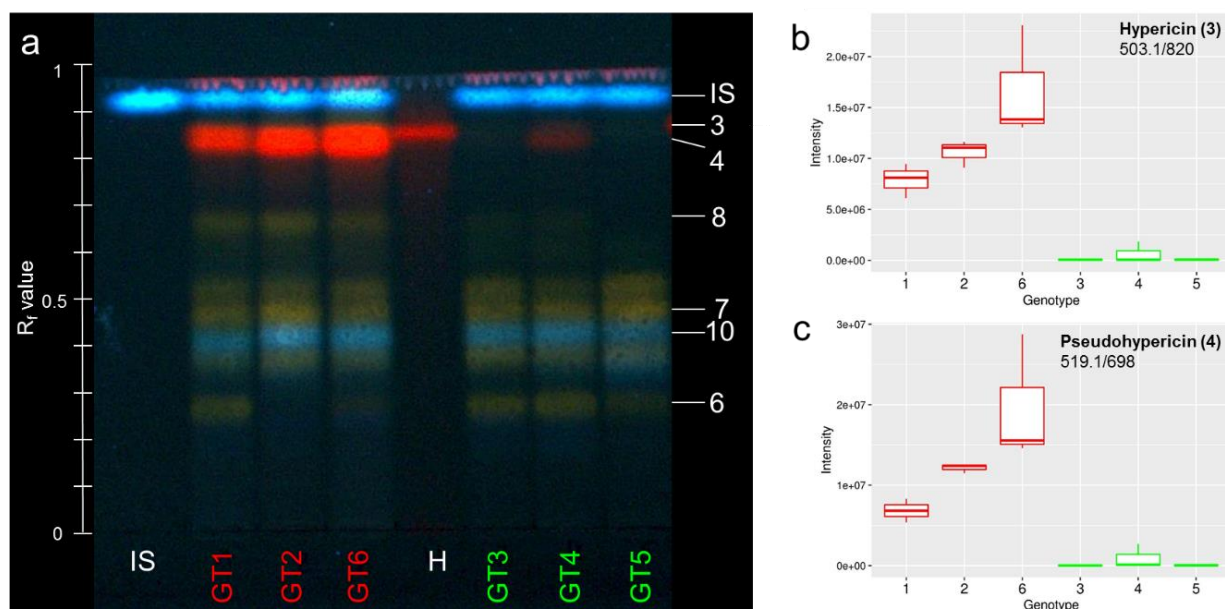


Figure 3.2 Analytical detection of hypericin (**3**) and pseudohypericin (**4**) in G++ PT (red: GT1 = H06-1988, GT2 = HyPR-03, GT6 = HyPR-01) and G- PT genotypes (green: GT3 = H06-1489, GT4 = H06-1369, GT5 = H06-3220). (a) Thin-layer chromatogram including the internal standard umbelliferone (IS, 8 $\mu\text{g/mL}$) and hypericin (H, 1 mg/mL). (b) Boxplots of MS intensities: hypericin (**3**, feature 503.1/820) and pseudo-hypericin (**4**, feature 519.1/698). Each boxplot is based on three biological replicates each composed of 10 pistils of the same genotype.

In accordance with the TLC results, the amounts of hypericin (**3**) and the related naphthodianthrone pseudohypericin (**4**) as well as the precursor compounds penicilliopsin (**23**) or emodin dianthrone (**24**), hydroxyenicilliopsin (**25**) or hydroxyemodin dianthrone (**26**), protohypericin (**18**) and protopseudohypericin (**27**) are significantly higher in G++ PTs than in G- PTs (Figure 3.3b). These findings are in line with the detection of hypericin (**3**) and related phytochemicals in dark glands of *Hypericum* species with different MS imaging techniques.^{96,117} The elevated feature with m/z 313 and retention time 501 s was identified as endocrocin (**28**). This anthraquinone differs from emodin (**20**), the proposed monomeric precursor of hypericin, by a carboxyl group. Moreover, two glycosides of the dimeric anthraquinone skyrin (**29** and **30**) are equally increased in the G++ PT materials. Those were already observed by Wirz et al. (2000)¹¹⁸ and more recently by Kimáková et al. (2018)¹⁰¹ in *Hypericum* extracts. If interpreted as side products of hypericin synthesis, these compounds suggest that biosynthesis of hypericin proceeds preferably via initial formation of the C5-C5' bond between the naphthodianthrone halves (Figure 3.4) and not via the C10-C10' bond as postulated by Soták et al. (2016a)²⁷. Already Falk (1999)¹¹⁹ has considered the C5-C5' bond formation as an alternative to an initial C10-C10' bond. Since emodin dianthrone (**24**) and its hydroxylated derivative (**26**), which indicate a C10-C10' bond, possess the same molecular formulas as penicilliopsin and hydroxyenicilliopsin with a C5-C5' bond, respectively, we cannot differentiate between these substances in the absence of reference compounds defining the retention time on the UHPLC C18 column. Currently, we are also not able to define whether emodin anthrone (**21**), emodin (**20**), and atrochryson (**31**) (as proposed by Gill and Gimenez (1991)²⁶ for the synthesis of austrovenetin) are the substrates for this reaction, since none of these potential intermediates could so far be detected in our analyses.

Surprisingly, the flavonol quercetin (**5**), some quercetin glycosides (**7/9** and **8**) and phenolic acids such as *O*-*p*-coumaroylquinic acid are also more abundant in glanded tissues. A correlation of the amounts of quercetin derivatives (**5**, **8**) with those of hypericin (**3**) and pseudohypericin (**4**) was already described by Kusari et al. (2009)⁷⁷ and will be discussed in detail in chapter 4.3.5. Further, different imaging techniques could show the location of quercetin derivatives at dark glands.^{96,120}

To identify additional differentiating features, G- PT samples were checked for analogous concentration increases. Among the compounds identified this way was the phloroglucinol derivative ($[M-H]^- C_{20}H_{27}O_4^-$) and the quercetin diglycosid rutin (**6**) (Table 3.1, negatively correlating features). The other contributing features are not identified, but their chromatographic behavior suggests highly apolar compounds possibly related to phloroglucinols or terpenoids.

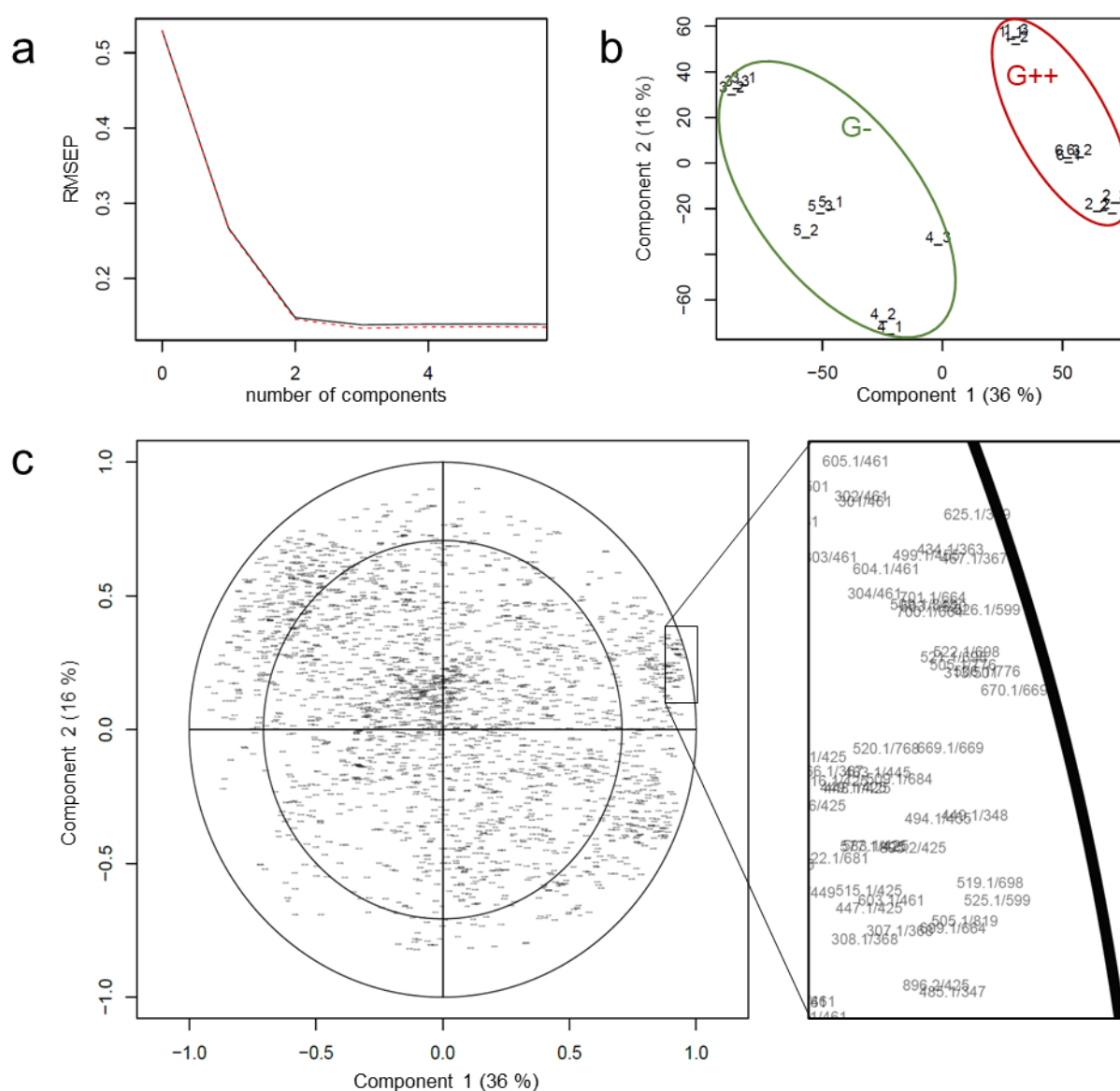


Figure 3.3 PLS analysis of glanded versus glandless *H. perforatum* pistils analyzed by UHPLC-ESI-HRMS in negative mode. **(a)**: root mean squared error of prediction curve (black line = cross validation estimate, red dotted line = bias corrected cross validation estimate); **(b)**: score plot (G++PT accessions (red): 1 = H06-1988, 2 = HyPR-03, 6 = HyPR-01; G- PT accessions (green): 3 = H06-1489, 4 = H06-1369, 5 = H06-3220); **(c)**: correlation loading plot with identified features in Table 3.1.

Table 3.1 Extract from the highest positive and negative correlating features determined by PLS analysis of UHPLC-ESI-HRMS metabolite profiles of glanded versus glandless *H. perforatum* pistils (excluding isotopes); sort by “Var1” the variance of the feature that is explained by component 1. Due to low concentration not all analytical parameters could be determined (-).

Var1	Feature	m/z [M-H] ⁻	Rt [s]	Molecular Formula [M-H] ⁻	Δ ppm	RDB	Fragment ions (intensity)	abs max [nm]	Assignment
0.941	525.1/599	525.1185	599	C ₃₀ H ₂₁ O ₉ ⁻	-1.077	20.5	-	-	Hydroxyemodin dianthrone (26) or Hydroxyphenicilliosin (25)
0.939	519.1/698	519.0711	698	C ₃₀ H ₁₅ O ₉ ⁻	-2.052	23.5	519.0716 (100) 487.0454 (11) 503.0403 (5)	283 331 366	Pseudohypericin (4)
0.933	449.1/348	449.1081	348	C ₂₁ H ₂₁ O ₁₁ ⁻	-1.792	11.5	241.0503 (20) 285.0400 (100) 323.0768 (7)	-	n.a
0.931	313.0/501	313.0348	501	C ₁₆ H ₉ O ₇ ⁻	-1.808	12.5	269.0463 (100)	-	Endocrocin (28)
0.928	505.1/776	505.0929	776	C ₃₀ H ₁₇ O ₈ ⁻	0.018	22.5	-	-	Protohypericin (18)
0.924	699.1/664	699.1354	664	C ₃₆ H ₂₇ O ₁₅ ⁻	-0.233	23.5	519.0712 (100)	285 333	Skyrin glucoside (27)
0.923	669.1/669	669.1250	669	C ₃₅ H ₂₅ O ₁₄ ⁻	0.017	23.5	519.0729 (100)	285 333	Skyrin-xylopyranoside /riboside (28)
0.917	503.1/820	503.0768	820	C ₃₀ H ₁₅ O ₈ ⁻	-0.896	23.5	503.0767 (100) 459.0870 (8)	-	Hypericin (3)
0.914	499.1/455	499.0849	455	C ₃₁ H ₁₅ O ₇ ⁻	5.097	24.5	-	-	n.a.
0.910	532.1/426	531.0540	426	C ₂₇ H ₁₅ O ₁₂ ⁻	-5.403	20.5	-	-	n.a.
0.905	337.1/191	337.0923	191	C ₁₆ H ₁₇ O ₈ ⁻	-1.663	8.5	119.0503 (5) 163.0400 (100) 173.0455 (5) 191.0559 (7)	288 309	O- <i>p</i> -Coumaroylquinic acid
0.904	509.1/684	509.1240	684	C ₃₀ H ₂₁ O ₈ ⁻	-0.336	20.5	-	-	Emodin dianthrone (24) or Penicilliosin (23)
0.897	337.1/172	337.0921	172	C ₁₆ H ₁₇ O ₈ ⁻	-2.197	8.5	119.0503 (5) 163.0400 (100) 173.0455 (5) 191.0559 (6)	288 309	O- <i>p</i> -Coumaroylquinic acid
0.895	463.1/445	463.0877	445	C ₂₁ H ₁₉ O ₁₂ ⁻	-1.035	12.5	301.0348 (100)	368	Quercetin-glycosid (7/9)
0.892	447.1/425	447.0922	425	C ₂₁ H ₁₉ O ₁₁ ⁻	-2.471	12.5	301.0350 (100) 284.0402 (3)	266 350 368	Quercetin-3- <i>O</i> -rhamnosid (8)
0.891	301.0/461	301.0353	461	C ₁₅ H ₉ O ₇ ⁻	-0.119	11.5	151.0037 (83) 178.9985 (100) 193.0141 (5) 257.0454 (11) 273.0403 (13)	368 271	Quercetin (5)
-0.925	331.2/727	331.1911	727	C ₂₀ H ₂₇ O ₄ ⁻	-1.306	7.5	259.0974 (5) 219.0665 (7) 194.0591 (89) 287.2014 (100)	-	Phloroglucinol derivative
-0.780	609.1/399	609.1455	399	C ₂₇ H ₂₉ O ₁₅ ⁻	-1.014	13.5	301.0350 (100) 271.0248 (5) 343.0457 (7)	254 365	Rutin (6)

3.3.3 Identification of genes involved in the hypericin biosynthesis

One main aim of this study was to identify putative regulators of dark gland differentiation and candidate genes involved in hypericin biosynthesis. The results of the transcriptomic analysis are discussed in detail in the original publication.⁶⁷ This chapter gives a short overview of strong upregulated genes at the developmental stage FB45 of G++ placental tissues, that take part in the secondary metabolism.

With respect to the biosynthesis of flavonoids, it is noteworthy that a 2-oxoglutarate, Fe(II)-dependent oxygenase (2-ODD)¹²¹ and a cytochrome P450 enzyme (CYP)¹²², which are typical for hydroxylation and dehydration reactions in that pathway, are encoded by differentially expressed genes (DEGs). Further, an UDP-glucosyltransferase (UGT) for the transfer of sugar moieties to flavonols was induced,¹²³ leading to metabolites such as the observed quercetin glycosides (**7**, **8**, **9**). For translocation of the cytosolically synthesized flavonoids across membranes and transport to the vacuole, mechanisms including glutathione S-transferases (GST)¹²⁴⁻¹²⁶ and ABC transporters (ABC)¹²⁷ are discussed.¹²⁸ Genes encoding both types of proteins are within the list of highly induced DEGs. In addition, a β -glucosidase targeted for secretion (BGLU)¹²⁹ and a membrane integral protein of the major facilitator superfamily (MFP)¹³⁰ are observed, which suggest cleavage of sugar moieties from vacuolar or secreted compounds or cell wall components and sugar import (MFP). Since no further highly induced DEGs encoding transport proteins are observed, naphthodianthrone transport might be similar to flavonoid transport and may involve GST-associated vesicles. This would confirm the observations by Onelli et al. (2002) who extensively reported the accumulation of vesicles in and around the developing dark glands of leaves.¹⁰⁴

With respect to naphthodianthrone synthesis (Figure 3.4), we observed several DEGs encoding the well-described octaketide synthase HpPKS2 (OKS), executing the first committed step of hypericin synthesis.¹⁰² Remarkable is the highly induced potential polyketide cyclase (PKC), which might encode a missing function for correct cyclization of the octaketide proposed by Karppinen et al. (2008).¹⁰² Two DEGs annotated as dihydrofolate reductases belonging to the superfamily of serine hydrolases may encode thioesterases (TER) involved in releasing octaketide and tetraketide precursors of hypericin (**3**) and flavonoids from their coenzyme A conjugates.¹²⁹ In addition, we detected DEGs encoding phenoloxidative coupling proteins (POCPs) discovered by Soták et al. (2016a)²⁷ to be induced in dark gland-containing tissues and suggested to be involved in C-C bond formation between the naphthodianthrone halves of hypericin (**3**) and its derivatives. To our knowledge, no experimental proof of the functionality of the PR-10-related POCPs exists, but a PR-10-related protein has been shown to exhibit norcoclaurine synthase activity in opium poppy by performing a condensation reaction between dopamine and 4-hydroxyphenyl-acetaldehyde in benzyloquinoline alkaloid synthesis.¹³¹ Besides the POCP genes, we identified a berberine bridge enzyme (BBE)¹³², which suggests an additional or alternative function to POCPs for the formation of one of the three C-C linkages between the

naphthodianthrone halves of hypericin (**3**). Since the encoded BBE is targeted for secretion by a N-terminal signal peptide,¹³³ it will most likely close the last C-C bond for which an enzyme is required. It could be similar to THCS and CBDS in tetraketide-based cannabinoid synthesis,^{134,135} which belong to the same protein family and are predicted to contain N-terminal signal peptides as well. Our observation that skyrin glycosides (**27**, **28**) are correlated with the presence of dark glands suggests that the C5-C5' bond is formed first. This view is supported by the previous detection of skyrin glycosides in *H. perforatum*¹¹⁸ and the recent notion that their presence is correlated with occurrence of hypericin (**3**) in *Hypericum* species.¹⁰¹ As mentioned above, initial C5-C5' bond formation has been suggested earlier as an alternative to C10-C10' bond formation.¹¹⁹ Therefore, POCPs and the berberine bridge enzyme might be involved in the formation of the C5-C5' bond and the C10-C10' double bond, respectively, while the C4-C4' bond might be formed in a last step nonenzymatically. Since Kusari et al. (2015)⁹⁶ and Kimáková et al. (2018)¹⁰¹ noted that the presence of emodin anthrone (**21**) and emodin (**20**) does not correlate with the occurrence of hypericin in *Hypericum* species, the biosynthesis of hypericin might proceed via atrochryson (**31**), atrovirin B, austrovenetin and penicilliopsin (**23**) (Figure 3.4; compare Gill and Gimenez (1991)²⁶) completely circumventing emodin (**20**) and emodin anthrone (**21**), but retaining the potential to yield skyrin.[^]

3.4 Conclusion

H. perforatum promises important medical applications some of which maybe related to hypericin. This lead the scientific community to focus on the biosynthesis of hypericin. In contrast to other studies, the development of dark glands was observed in placental tissue. This enabled the identification of candidate genes considered responsible for dark gland (DG) differentiation. The varying occurrence of dark glands in the pistil tissue of *H. perforatum* GTs was used for a comparable metabolite analysis. The content of hypericin, pseudohypericin, endocrocin, skyrin glycosides, and several flavonoids correlated with the occurrence of dark glands. The combination of the metabolite profiling with gene expression data and DG development was used to generate a revised model for hypericin biosynthesis. Based on the transcriptomic data, genetic engineering of the identified candidate genes might lead to complete shutdown or alternatively, hyper-activation of dark glands. Such artificial lines have the potential to provide new crude extracts to be tested both in Alzheimer and cancer research where the comparison of hypericin-free versus high hypericin extracts is already in high demand. Furthermore, the candidates shown in this study, once validated, will be very valuable parents for future breeding programs of *H. perforatum*.

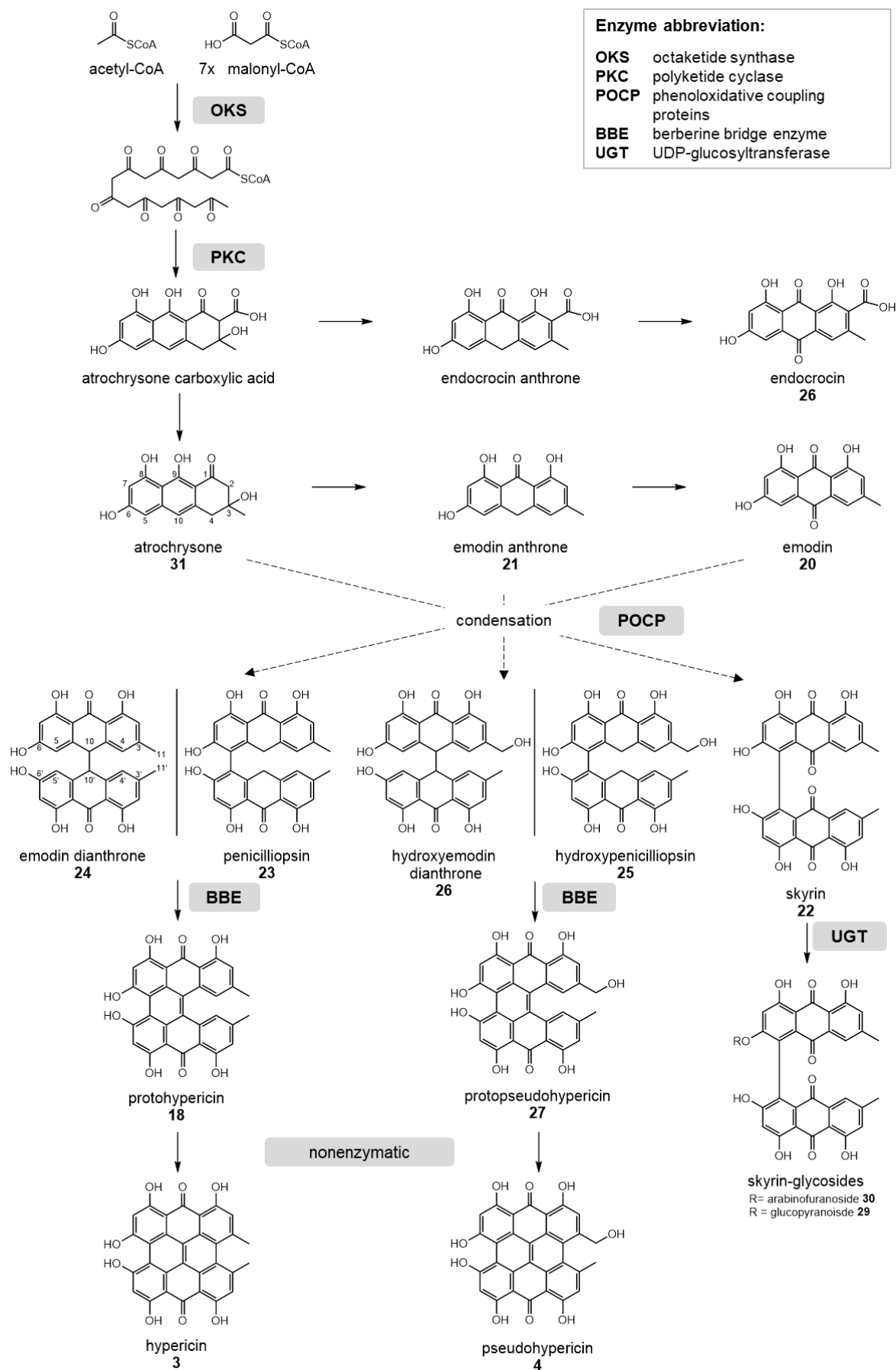


Figure 3.4 Proposed hypericin biosynthetic pathway based on our metabolomics, transcriptomics and functional annotation, adapted from Rizzo et al. (2019) ⁶⁷.

4 Intraspecific variance of *H. perforatum* genotypes – a metabolomics approach

Abstract

In this study, comparative metabolite profiles of 93 *H. perforatum* accessions with different genetic backgrounds were investigated to estimate the intraspecific diversity. The genotypes were cultivated under identical conditions and harvested during the flowering period to minimize the impact of environmental factors. Flowers and leaves were sampled separately and were subsequently chemically characterized by UHPLC-ESI-HRMS. The untargeted experiment enables the simultaneous detection of hundreds of metabolites. Using multivariate data analysis, the chemical diversity of the extracts was evaluated. The genotypes differ especially in the content of the phloroglucinol hyperforin. One genotype was found not to contain this active compound and showed less prenylated hyperforin precursors. Additionally, the substitution pattern of quercetin derivatives varied widely between all genotypes. The examination of the required marker compounds of *Hyperici herba* revealed that more than 20% of the extracts did not fulfill the European pharmacopeia specification.

This untargeted metabolomics concept combined with multivariate data analysis enables the investigation of the natural product diversity within the species, elucidation of biosynthetic pathways, and the selection of genotypes with a suitable compound composition according to the proposed pharmacological application of the economically important species.

4.1 Introduction

The genus *Hypericum* comprises around 450 species, of which *Hypericum perforatum*, commonly St. John's Wort, is the best known. Extracts of the aerial parts of this herb are prescribed to treat mild to moderate depressions, but recently also anti-microbial, anti-inflammatory, and anti-Alzheimer effects were reported. For these reasons it is among the ten best-selling herbal products in Europe and the USA. The medicinal effects of the extracts highly depend on the chemical composition, which is genetically predetermined and strongly influenced by the environment. Major active constituents are phloroglucinols (hyperforin), naphthodianthrone (hypericin), and flavonoids. The economic importance and the chemical diversity of the secondary metabolites make *H. perforatum* an interesting target of phytochemical investigations.

4.2 Experimental

4.2.1 Plant material

The 93 wild accessions of *Hypericum perforatum* L. used in this study cover the complete range of ploidy, mode of reproduction, and genetic backgrounds as determined by Koch et al. (2013) and Molins et al. (2014) (Appendix 3.1).^{64,107} Plants were cultivated in pots on three fields in the IPK Gatersleben. Each genotype was represented on each field with up to twelve plants, originated from cuttings (Figure 4.1). Flowers and leaves were harvested separately. Flowers of early flowering accessions were taken in July 2018 (Appendix 3.1). Flowers of the remaining accessions were harvested parallel to leaves during the flowering stage in August. All samples were collected into 25 ml Eppendorf tubes and immediately frozen in liquid nitrogen. A second sample set was collected in parallel for genetic analysis. The sampling of each field took one day, and all fields were sampled on three consecutive days with comparable climate conditions. In order to minimize the influence of the time of day, the harvesting sequence was changed between days. A detailed list of the accession numbers and the sampling information are listed in the Appendix 3.1.



Figure 4.1 Picture of the sampling site (left) and a simplified representation of the experimental set up (right).

4.2.2 Extraction and Sample preparation

In total, 346 leaf samples and 291 flower samples were processed. The frozen plant material was lyophilized (72 h) until dryness, and flower samples were weighted to calculate the average flower weight. Afterwards, all samples were powdered using a Cryogenic Plant Grinder (Labman). Therefore, each sample was transferred to a 20 ml grinder polyvial (Zinsser Analytic) containing one stainless steel ball (5 mm). With -75 °C and 5% humidity, the grinding was performed twice for 15 s with 30 Hz. A break of 30 s in between grindings avoided warming of the samples. For each organ separately, a quality control (QC) sample was prepared, which is composed of equal parts of all individual samples. These pooled samples are considered as data set representatives for quality control and method development. Each sample was extracted with methanol (LC-MS grade, Chromasolv, Honeywell) containing 8 µg ml⁻¹ umbelliferone as an internal standard. For extraction, the material/solvent ratio was adjusted to the final concentration of 2 mg ml⁻¹. After 30 s mixing with vortex an ultrasonic extraction for 15 min took place. After 10 min centrifugation (16.000 rpm), the supernatant was diluted with the extraction solvent to reach the concentration of 1 mg ml⁻¹ for the leaf samples and 0.5 mg ml⁻¹ for flower samples. The samples were immediately frozen at -20 °C for at least 12 h and maximum 36 h.

4.2.3 UHPLC-ESI-MS/PDA analysis

Negative ion high-resolution ESI mass spectra were obtained from an Orbitrap Elite mass spectrometer (ThermoFisher Scientific, Bremen, Germany) equipped with a HESI electrospray ion source (spray voltage 3.5 kV; source heater temperature: 350 °C; capillary temperature 300 °C; FTMS resolution 15.000). Nitrogen was used as the sheath and auxiliary gas. The MS system was coupled with an ultra-high-performance liquid chromatography (UHPLC) system (Dionex UltiMate 3000, Thermo Scientific), fitted with a RP-C18 column (1,9 µm; 50 x 2,1 mm; Hypersil GOLD; Thermo Scientific; column temperature: 40 °C), and a photodiode array detector (PDA, Thermo Scientific; 220–650 nm). For UHPLC separation a three solvent gradient system with A: 1 mM ammonium acetate in water (pH 5), B: MeOH with 1% 1 mM ammonium acetate and C: MeOH at a flow rate of 500 µl min⁻¹ was applied (95:5:0 for 1 min, 6 min gradient to 0:100:0, 0.5 min to 0:1:99, isocratic hold for 3.5 min and returning to 95:5:0 in 1 min, isocratic hold for 4 min). For the LC-MS measurements, a complete randomized design was applied for each plant organ. The leaf samples were divided into five batches, and the flower samples in four. For each organ, a pooled sample was prepared containing all samples in equal parts. Each batch started with 4 injections of the QC sample to equilibrate the column. After seven study samples the QC was injected every 8th samples (2µl injection volume). In addition, an external standard mix (MM8) was injected every 16th sample (1µl injection volume). The data were evaluated with the Xcalibur software 2.2 (Thermo Fisher).

4.2.4 UHPLC-MS data processing

The raw spectra of samples and QCs were converted to mzML format using ProteoWizard (3.0.11110) to get centroided data. Further processing was done in R (version 3.5.1) using xcms (version 3.2.0)⁶⁹. Peak picking based on the CentWave Algorithm was performed using the following parameters: findChromPeaks(ppm=15, peakwidth=c(7,25), sntresh=10, noise=10000, prefilter=c(3,1000)). After that, the peaks were grouped over all samples with groupChromPeaks(minFraction=1, bw=5). The resulting feature lists of leaf and flower samples comprise 1651 and 1396 features, respectively.

Batch correction was applied to overcome the high variance introduced by the measurement procedure. Different methods for batch correction available in the R package BatchCorrMetabolomics (0.1.14) were tested, and the result validated with the parameters “duplo” and “PCA” (Bhattacharyya distance) introduced by Wehrens et al. (2016).¹³⁶ The best result was received by applying a linear model regression on the QC samples.

4.2.5 Isolation procedure

The *H. perforatum* plant material is described in chapter 5.2.6. The frozen crushed leaf material (90.2 g) was sonicated with MeOH and further extracted under shaking at room temperature with 95% aqueous MeOH overnight. After drying in *vacuo*, the extract (9.34 g) was partitioned via liquid-liquid extraction between CHCl₃ and water. The green chloroform phase (2.12 g) was applied to silica column chromatography. Gradient elution with MeOH:CHCl₃ from 5:95 to 100:0 yielded 17 fractions, based on TLC (MeOH:CHCl₃, 1:4.9 v/v) and UHPLC-MS profiling. Fraction 16 (5.1 mg) was applied to Sephadex LH20 column chromatography (MeOH) resulting in the isolation of compound **30** (5.3 mg). Fraction 17 (176.9 mg) was applied to Diaion column chromatography. The column was successively flushed with water, MeOH, acetone, and acidified acetone (2N HCl). The MeOH and acetone fractions were combined and purified with a Sephadex LH20 column chromatography (MeOH) yielding 0.7 mg of **29**.

S-Skyrin-6-O-β-glucopyranoside (**29**): orange solid; negative ion ESI-HRMS *m/z* 699.1350 [M-H]⁻ (calcd for C₃₆H₂₇O₁₅⁻, 699.1355); MS² see Table 5.1; CD (MeOH) λ_{max} (Δε): 310 (-19055), 298 (-11547), 290 (-15443), 262.5 (104488), 248 (-43581), 237.5 (-15454), 215.5 (-49987) nm, UV-Vis (MeOH) λ_{max} 221 (5.60), 258 (5.50), 295 (5.25), 451 (5.07) nm; ¹H-NMR (400 MHz, methanol-*d*₄) δ 7.33 (*d*, *J*=1.7 Hz, 1H, H-4), δ 7.22 (*d*, *J*=1.7 Hz 1H, H-4'), δ 7.08 (*s*, 1H, H-7), δ 7.06 (*d*, *J*=1.5 Hz, 1H, H-2), δ 7.00 (*d*, *J*=1.5 Hz 1H, H-2'), δ 6.45 (*s*, 1H, H-7'), δ 5.02 (*d*, *J*=7.6 Hz, 1H, H-1''), δ 3.81(*dd*, *J*=2.3, 12.2 Hz, 1H, H-6''a), δ 3.61 (*dd*, *J*=5.9, 12.4 Hz, 1H, H-6''b), δ 3.46 (*m*, 1H, H-5''), δ 3.36 (*dd*, *J*=9.2 Hz, 1H, H-3''), δ 3.20 (*dd*, *J*=9.4 Hz, 1H, H-4''), δ 3.09 (*dd*, *J*=7.6, 9.2 Hz, 1H, H-2''), δ 2.35 (*s*, 3H, Me-3), δ 2.32 (*s*, 3H, Me-3'); ¹³C-NMR (400 MHz, methanol-*d*₄, measured with HMBC) 183.7 (C-10), 167.5 (C-8'), 166.0 (C-8), 164.1 (C-6), 149.9 (C-3), 148.1 (C-3'), 127.8 (C-5), 124.4 (C-2, C-2') 121.8 (C-4), 121.0 (C-4'), 115.3 (C-9a'), 114.8 (C-9a), 112.8 (C-8a), 108.5 (C-8a').

S-Skyrin-6-O-β-arabinofuranoside (30): orange solid; negative ion ESI-HRMS m/z 669.1245 [M-H]⁻ (calcd for C₃₅H₂₅O₁₄⁻, 669.1250); MS² see table Table 5.1; CD (MeOH) λ_{\max} ($\Delta\epsilon$): 306.5 (-11551), 297.5 (-10697), 289 (-13304), 260.5 (67587), 247 (-20703), 236 (-1719), 217 (-25923) nm, UV-Vis (MeOH) λ_{\max} 221 (5.60), 258 (5.50), 295 (5.25), 451 (5.07) nm; ¹H-NMR (400 MHz, methanol-*d*₄) δ 7.35 (*d*, $J=1.6$ Hz, 1H, H-4), δ 7.30 (*d*, $J=1.7$ Hz, 1H, H-4'), δ 7.07 (*s*, 1H, H-7), δ 7.08 (*s*, 1H, H-2), δ 7.00 (*s*, 1H, H-2'), δ 6.64 (*s*, 1H, H-7'), δ 5.57 (*s*, 1H, H-1''), δ 3.51 (*dd*, $J=4.4$ 1H, H-5''a), δ 3.60 (*dd*, $J=3.6$ Hz, 1H, H-5''b), δ 3.75 (*m*, 1H, H-3''), δ 3.65 (*m*, 1H, H-4''), δ 3.78 (*dd*, $J=1.3, 3.3$ Hz, 1H, H-2''), δ 2.37 (*s*, 3H, Me-3), δ 2.35 (*s*, 3H, Me-3'); ¹³C-NMR (400 MHz, methanol-*d*₄, measured with HMBC and HSQC) 183.8 (C-10), 166.2 (C-8'), 166.2 (C-8), 163.4 (C-6), 149.9 (C-3), 149.4 (C-3'), 124.5 (C-2'), 124.7 (C-2), 121.7 (C-4), 121.5 (C-4'), 111.9 (C-8a), 22.1 (Me-3), 22.0 (Me-3'), 107.4 (C-1''), 83.5 (C-2''), 78.1 (C-3''), 87.2 (C-4''), 62.4 (C-5'').

4.2.6 Chemometrics

Different methods of multivariate data analysis were utilized in R. The analysis were performed on base of the generated feature tables (chapter 4.2.4). Packages that were used are: *pcaMethods* (1.78.0)⁷⁰, *Heatmaply* (1.0.0)¹³⁷, *NetworkD3* (0.4), and *Corrplot* (0.84)¹³⁸.

4.3 Results and discussion

4.3.1 Experiment design and method development

4.3.1.1 Experimental setup and prestudies

After the promising study shown in chapter 3, an advanced experiment adapted to the requirements of metabolomic investigations was performed. In order to increase the experimental significance, all available and cultivable *H. perforatum* accessions of the IPK Gatersleben were included (appendix 3.1). The total number of 93 lines and a sufficient number of biological replicates made it necessary to conduct the experiment with plants field-grown on three plots (Figure 4.1). Since *H. perforatum* is also cultivated in the field for the production of pharmaceutical products, this corresponds to authentic conditions. Nevertheless, the environmental influences are identical for all GTs, and therefore comparable.

An enormous advantage of this study is the presence of biological replicates because the differences within a GT can be compared to the intraspecific variance. The plants were cultivated in pots to avoid positive and negative plant-plant interactions by roots, which can have a strong influence on the composition of secondary metabolites. The three fields were located side by side, with plot 1 being shadier due to nearby trees than plots 2 and 3, which were sunny all day (Figure 4.1).

In contrast to the previous experiment (chapter 2), leaves and flowers were sampled separately, and not the complete aerial parts were used. Due to the absence of stems and smaller sample amounts, a cryogenic grinder was employed. Thus, the grinding process could be accelerated, standardized, and more homogeneous samples were obtained. The separate investigation of flowers allows the first comparison of different GTs based on this plant organ, which usually contains higher contents of the valuable ingredients.¹³⁹ In parallel to the metabolomics study, samples for the analysis of genome and transcriptome were taken that will be evaluated at the IPK Gatersleben. This comprehensive approach allows further knowledge of the genetic variance, which causes the metabolite variation.

4.3.1.2 Extraction and UHPLC-ESI-HRMS method optimization

The preliminary investigation in chapter 2 concentrated on the comparison of extracts produced according to pharmaceutical procedures. However, this production is extremely time-consuming due to the included maceration and drying steps. Furthermore, each working step favors technical errors and chemical changes of the extract due to the instability of some compounds. Hyperforin (**1**), in particular, is only partially stable so that oxidation products are formed quickly.¹⁴⁰⁻¹⁴² Under light exposure, the proto-forms of naphthodianthrones are converted to the oxidized hypericin (**3**) and pseudohypericin (**4**).¹⁴³ To counteract these processes, in the present study a direct ultrasonic bath extraction with 100% methanol without light exposure was used. This enables a fast and standardized extraction without error-prone working steps. To treat all samples equally, the freshly prepared extracts were stored in the freezer before measurement. According to preliminary tests (appendix A.4.1), the frozen samples were more comparable with each other over the measuring time in the autosampler than freshly prepared samples.

Furthermore, the UHPLC method was adapted for high sample throughput. In particular, the pH sensitivity of naphthodianthrones led to retention time shifts within a batch, which made automatic evaluation difficult. Therefore ammonium acetate buffer (pH 5) was used as a modifier instead of formic acid in the mobile phase, which facilitates the comparability between different batches. This buffer system was already used in higher concentrations by Smelcerovic and Spiteller (2006)¹⁴⁴, and Farag and Wessjohann (2012)⁵⁷.

The final measuring concentrations of leaf and flower samples were determined separately by investigating the linear dynamic range (LDR). Due to the buffer usage, the ionization potential of phloroglucinols raised, compared to the previous experiment (chapter 2), resulting in a reduction of the LDR. Therefore, lower concentrations had to be applied. These were finally adjusted to 1 mg material per ml methanol for leaf samples and, due to the very high hyperforin (**1**) content, to 0.5 mg ml⁻¹ for flowers.

The measuring time should generally be as short as possible to reduce the average stay of the samples in the autosampler. In the final experiment, the runtime, including the equilibration time, was shortened to 15 min. To maintain a good separation, the flow rate was increased to

0.5 ml min⁻¹ so that the column had an optimized separation performance at a pressure of 300–700 bar.

The mass accuracy of the Orbitrap mass spectrometer decreases continuously with increasing sample numbers, and, from experience, it was no longer within the acceptable range of +/- 3 ppm after about 120 runs in one uninterrupted sequence (batch). Since the number of samples significantly exceeded the number of possible runs within a batch, the samples were randomly measured in batches (leaf samples five batches, flower samples four batches) with maximal 100 runs each. In between two batches, the sprayer and the transfer capillary were cleaned, and recalibration was carried out.

4.3.1.3 UHPLC-MS data processing optimization and batch correction

The first step in processing LC-MS data is the correct creation of a feature-table, which shows peak intensities as entries for each sample. A feature represents the combination of a mass-to-charge ratio (m/z) with a retention time. To determine a feature correctly in all samples, the parameters for peak picking and grouping were optimized (chapter 4.2.4), whereby, in the beginning, all features with intensities above 1000 were considered.

The resulting PCA is shown in Figure 4.2a. By coloring each sample according to the batch number, it becomes clear that the measurement conditions explain the main variance. These batch-to-batch differences are a common phenomenon in extensive untargeted LC-MS metabolomics experiments. This so-called batch effect is caused by unavoidable external influences on the chromatography and MS instruments. To ensure the comparability of samples from different batches, a batch correction is required. Different methods were applied to validate the most appropriate batch correction (appendix 4.3). The success was evaluated by visual assessment of the PCA and by two criteria introduced by Wehrens et al. (2006), the Bhattacharyya distance and the parameter "duplo". "Duplo" describes the repeatability, and is calculated as proportion of the variance within the replicates of a sample, to the total variance.¹³⁶

The simplest correction relies on the normalization of the internal standard umbelliferone. But with this approach, the batch effect was not removable (Figure 4.2b) since the ionization behavior of the diverse compounds, present in the analyzed crude extract, vary considerably. Thus, only substances similar to the internal standard can be adequately corrected by this method, whereas compounds with very different properties experience an additional error.

A frequently described method for the correction of batch effects is based on the use of QC samples - a pooled sample that has been prepared in the same way as the study samples and is measured regularly within the sequence.^{136,145,146} Since aliquots of the QC have been sufficiently measured (every eighth sample) throughout the experiment, this correction can be applied using different regression methods.¹³⁶ In the resulting PCA (Figure 4.2c) replicates and the QC samples of different batches cluster together, which indicates a good correction.

Another tested correction method is independent of the QCs and exclusively based on the study samples. It is assumed that due to proper randomization, the samples show, on average, the same batch dependent effect. This seems to be an advantage because more metabolites can be corrected, even those, which cannot be detected in the QCs due to dilution effects. The result was comparable to the correction based on QCs (data not shown).

It turned out that there are a large number of non-detects in the used feature tables, which frequently cause problems for corrections. These features often show a low intensity, so that it depends strongly on the batch, whether they are detected at all. Therefore, the minimum intensity for the peak picking has been increased to 10000. This resulted in a feature reduction (from 2829 to 1651), and a good batch correction result was achieved after applying the QC correction so that this processing method was used for further data evaluation (Figure 4.2d).

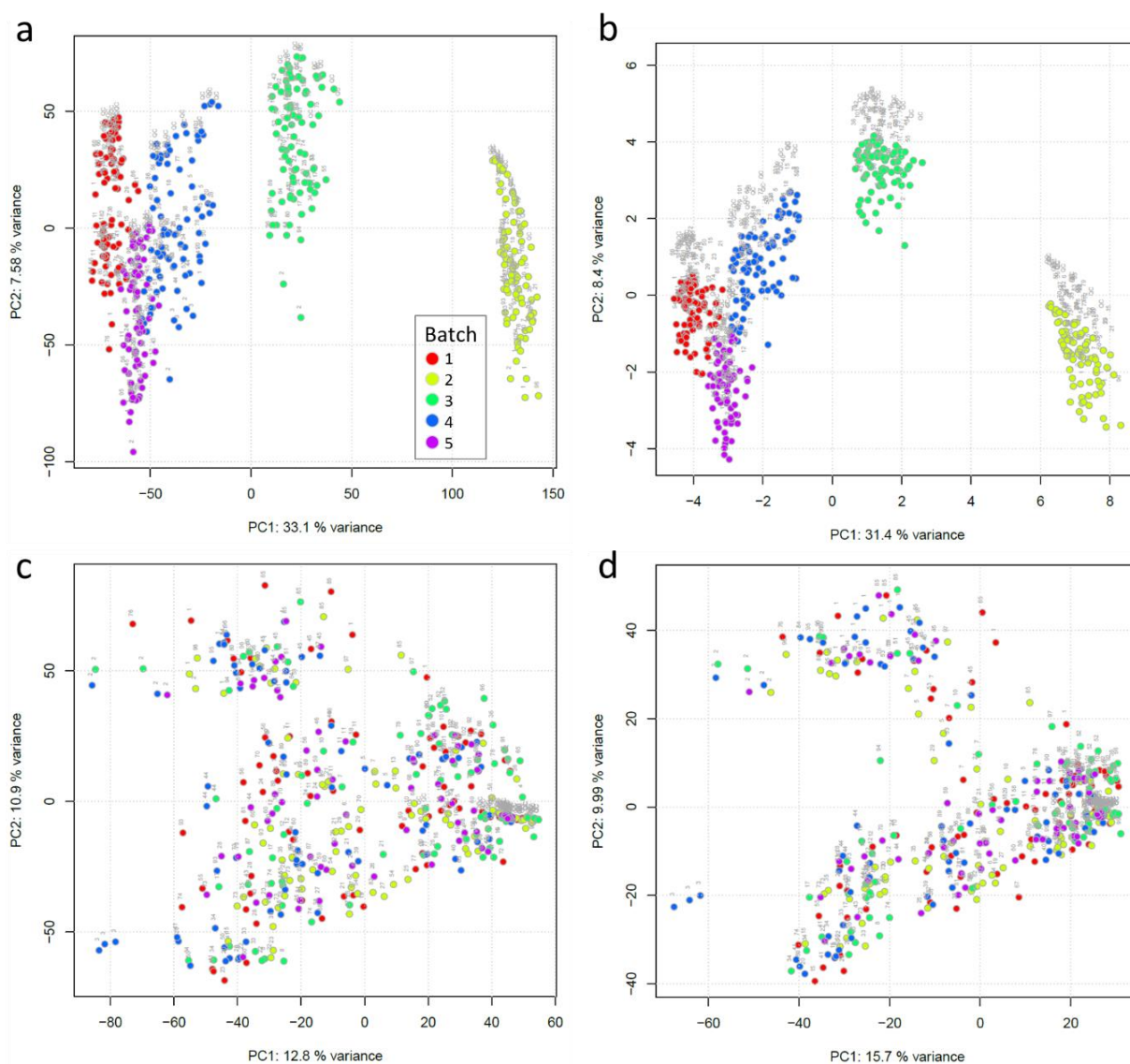


Figure 4.2 PCA colored by batch number of the logarithmized leaf dataset of *H. perforatum* genotypes, labeled with GT number (grey): a) without batch correction (noise < 1000) b) internal standard normalization (noise < 1000) c) QC correction (noise < 1000) d) QC correction (noise < 10000).

4.3.2 Characterization of the intraspecific variance of *H. perforatum* by multivariate data analysis

An experimental setup was created in which genotypes (GTs) with different origins were cultivated under the same conditions, to address the genetic variation of secondary metabolites within the species *H. perforatum* from Europe and North America. The chemical compositions of the methanolic extracts were analyzed with UHPLC-ESI-MS in an untargeted metabolomics experiment. The good sensitivity allows the detection of hundreds of constituents over a wide concentration range. The widely used Principal Component Analysis (PCA) projects the high-dimensional data in a two-dimensional space showing the highest variance in the data. This method gives a good overview of the overall variation in the GTs and helps to understand patterns.

Since the constituents of the two organs, flowers, and leaves, differ in the amounts of secondary metabolites,¹³⁹ the sample sets were measured with different concentrations (chapter 4.3.1.2). Therefore, the sample sets are considered separately for data evaluation. For the analysis of the leaves, only samples from the consecutive harvest days 3, 4, and 5 were used, because the weather conditions were comparable.

The PCA scores plots based on the leaf data are shown in Figure 4.3a,c. The first two PCs represent only 26.3% of the total variance, which proves that there are no samples fundamentally different from the others. However, this was to be expected as all samples belong to the same species. In PC1, a single genotype (GT3) is separated from the other GTs, while PC2 divides the GTs into two groups. The features responsible for the separation were investigated in the loadings plot (Figure 4.3b,d). Characteristic features and tentatively assigned compounds (numbered according to increasing retention time) are listed in Table 4.1. Mainly responsible for the separation are flavonoids, marked in green in the loadings plots. Especially astilbin (**19**/P15), rutin (**6**/P17), and acetylated quercetin glycosides (P12, P13, P23, P24) vary. GTs in the positive quadrants of PC1 possess an increased content of acetylated flavonoids. Group A is mainly characterized by increased astilbin (**19**) and decreased rutin (**6**) content, while it is vice versa in group B. Group B additionally contains increased amounts of the sulfonated xanthone P21 (marked blue), whose isolation is described in chapter 5. GT3 seems to contain increased amounts of two flavonoids (P22: C₂₂H₂₁O₁₂⁻, P26: C₂₂H₂₁O₁₁⁻) in particular. Based on the molecular formulas, it is assumed that these are simple methylated quercetin glucosides. Furthermore, prenylated phloroglucinols, marked in yellow in the loadings plots, exert a greater influence. Higher prenylated compounds (**1**/P49, **2**/P50, P57) are found in the positive range of PC1 as well as phloroglucinol **17** (P47) and the isomer P51, which suggests that GT3 contains less of these compounds. In contrast, phloroglucinols represented by compounds such as P35 (*m/z* 331), P41 (*m/z* 501), and P54 (*m/z* 413) are found in GTs shifted in the negative range of PC2.

PC3 and PC4 separate in addition to the GT3, two further GTs (21 and 29). Responsible for the partition towards PC4 is a reduced content of rutin (**6**) and prenylated phloroglucinols, in particular hyperforin (**1**), adhyperforin (**2**) and their degradation product furohyperforin (**32**/P57).

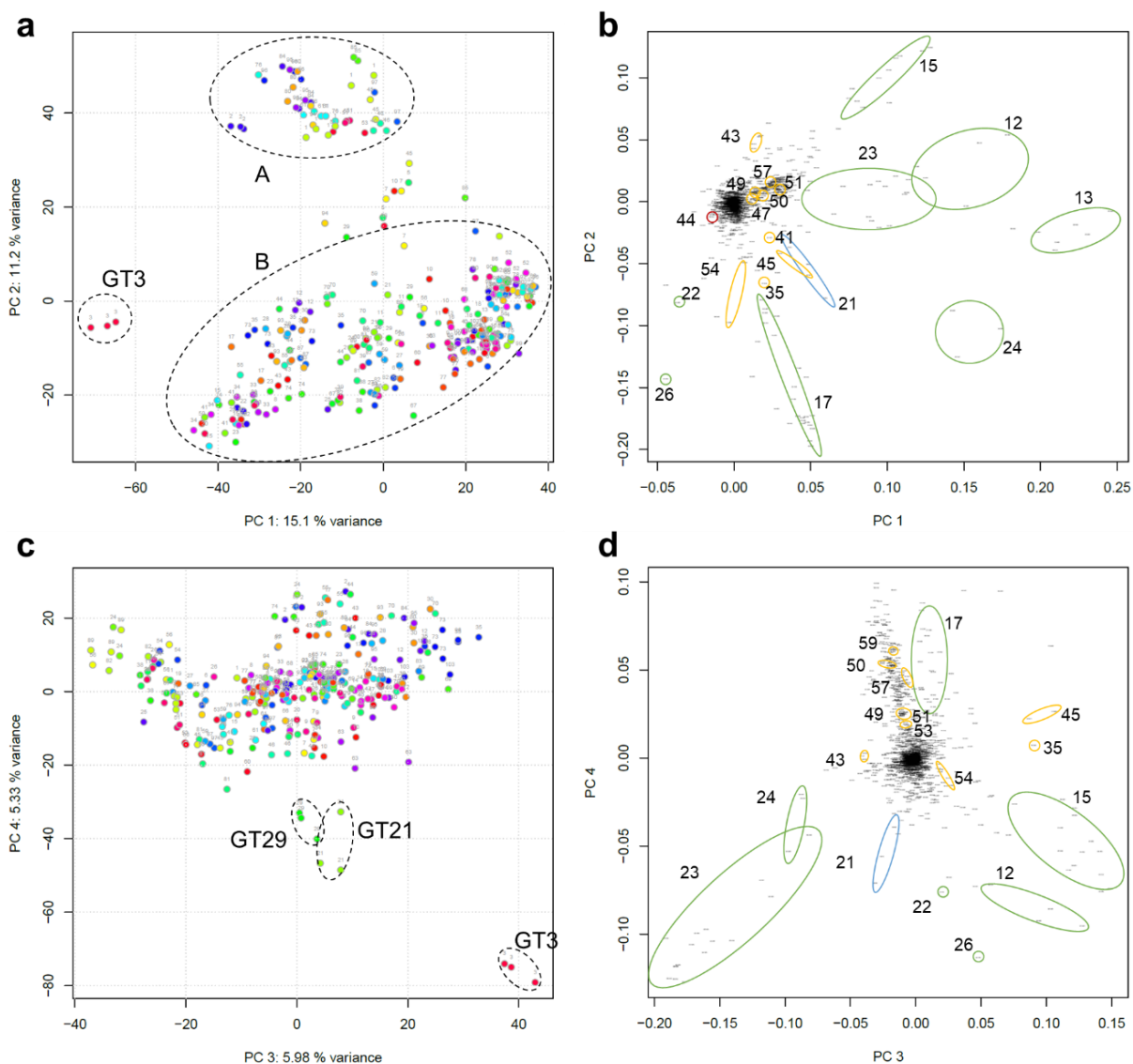


Figure 4.3 PCA of UHPLC-MS data of *H. perforatum* leaf samples: scores plots are coloured by GT and show a) PC1 and PC2 and c) PC3 and PC4. The corresponding loadings plots with assigned loadings (peak numbers according to Table 4.1) coloured by compound class (green = flavonoids, yellow = prenylated phloroglucinols, blue = xanthenes, red = naphthodianthrone) b) PC1 and PC2, and d) PC3 and PC4.

Table 4.1 Peaktable of major and/or discriminant compounds identified in *H. perforatum* genotypes for leaf (L) and flower (F) samples.

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P1	191.1/13 ^{L,F}	191.0557	0.2	C ₇ H ₁₁ O ₆	-2.310	173.0455 (16), 127.0401 (18), 111.0296 (100)	quinic acid
P2	353.1/38 ^{L,F}	353.0873	0.62	C ₁₆ H ₁₇ O ₉	-1.431	191.0560 (100), 179.0350 (37), 135.0453 (7)	chlorogenic acid (10) ^a
P3	337.1/63 ^L 337.1/63 ^F	337.0925	1.01	C ₁₆ H ₁₇ O ₈	-1.248	191.0560 (31), 173.0457 (5), 163.0403 (100), 119.0504 (6)	coumaroylquinic acid I
P4	289.1/165 ^L 289.1/159 ^F	289.0714	2.20	C ₁₅ H ₁₃ O ₆	-1.458	245.0820 (100), 205.0505 (34), 179.0351 (15)	epicatechin
P5	577.1/231 ^L 577.1/230 ^F	577.1343	3.84	C ₃₀ H ₂₅ O ₁₂	-1.350	451.1032 (29), 425.0876 (100), 407.0771 (54), 289.0718 (21),	procyanidin
P6	289.1/248 ^L 289.1/246 ^F	289.0715	4.09	C ₁₅ H ₁₃ O ₆	-1.008	245.0819 (100), 205.0507 (32), 179.0351 (14)	catechin (14)
P7	161/263 ^L 161/261 ^F	161.0245	4.36	C ₉ H ₅ O ₃	0.203	161.0205 (23), 133.0298 (100), 117.0345 (40)	umbelliferone (internal standard)
P8	421.2/269 ^L 421.1/261 ^F	421.0784	4.43	C ₁₉ H ₁₇ O ₁₁	1.770	403.0670 (13), 331.0456 (83), 301.0352 (100)	n.a.
P9	313/275 ^F	313.0345	4.57	C ₁₆ H ₉ O ₇	-2.862	269.0454 (100)	endocrocin (28) ^a
P10	399.1/279 ^L 399.1/278 ^F	399.1291	4.63	C ₁₈ H ₂₃ O ₁₀	-1.328	309.0977 (24), 287.0772 (61), 279.0873 (100), 167.0350 (47)	n.a.
P11	477.1/282 ^{L,F}	477.0670	4.67	C ₂₁ H ₁₇ O ₁₃	-0.972	301.0353 (100)	miquelianin ^a (quercetin-3-O-glucuronide)
P12	505.1/287 ^{L,F}	505.0982	4.76	C ₂₃ H ₂₁ O ₁₃	-1.195	463.0877(33), 445.0771 (11), 300.0272 (100), 301.0349 (64)	acetyl quercetin-3-O-hexoside I *
P13	651.2/291 ^{L,F}	651.1547	4.83	C ₂₉ H ₃₁ O ₁₇	-1.963	609.1457 (41), 591.1348 (100), 301.0353 (52), 300.0276 (91), 299.0196 (21), 271.0248 (18)	acetyl rutin I *
P14	549.1/296 ^{L,F}	549.0882	4.92	C ₂₄ H ₂₁ O ₁₅	-0.679	505.0984 (100)	n.a.
P15	449.1/300 ^{L,F}	449.1088	4.99	C ₂₁ H ₂₁ O ₁₁	-0.411	303.0506 (100), 285.0401 (82), 151.0038 (23), 323.0768 (12)	astilbin (taxifolin-3-O-rhamnoside) (19)
P16	463.1/302 ^{L,F}	463.0876	5.0	C ₂₁ H ₁₉ O ₁₂	-0.629	301.0350 (100), 343.0457 (4)	hyperoside (7) ^a (quercetin-3-O-galactoside)
		463.0876	5.02	C ₂₁ H ₁₉ O ₁₂	-0.629	301.0350 (100), 343.0457 (5)	isoquercitrin (9) ^a (quercetin-3-O-glucoside)
P17	609.1/305 ^L 609.1/304 ^F	609.1459	5.06	C ₂₇ H ₂₉ O ₁₆	-0.308	301.0348 (100), 343.0457 (7), 271.0246 (6), 255.0297 (3), 300.0279 (44)	rutin (6) ^a
P18	433.1/312 ^L 433.1/311 ^F	433.0768	5.18	C ₂₀ H ₁₇ O ₁₁	-1.488	343.0455 (3), 301.0353 (83), 300.0276 (100), 271.0613 (5)	avicularin (quercetin-3-O-arabinofuranoside)
P19	447.1/321 ^L 447.1/320 ^F	447.0929	5.33	C ₂₁ H ₁₉ O ₁₁	-0.929	301.0350 (100), 300.0281 (35)	quercitrin (8) (quercetin-3-O-rhamnoside)
P20	569.2/332 ^L 569.2/331 ^F	569.1863	5.48	C ₂₆ H ₃₃ O ₁₄	-2.194	509.1647 (63), 405.1178 (46), 363.1080 (19), 345.0974 (35), 243.0658 (100)	n.a.
P21	337/332 ^L 337/343 ^F	337.0019	5.51	C ₁₄ H ₉ O ₈ S	-1.339	321.9788 (100), 257.0454 (68)	1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid
P22	477.1/333 ^L	477.1036	5.55	C ₂₂ H ₂₁ O ₁₂	-0.585	357.0604 (8), 316.0543 (11.06%), 315.0506 (100), 314.0430 (49)	O-methylquercetin hexoside
P23	505.1/334 ^L 505.1/333 ^F	505.0984	5.55	C ₂₃ H ₂₁ O ₁₃	-0.8	463.0875 (30), 445.0769 (17), 300.0272 (100), 301.0349 (65)	acetyl quercetin-3-O-hexoside II *
P24	651/337 ^L	651.1551	5.60	C ₂₉ H ₃₁ O ₁₇	-2.430	609.1413 (43), 591.1331 (100), 301.0353 (55), 300.0276 (99), 271.0248 (8)	acetyl rutin II *

4. Intraspecific variance of *H. perforatum* genotypes – a metabolomics approach

No	Feature	[M-H]-	Rt (min)	Moecular Formula [M-H]-	Δ ppm	Fragments (Intensity)	Identification
P25	263.1/345 ^L 263.1/340 ^F	263.1289	5.63	C ₁₅ H ₁₉ O ₄	-0.769	194.0583 (100), 166.0636 (13), 151.0039 (3)	dimethylallyl-phlorisobutyrophenone (15) ^a
P26	461.1/344 ^L	461.1079	5.67	C ₂₂ H ₂₁ O ₁₁	-2.179	341.0654 (14), 299.0548 (46), 298.0475 (100), 269.0449 (8)	n.a.
P27	301/344 ^L 301/343 ^F	301.035	5.72	C ₁₅ H ₉ O ₇	-1.182	178.9988 (100), 151.0040 (80), 273.0407 (13), 257.0453 (11)	quercetin (5) ^a
P28	537.1/370 ^{L,F}	537.0825	6.15	C ₃₀ H ₁₇ O ₁₀	-0.335	443.0409 (100), 385.0718 (55), 417.0617 (9)	l3,1l8-biapigenin (33)
P29	299.1/393 ^L 299.2/389 ^F	299.0921	6.51	C ₁₇ H ₁₅ O ₅	-1.394	238.0627 (26), 205.0502 (18), 195.0081 (25), 193.0504 (12), 179.0347 (100), 135.0450 (11)	n.a.
P30	685.1/402 ^F	685.1195	6.67	C ₃₅ H ₂₅ O ₁₅	-0.559	535.0671 (100), 649.3593 (2)	n.a.
P31	277.1/407 ^F	277.1076	6.78	C ₁₅ H ₁₇ O ₅	-1.901	249.1131 (100), 158.9243 (4)	n.a.
P32	327.1/411 ^L	327.087	6.86	C ₁₈ H ₁₅ O ₆	-1.136	327.0873 (25), 309.2050 (37), 308.2230 (47), 284.0324 (22), 272.0323 (88), 271.0249 (100), 258.0169 (17)	tetrahydroxy-prenyl-xanthone
P33	341.1/418 ^L	341.1025	6.96	C ₁₉ H ₁₇ O ₆	-1.734	326.0794 (100), 309.0767 (93), 272.0236 (29), 217.0507 (68), 190.0637 (21)	trihydroxy-methoxy-prenylxanthone
P34	525.1/424 ^{L,F}	525.1191	7.05	C ₃₀ H ₂₁ O ₉	-0.055	507.1083 (8), 270.0534 (100), 254.0586 (21)	hydroxyemodin dianthrone (26)
P35	331.2/424 ^L 331.2/425 ^F	331.1912	7.08	C ₂₀ H ₂₇ O ₄	-0.944	194.0583 (100), 166.0636 (41), 151.0037 (11)	O-geranyl-methylpropanoyl-phloroglucinol
P36	325.1/445 ^L	325.0718	7.34	C ₁₈ H ₁₃ O ₆	0.150	325.0716 (20), 310.0480 (100), 309.0402 (21), 307.0610 (37), 295.0246 (11), 173.0609 (6), 151.0039 (5)	n.a.
P37	269.2/456 ^L 269/453 ^F	269.0455	7.52	C ₁₅ H ₁₀ O ₅	-0.359	269.0452 (100), 241.0503 (7), 169.1233 (4), 225.0557 (42)	emodin (20) ^a
P38	699.1/459 ^{L,F}	699.135	7.63	C ₃₆ H ₂₇ O ₁₅	-0.834	519.0727 (100)	S-skyrin-6-O- β -glucopyranoside (29) ^a
P39	521.1/458 ^L 521.1/459 ^F	521.0876	7.65	C ₃₀ H ₁₇ O ₉	-0.432	521.0878 (100)	protopseudohypericin (27)
P40	669.1/463 ^{L,F}	669.1246	7.7	C ₃₅ H ₂₅ O ₁₄	-0.521	519.0720 (100)	S-(+)-skyrin-6-O- β -arabinofuranoside (30) ^a
P41	501.3/463 ^L 501.3/464 ^F	501.3014	7.7	C ₃₃ H ₄₁ O ₄	-0,732	501.0628 (90), 486.0382 (16), 457.0731 (10), 432.2307 (100), 383.2227 (14)	Unidentified phloroglucinol I
P42	519.1/466 ^L 519.1/467 ^F	519.0718	7.75	C ₃₀ H ₁₅ O ₉	-0.800	519.0722 (100), 503.0408 (7), 487.0458 (13)	pseudohypericin (4)
P43	467.3/470 ^L 467.3/471 ^F	467.3163	7.82	C ₃₀ H ₄₃ O ₄	-0.863	398.2461 (100), 383.2227(5), 329.1757 (5), 327.1965 (5)	hyperifirin (40)
P44	509.1/474 ^{L,F}	509.1239	7.88	C ₃₀ H ₂₁ O ₈	-0.611	491.1131 (6), 254.0584 (100)	Emodin dianthrone (24)
P45	413.3/475 ^L 413.3/476 ^F	413.2702	7.91	C ₂₆ H ₃₇ O ₄	1.275	369.2793 (19), 369.2424 (13), 276.1365 (54), 275.1289 (32), 233.0819 (100), 221.0819 (19), 208.0741 (38)	deprenylhyperpolyphyllirin (38)
P46	481.3/483 ^L 481.3/482 ^F	481.3326	8.04	C ₃₁ H ₄₅ O ₄	0.638	437.3430 (100), 411.2922 (79), 343.1918 (24), 276.1368 (27), 233.0821 (56), 208.0742 (15), 207.0666 (26)	hyperpolyphyllirin (39)
P47	331.2/487 ^L 331.2/489 ^F	331.1914	8.14	C ₂₀ H ₂₇ O ₄	-0.189	331.1919 (24), 313.1812 (11), 287.2020 (100), 262.1214 (11), 261.1501 (17), 207.0666 (24), 194.0587 (6)	3-geranyl-methylpropanoyl-phloroglucinol (17) ^a
P48	345.3/487 ^L 345.3/488 ^F	345.2070	8.21	C ₂₁ H ₂₉ O ₄	-0.153	327.1967 (13), 301.2173 (100), 276.1368 (11), 261.1496 (18), 221.0820 (26)	3-geranyl-methylbutanoyl-phloroglucinol ^a
P49	535.4/496 ^L 535.4/497 ^F	535.3786	8.25	C ₃₅ H ₅₁ O ₄	-1.37	466.3095 (100), 397.2382 (34), 383.2234 (64), 315.1606 (42), 313.1814 (30)	hyperforin (1)

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P50	549.4/499 ^L 549.4/501 ^F	549.3945	8.31	C ₃₆ H ₅₃ O ₄	-0.734	480.3243 (100), 411.2539 (21), 397.2382 (38), 329.1756 (32), 313.1807 (27)	adhyperforin (2)
P51	331.2/500 ^{L,F}	331.1914	8.33	C ₂₀ H ₂₇ O ₄	-0.038	287.2018 (78), 261.1099 (5), 207.0664 (35) 194.0586 (100), 152.0118 (13)	2-O-geranyl-methylpropanoyl-phloroglucinol (41)
P52	505.1/502 ^L 505.1/503 ^F	505.0924	8.40	C ₃₀ H ₁₇ O ₈	-1.071	505.0930 (100)	protohypericin (18)
P53	345.2/507 ^{L,F}	345.207	8.44	C ₂₁ H ₂₉ O ₄	-0.326	316.1996 (14), 301.2169 (64), 209.0818 (17), 208.0740 (100), 152.0116 (18)	2-O-geranyl-1-methylbutanoyl-phloroglucinol (42)
P54	413.2/508 ^{L,F}	413.2333	8.46	C ₂₅ H ₃₃ O ₅	-0.236	343.1913 (100)	maculatoquinone (37)
P55	503.1/515 ^L 503.1/517 ^F	503.0775	8.56	C ₃₀ H ₁₅ O ₈	0.555	503.0773 (100), 459.0874 (11), 461.0668 (4)	hypericin (3) ^a
P56	485.3/522 ^{L,F}	485.3274	8.7	C ₃₀ H ₄₅ O ₅	0.396	457.3325 (100)	unidentified phloroglucinol II
P57	551.4/528 ^{L,F}	551.3741	8.71	C ₃₅ H ₅₁ O ₅	-0.250	483.3087 (25), 482.3041 (100), 467.2800 (12), 413.2335 (42), 411.2545 (88), 399.2181 (87), 331.1553 (65), 329.1766 (53)	furohyperforin (32)
P58	401.3/541 ^F	401.2697	9.01	C ₂₅ H ₃₇ O ₄	-0.207	401.2704 (100), 383.2593 (26), 195.0662 (8)	unidentified phloroglucinol III
P59	567.4/542 ^L 567.4/543 ^F	567.3696	9.05	C ₃₅ H ₅₁ O ₆	0.824	471.3116 (100), 497.3272 (14), 453.3007(15), 415.3218 (12)	unidentified phloroglucinol IV

^a identified with standard compound

* acetylation at the sugar moiety

^L feature in leaf data

^F feature in flower data

Hypericum is harvested for pharmacological purposes during the flowering period to achieve the highest possible secondary metabolite content. That is why the study considers not only the leaf material but also the flowers. The analysis of the flower samples was performed independently from the leaf samples because a joint data processing was not possible. Since both data sets were measured at different time points, minimal retention time shifts occur between the features. The PCA shown in Figure 4.4 represents the main variances within the flower data set. In accordance with the PCA of the leaves (Figure 4.3c), GT 3, 21, and 29 are separated from the other GTs, with GT 3 and one replicate of GT 29 being separated most clearly. Responsible for this are, as shown in the loadings plot (Figure 4.4b), flavonoids, and prenylated phloroglucinols. For the separation of the three GTs an increased content of smaller prenylated phloroglucinols, such as **15**, P35, P45, P46, and P58, and the unidentified compound P30 (C₃₅H₂₆O₁₅) are responsible. Hyperforin (**1**), adhyperforin (**2**), and their degradation product furohyperforin (**32**), however, appear to be reduced in these lines. The separation towards PC2 is mainly characterized by the presence of acetylated flavonoids (P12, P13, P23, P24) and astilbin (**19**). Whereby the GTs shifted in a negative direction, have increased values.

The comparison of PC3 and 4 (Figure 4.4c) represents a separation into two groups. The shift towards positive PC3 values of group A correlates with reduced rutin (**6**) values and increased values of various small phloroglucinols (P25, P31, P43, P45, P46, P58).

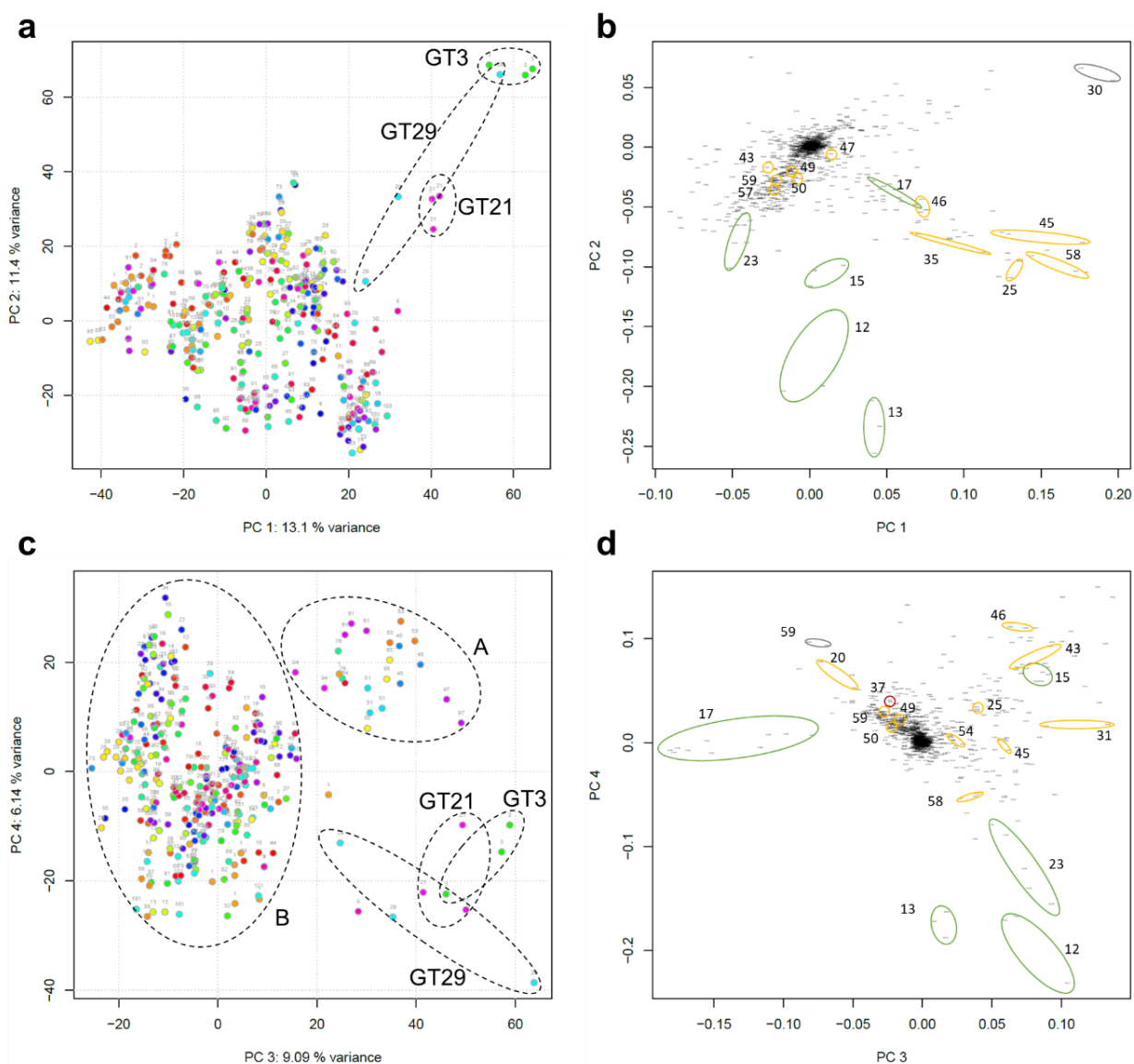


Figure 4.4 PCA of UHPLC-MS data of *H. perforatum* flower samples: scores plots are coloured by GT and show a) PC1 and PC2 and c) PC3 and PC4. The corresponding loadings plots with assigned loadings (peak numbers according to Table 4.1) coloured by compound class (green = flavonoids, yellow = prenylated acylphloroglucinols, blue = xanthenes, violet = organic acid, red = naphthodianthrone) b) PC1 and PC2, and d) PC3 and PC4.

4.3.3 Relation of metabolite profiles and GT characteristics

As described above, the different GTs vary in their secondary metabolite composition. The applied experimental design was intended to largely exclude environmental factors as a source of metabolite changes. In conclusion, the remaining variance is caused by genetic differences. Therefore, in this chapter, possible parallels between the metabolite composition and genetic or phenotypic characteristics such as ploidy, reproduction strategy, flowering behavior, or the origin will be discussed. Metadata were provided by our project partner at IPK Gatersleben (Appendix 3.1). The genetic parameters were determined according to Molins et al. (2014).⁶⁴ The investigated *H. perforatum* GTs are mainly tetraploid or hexaploid. Just three diploid lines are included. By staining the PCAs based on leaf data with the ploidy level (Figure 4.5b), no correlation to the

metabolite composition was visible. This observation is not surprising since Molins et al. (2014)⁶⁴ and Koch et al. (2013)¹⁰⁷ could show that genetic clustering is also largely independent from the ploidy level.

By examining the microsatellite variation, Molins et al. (2014)⁶⁴ were able to assign the GTs to three genetic clusters. The three genetic clusters “red”, “green”, and “blue” can also occur mixed in some GTs (Appendix 3.1) and are spread over North America and Europe equally. Interestingly, it is assumed that the “blue” genetic cluster is possibly a genetic variation of *H. maculatum*, which is intercrossed in *H. perforatum* over time.^{64,107} Staining the PCA based on the genetic clusters (Figure 4.5a), revealed no clear separation between the backgrounds. Nevertheless, the GT group A, formed in the positive direction of PC1, doesn't contain GTs with the “blue” background. Only the “red”, “green” and “mixed” (marked in yellow) forms are included. Group A is mainly characterized by a low amount of rutin (**6**) (Figure 4.4). These findings are in contrast to most of the literature data, which report the absence of rutin (**6**) in *H. maculatum* compared to *H. perforatum*.^{75,147-149} Nevertheless, Umek et al. (1999)⁷⁶ and Rusalepp et al. (2017)⁶¹ reported higher amounts in *H. perforatum*, and a number of studies were able to detect rutin (**6**) in *H. maculatum* as well.^{60,77,150,151} However, regarding the literature, which compared the composition of the major compounds of *H. maculatum*, the results are inconsistent. A summary of the literature is given in Appendix 4.4. However, only Raclariu et al. (2017)⁵⁶, Bagdonaitė et al. (2012)¹⁴⁷ and Radusienė et al. (2004)¹⁵² compared major metabolites of the species under comparable conditions in the field. The other studies used *in vitro* cultures^{117,151,153} or plants from the wild^{60,61,76,77,92,144,148,149,154,155}, which are highly influenced by culture conditions and environmental factors. Nevertheless, all studies included just one or a small number of genetic lines representing each species, and could, therefore, not appropriately cover the diversity of the highly varying species. Apart from the major constituents, Nedialkov et al. (2015) identified in *H. maculatum* new prenylated phloroglucinols and maculatoquinones.¹⁵⁶ Zheleva-Dimitrova et al. (2012) detected benzophenone derivatives.¹⁵⁷ However, by checking for possibly fitting features of these compounds, no significant difference between lines with different genetic backgrounds could be observed. This leads to the conclusion that although the blue background originates from *H. maculatum*, the secondary metabolite levels are not significantly different.

90 GTs were originally collected at different places in Europe and North America from the wild before cultivation. In our sample set, no differences between the metabolite profiles according to the continent were detected. *H. perforatum* is not endemic to North America and was probably introduced at the beginning of 18th century.¹⁵⁸ Molins et al. (2014) proposed that *H. perforatum* was introduced multiple times because all three gene clusters from Europe were also found in North America.⁶⁴ The same genetic pools on both continents are reflected in a similar metabolite variance.

However, often different GTs were sampled per site. The PCA based on the sample origin is shown in Figure 4.6. In most cases, sample sites include GTs with similar metabolite composition, for

example GTs from Silwood in England (violet square) or La Selva de Mar in Spain (blue square). Normally GTs from a similar sample site are closely related because of the facultatively apomictic reproduction of *H. perforatum*.^{64,107} The asexual reproduction via seeds allows the fast-spreading of progeny with identical characteristics like the mother plant. This reproduction strategy represents one reason for the invasive behavior of this species. Interestingly, different sample sites close to the Great Lake region (Michigan MI and Wisconsin WI) do not cluster together. Although the GTs possess geographical proximity, they are separated in PCA by the specific sampling sites. This reflects the findings of Molins et al. (2014), who identified the Great Lake region as a center of genetic variability.⁶⁴

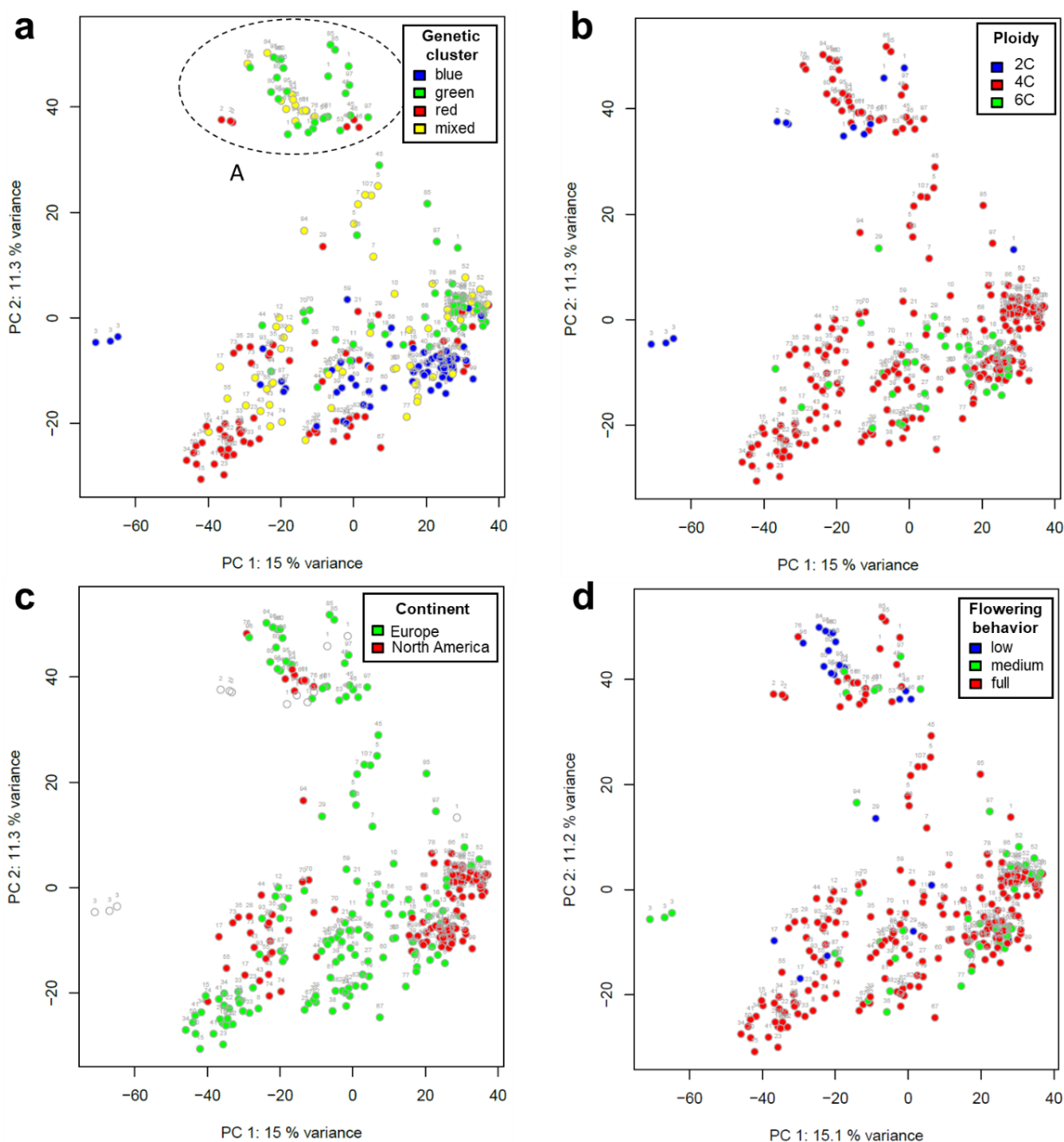


Figure 4.5 PCA of UHPLC-MS data of leaf samples: scores plots are colored by a) genetical background, b) ploidy, c) continent d) flowering behavior (full = harvest of 30 flowers for each replicate, medium = two replicates between 20-10 flowers, low = two replicates lower than 10 flowers).

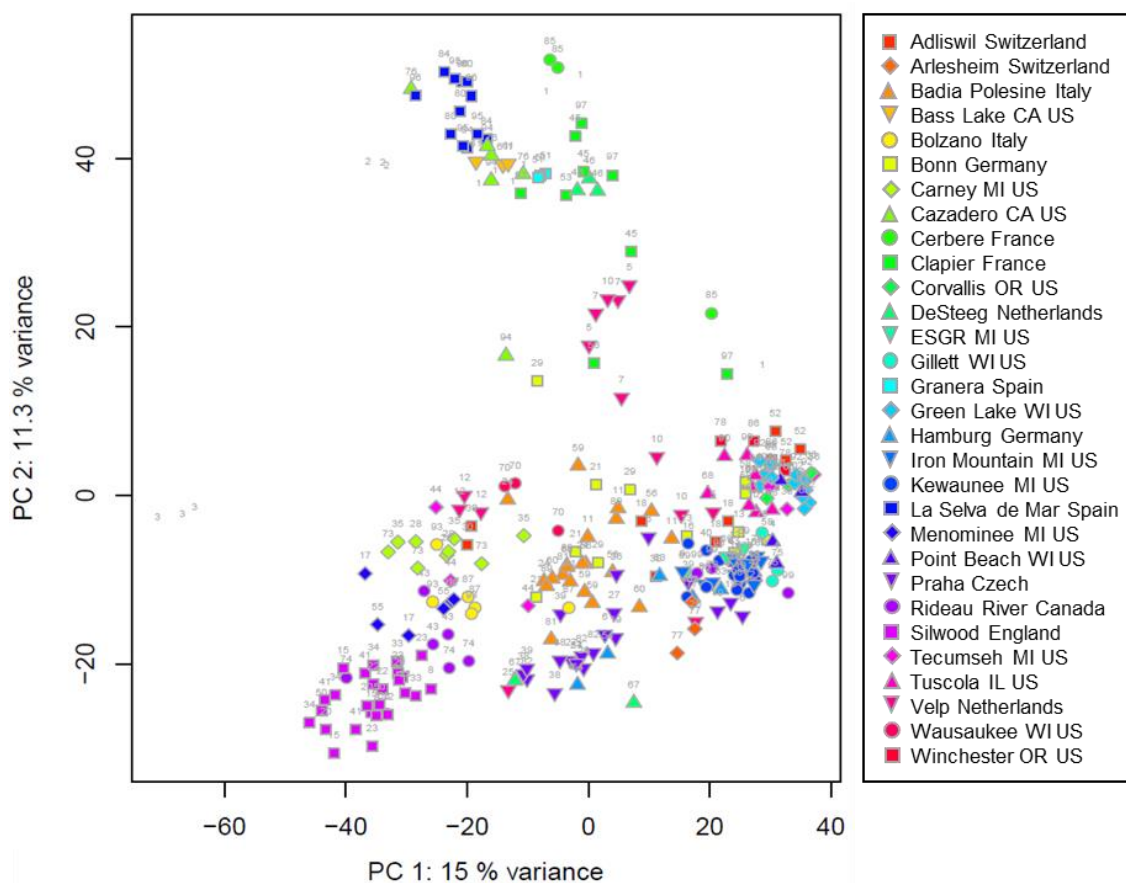


Figure 4.6 PCA of UHPLC-MS data of *H. perforatum* leaf samples: scores plot is colored by sampling site. Not colored are GTs 1, 2, and 3 originated from botanical gardens.

4.3.4 Comparison of major compounds and the suitability in the pharmaceutical industry

4.3.4.1 European Pharmacopoeia criteria

Hypericum perforatum is applied against mild to moderate depression, and therefore became one of the best-selling herbs in the Western World. To fulfill the quality criteria as *Hyperici herba* according to the European Pharmacopoeia, the presence of certain marker compounds has to be analyzed. In this chapter, the investigated GTs were checked for these quality characteristics. In contrast to the prescribed HPLC-DAD method, our methodology is not used for quantification, so that just relative intensities are compared. Considered compounds are the two phloroglucinols, hyperforin (1), and adhyperforin (2), the flavonoids, quercetin (5), rutin (6), hyperoside (7), quercitrin (8), isoquercetin (9), and biapigenin (33), as well as the naphthodianthrone, hypericin (3) and pseudohypericin (4). The compounds are marked in the chromatogram, shown in Figure 4.7. Due to only a small retention time difference between hyperoside (7) and quercitrin (8), these isomers cannot be separated by the automated processing. Therefore, these two marker compounds were considered with one value and the peak number P16. However, the ratio between these two compounds can be evaluated in the chromatogram visually (shown in Figure 4.7c).

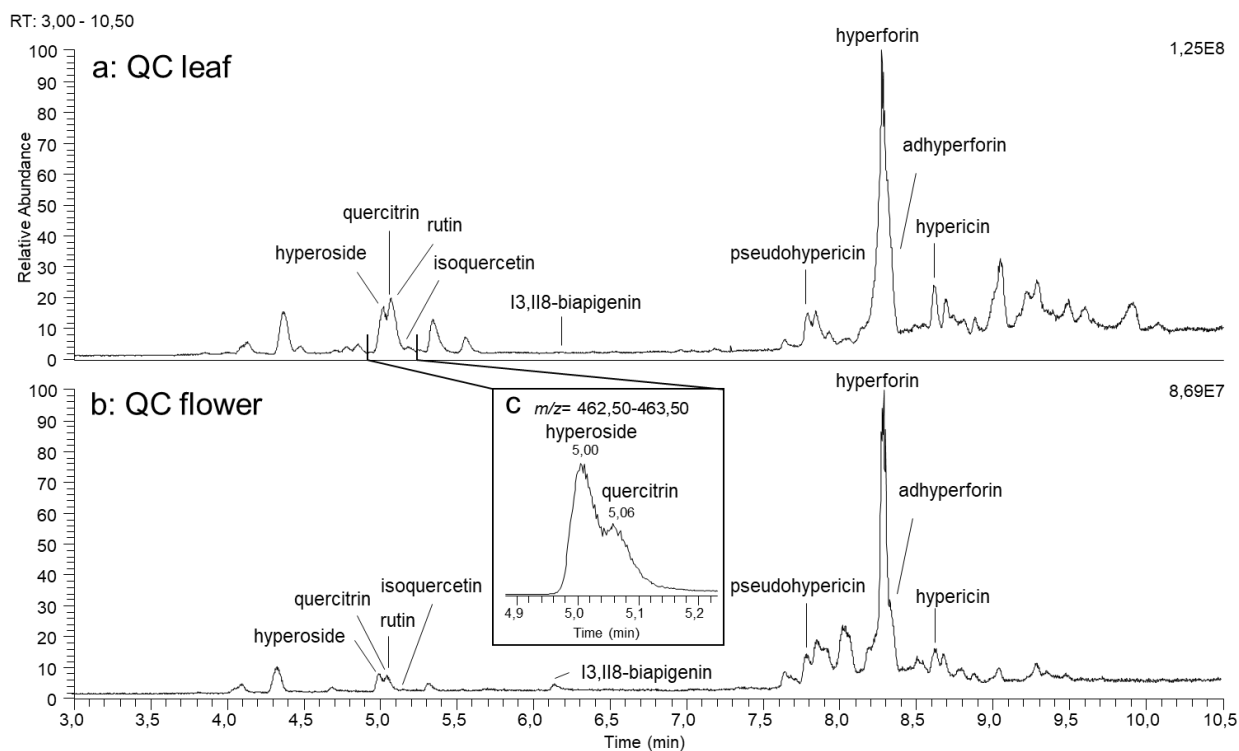


Figure 4.7 LC-MS chromatogram (TIC) of: a) pooled quality control (QC) of leaf samples and b) pooled quality control (QC) of flower samples. c) Mass range chromatogram of m/z 463 of leaf data with Rt window 4.9–5.2 min.

The average intensities of the marker compounds for each GT were calculated and displayed in heatmaps of the leaf and flower data in Figure 4.8. First of all, it is noticeable that the phloroglucinols, hyperforin (**1**) and adhyperforin (**2**), exhibit the highest intensities in both organs. In leaves, the naphthodianthrones and quercetin derivatives possess comparable intensities, whereas I3,II8-biapigenin (**33**) is much lower. In the flowers, I3,II8-biapigenin (**33**) reaches the intensities of quercetin derivatives and naphthodianthrones.

Within the heatmap, a hierarchical clustering of the GTs based on the marker compounds was performed. In the leaf data, the GTs cluster into four groups (L1–L4). Cluster L1 is separated from the main group, which is subdivided into the three cluster L2–L4. Cluster L2 comprises the highest number of GTs. The right subgroup of the cluster L2 possesses identical profiles. Interestingly all these GTs originate from England. Obviously, the geographical separation of the GTs resulted in the same compound pattern, which indicates that they descended from the same relatives and were possibly produced by asexual reproduction.

The average metabolite profile of each cluster was calculated to simplify and present the characteristics in a radar chart (Figure 4.9a). The scale for each compound is based on the minimum and maximum value measured among all GTs, so that a comparison between compounds is not feasible. Cluster L1 is characterized by a reduced I3,II8-biapigenin (**33**) content in comparison to the others. Apart from that, L1 is similar to L2. GTs in cluster L3 and L4 possess a reduced rutin (**6**) content. Additionally, L3 has a slightly increased content of hyperoside/isoquercitrin (**7/9**), quercitrin (**8**), hypericin (**3**), and I3,II8-biapigenin (**33**). L4

represents just the GT3, which was already separated in the untargeted PCA analysis (Figure 4.3). The profile of L4/GT3 show many abnormalities (appendix 4.5). The major difference to the other GTs is a minimal phloroglucinol content in GT3. Compared to all GTs also the quercitrin (**8**) and the hypericin (**3**) contents are low. However, the differences are not significant (compare Figure 4.8a) because the fluctuation of these compounds over all samples is small. The hyperoside/isoquercitrin (**7/9**) content of L4/GT3 is slightly increased.

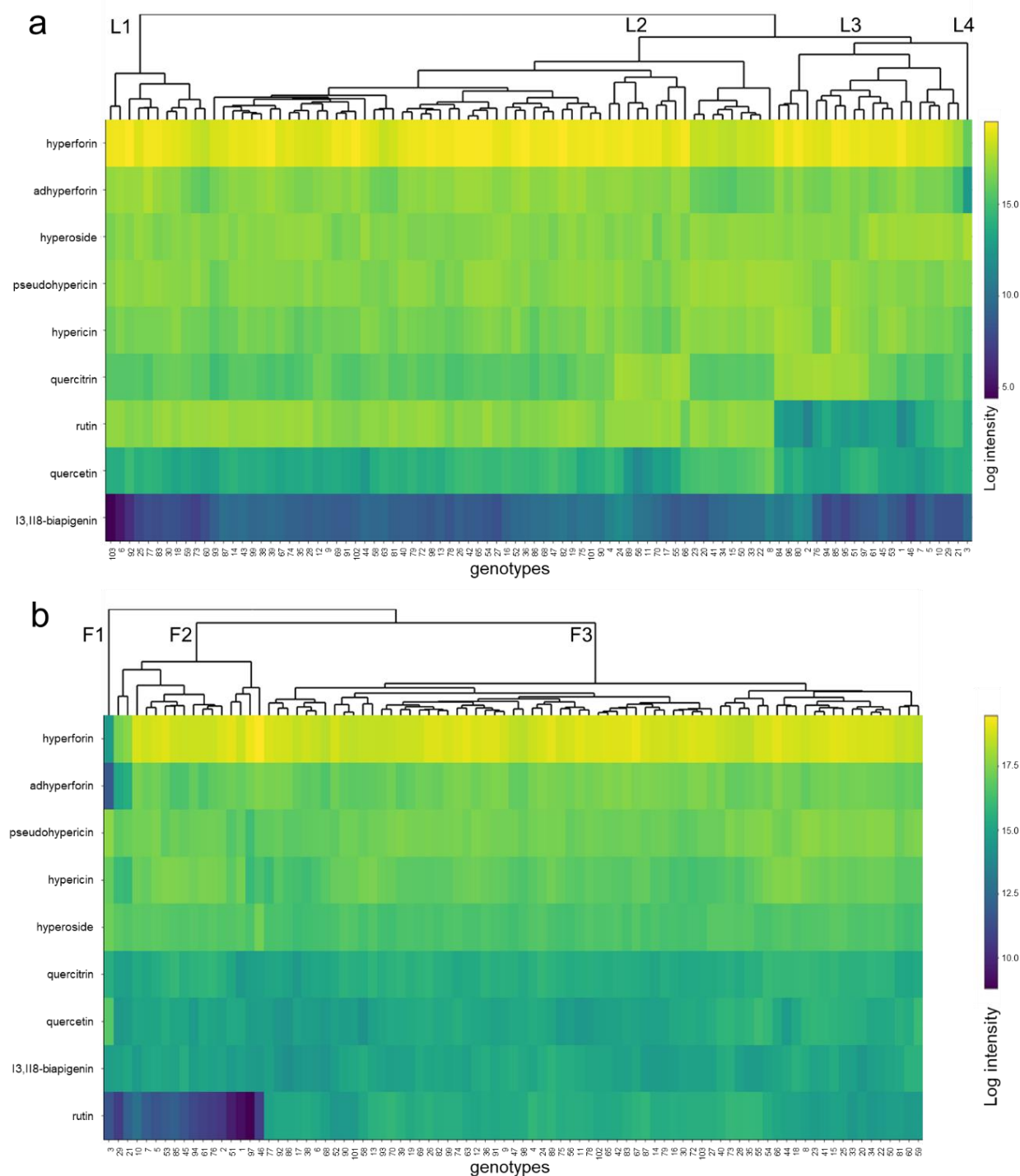


Figure 4.8 Heatmap of marker compounds intensities (log) for 93 *H. perforatum* GTs clustered with hierarchical cluster analysis: a) leaf data (cluster L1-L4), b) flower data (cluster F1-F3).

Figure 4.8b shows the heatmap based on the flower data. The HCA clusters the GTs in three strict groups, indicated by a long distance. The average content per compound of each cluster is presented as radar chart in Figure 4.9b. Cluster F1 contains solely the unique GT3, in which hyperforin (**1**), adhyperforin (**2**), and rutin (**6**) are almost absent. Further, GT3 shows a significantly higher quercetin (**5**) amount. The lack of rutin (**6**) is the reason for the separation of cluster F2. Notably, GTs 21 and 29 in this cluster have fewer phloroglucinols than the other GTs of the cluster F2. The biggest cluster F3 comprises more than 75% of the investigated GTs and contains all marker compounds in a comparable amount.

In general, the heatmaps show that the differences between the GTs are more prominent in flower data, but the same trend is found in the leaf data. The formed clusters of the flower data match with the clusters of the leaf data: F1 equals L4, F2 equals L3, and F3 equals L1 and L2. Especially GT3 (L4, F1) is conspicuous because it contains significantly less hyperforin and adhyperforin. In general, the chromatogram of GT3 differs significantly from an average *H. perforatum* profile (appendix 4.5). Also, in the untargeted analysis, this GT showed the highest deviation (Figure 4.3 and Figure 4.4). A GT with a drastic decrease of hyperforin and adhyperforin is, apart from the chapter 2, where another GT show that characteristics, not known from the literature.

13,II8-Biapigenin (**33**) varies in the leaf data but not in flower data. For flower data, similar results were obtained by Büter et al. (1998).¹⁵⁹ They compared the major compounds of flowers from seven *H. perforatum* GTs and reported no inter-accession changes of **33**. This compound occurs mainly in buds and flowers,¹⁶⁰ therefore, the content variation in leaves will not significantly affect the overall content in the herbal drug, where the whole upper flowering plant is used.

A significant intraspecific variation occurs for rutin (**6**). In flower data 17 GTs and in leaf data 21 GTs contain significantly reduced amounts of this quercetin derivative glycosylated with the disaccharide rutinose at position 3. The different numbers for leaf and flower samples emerge because four of the nonflowering GTs were lacking **6**. Rutin (**6**) is formed via the transfer of glucose and rhamnose to quercetin (**5**). Interestingly, the precursor quercetin (**5**) variates independently. The correlation between rutin (**6**) and the monoglycosidic precursor isoquercitrin (**9**) could not be evaluated because it coelutes with its isomer hyperoside (**7**) (Figure 4.7c). At least for the sum of both compounds, no correlation is observable. The last biosynthetic step, known from buckwheat,¹⁶¹ is catalyzed by a flavonol-3-*O*-glucoside L-rhamnosyltransferase. Whether the absence of this enzyme is responsible for the rutin-poor chemotype might be verified by transcriptome analysis experiments. This rhamnosyltransferase differs from the rhamnosyltransferase producing quercitrin (quercetin-3-*O*-rhamnosid). Thus quercitrin (**8**) was still present in all GTs.

The high intraspecific variability of rutin (**6**) was already shown in the preliminary study (chapter 2). Several studies of *H. perforatum* identified rutin as a highly genotype-dependent compound.^{59,60,63,159,162,163} Franke et al. (1999) performed the study with the highest number of accessions. They quantified the major compounds of 153 lines and characterized especially two

main groups: one with a high rutin (**6**) and hyperoside and isoquercitrin (**7/9**) content and the other one with low rutin (**6**) and high hyperoside and isoquercitrin (**7/9**) content.¹⁶³ The study of Scotti et al. (2018) comprised 71 samples of primary *H. perforatum* material, whereby 38% did not contain **6** in detectable amounts.⁵⁹ These studies describe in accordance with our findings rutin (**6**) as a major discriminant compound. Fillipini et al. (2010) suggested the compound as a taxonomic marker to distinguish between the *H. perforatum* subspecies: *perforatum*, *angustifolium*, and *veronense*. They stated that subspecies *veronense* contains **6** exclusively.¹⁶² However, because they have not included multiple accessions or biological replicates per subspecies, the result might be coincidental. Since in our investigation, the GTs were not differentiated in subspecies, this hypothesis can not be addressed.

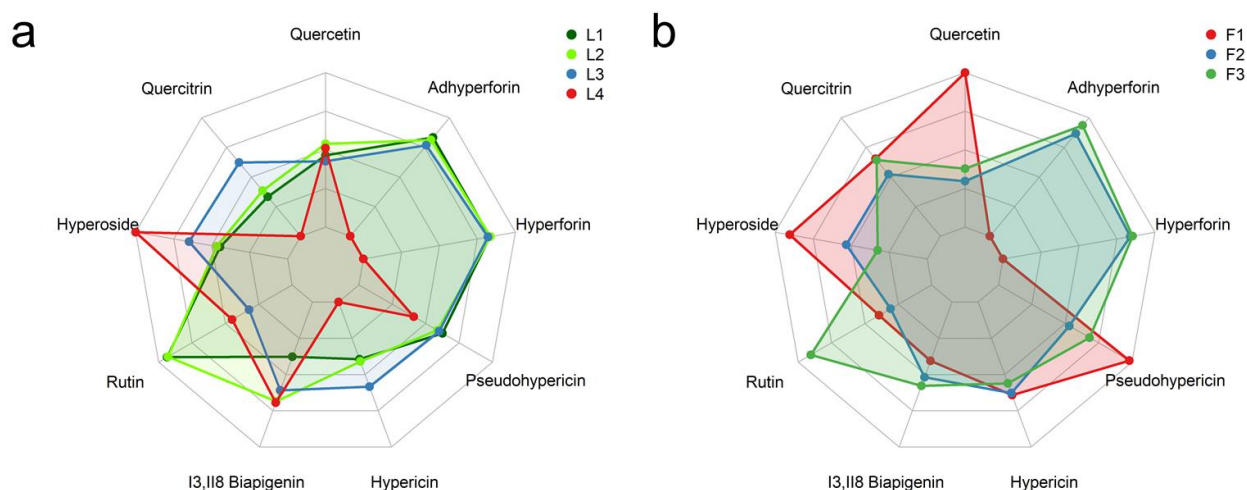


Figure 4.9 Radar chart representation of the mean intensity of each marker compound for the defined cluster (Figure 4.8) for a) leaf data and b) flower data. As scale, the minimum and maximum value for each compound over all GTs was used.

4.3.4.2 Suitability of the GTs for pharmaceutical usage

22% of the analyzed GTs are not in line with the European Pharmacopoeia 10 because of the absence of rutin (**6**). Therefore rutin should not be used as marker compound to distinguish *H. perforatum* from related species. GT 3 was found to be outstanding by lacking hyperforin and adhyperforin. The other marker compounds have not shown dramatic intraspecific changes.

Various activities are attributed to *Hypericum perforatum* extracts, summarized by Avato et al. (2005)¹⁶⁴ and Barnes et al. (2001,2019)^{10,165}.

Antidepressant properties

In the Western World, the extracts are mainly applied against mild to moderate depression. Nowadays, it is known that several compounds with different modes of action contribute to the antidepressant activity. Nevertheless, the complex mixture of the crude extract revealed better results than the isolated compounds. None of the identified compounds was able to explain the all effects of the crude extract.

Hyperforin (**1**) is assumed to be the main compound responsible for the effects, by impeding nonselectively the synaptosomal uptake of several neurotransmitters. Although the mechanism is not yet clarified, *in vitro* studies suggest that hyperforin does not inhibit the monoamine transporters themselves, but rather mediates a reduced reuptake by increasing the free intracellular Na⁺ concentration¹⁶⁶ and/or the neurotransmitter storage in synaptic vesicles¹⁶⁷. For compounds a dose-response relation is proven¹⁶⁷, which is why pharmaceutical *H. perforatum* medications are sometimes ideally also standardized to their hyperforin content. Special care has to be taken, with hyperforin contents higher than 1 mg hyperforin per day, because severe adverse effects were reported. The intake of hyperforin reduces the bioavailability of other drugs, like digoxin, ciclosporin, and oral contraceptives, which are substrates of cytochrome P450 enzymes (CYP3A4) or p-glycoprotein (ABCB1).⁴⁰

Flavonoids and naphthodianthrones were also shown to have antidepressant effects *in vitro*, by inhibiting monoamine oxidases (MAO).^{168,169} However, the results are questionable due to a much lower plasma concentration *in vivo*.¹⁷⁰ As a consequence, the application of an extract from GT 3 to depressive patients is not suitable because it may not reduce their symptoms.

Apart from the direct effects against depression, the higher activity of the crude extract compared to single compounds could be explained by a pharmacokinetic synergy. Especially flavonoids were reported as solubility and permeability modifiers of the poorly soluble hypericin (**3**). The addition of hyperoside (**7**) and rutin (**6**) can lead to a 400 times increased concentration of hypericin (**3**) in water.¹⁷¹ Furthermore, the permeability of hypericin across Caco-2 cells was enhanced after the addition of a combination of quercitrin (**8**) and rutin (**6**).¹⁷² *In vivo* the oral bioavailability of hypericin (**3**) was 58% higher in the presence of hyperoside (**7**) and procyanidin (P4) in rat plasma.¹⁷³ This leads to the conclusion that a high quercetin-glycoside content is desirable to produce an effective medicinal extract.

The study of Nöldner et al. (2002)¹⁷⁴ highlighted the importance of rutin (**6**) in the extract to treat depressive disorders. When they tested two different extracts, only the rutin-rich formulation was active. After adding rutin in a sufficient amount (3%) to the inactive extract, it showed activity in the forced swimming test on rats. The researchers found no dose dependency and assume that the effect is reached after a threshold limit.¹⁷⁴ Manufacturers should consider testing the plant material for the amount of rutin (**6**) and other major flavonoids because our results show that around 20% of the GTs lack compound **6**. Until now, the plant extracts are just standardized to their hypericin (**3**) and sometimes additional hyperforin (**1**) content. This leads to the variability of the flavonoid portion in the extracts and could be responsible for inconsistencies in clinical studies.

Alzheimer's disease

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive loss of memory and dementia. The connected histopathological features are amyloid- β -deposition ($A\beta$) and abnormal tau-protein. Accompanied by oxidative stress and neuro-inflammation, this leads to neuronal damage.¹⁷⁵

Several preclinical studies tested *H. perforatum* extracts *in vitro* and *in vivo* and revealed promising results against AD.^{11,176,177} However, there is no clinical evidence until now. As an example, Hofrichter et al. (2013) administered different extracts to APP-transgenic mice. They found a decrease of intracerebral $A\beta_{42}$ level and $A\beta$ -plaque size and number as well as microglia activation and a restored cognition in Morris water maze experiments. They assume that *H. perforatum* metabolites activate the ABCC1 transporter in the blood brain barrier so that monomeric $A\beta$ is removed from the brain.¹¹ A similar hypothesis had Brenn et al. (2014), who showed increased expression of the ABCB1 transporter.¹⁷⁶

Several studies tested pure compounds in diverse assays to identify the active ingredients of the extracts. Dinamarca et al. (2006)³⁸ showed that hyperforin (**1**) decreases dose- and time-dependent the $A\beta$ -aggregation. In contrast to that, Hofrichter et al. (2013)¹¹ mentioned that there is no correlation between the effects and the hyperforin (**1**) content. Another study showed that hypericin (**3**) interacts with the $A\beta$ -fibrils, without reducing the number.⁹⁸

Licensed medicaments against AD inhibit the Acetylcholinesterase (AChE), which reduces the neurotransmitter acetylcholine in the brain. Hernandez et al. (2010)⁸⁵ reported the highest AChE inhibition by hypericin (**3**) (5.7 $\mu\text{g/ml}$), whereas flavonoids like quercetin (**5**) exhibited moderate results. However, they did not include phloroglucinols, which showed promising results in the study of Guo et al. (2018)⁸². The acylphloroglucinol with the best AChE inhibition properties (6,9 μM) also inhibited β -secretase, responsible for the first step of the amyloidogenic processing.⁸²

The antioxidative compounds of *H. perforatum* were discussed to prevent neuron damage by reducing the oxidative stress in AD. In particular, the quercetin derivatives were able to decrease the formation of amyloid-induced reactive oxygen species in microglia. Catechin (**P6,14**) and epicatechin (**P4,14a**) have, in contrast, also cytoprotective effects that go beyond the antioxidant effect.⁸⁶ In our study, the GTs showed large fluctuations in acetylated flavonoids. This modification decreases their polarity and may facilitate to cross the blood-brain barrier and reach the place of action.

Guo et al. (2018)⁸² mentioned that for the treatment of a multifactorial disease like AD a multitargeted drug would be necessary. A complex compound mixture like an extract from *H. perforatum* addresses many different targets and may have a positive impact on AD patients. However, to obtain clinically reliable results, the metabolite profiles of the used plant extracts must be well investigated.

4.3.5 Correlation network analysis

The abundance of metabolites produced within the same biosynthetic pathway is often closely related. The correlation between involved compounds can be used to get more insights into biosynthetic processes in the plant. As an example, Kusari et al. (2009) could show that the abundance of hypericin (**3**) and the putative precursor emodin (**20**) highly correlates within different *Hypericum* species and organs.⁷⁷ In contrast to those studies, which used 30 samples maximum, the present study comprises for each organ over 200 samples, which increases the statistical power. Since the investigated GTs differ only slightly in their metabolite content, the data set is suitable for linear regression.

The following untargeted analysis assumes that pathway-connected compounds show the same relation to each other in each GT. The variation of compound intensities between the GTs and biological replicates enables the correlation of one feature across all samples to all other features. By enlarging the correlation network with unidentified features, a complex network is resulting, which offers the possibility to identify connected not yet identified compounds. The positive correlation (Pearson) analysis of the major identified compounds, shown in Table 4.1, were pictured in Figure 4.10 as correlation networks. The network, based on leaf samples (Figure 4.10a) displays three major groups of features, whereas the analysis of flower samples generates four groups (Figure 4.10b). A group is formed if many features correlate with each other, which indicates highly connected metabolites. The group marked red shows hypericin related compounds, and the yellow group exhibits phloroglucinols. The group marked green corresponds, in both organs, to catechin related compounds, like epicatechin and the dimer procyanidin. Additionally to that cluster, flavonoids are spread over the whole network and are often strongly connected to phloroglucinols and naphthodianthrones. Interestingly, the correlation network based on flower samples reveals two groups of phloroglucinols. These differences are discussed further in chapter 4.3.5.2.

4.3.5.1 Hypericin correlated metabolites

The hypericin pathway (Figure 3.4) is still not completely understood. The correlation network indicates a strong relationship between the hypericin biosynthesis connected compounds (red marked group Figure 4.10). To get further information, the intensities of hypericin (**3**) and pseudohypericin (**4**), as endproducts, were correlated to the whole feature list. The results of both organs are summarized in Table 4.2. Since no significant differences can be observed in the naphthodianthrone composition of leaves and flowers (Appendix 4.6), also the correlating features are mostly identical in both data sets (Table 4.2).

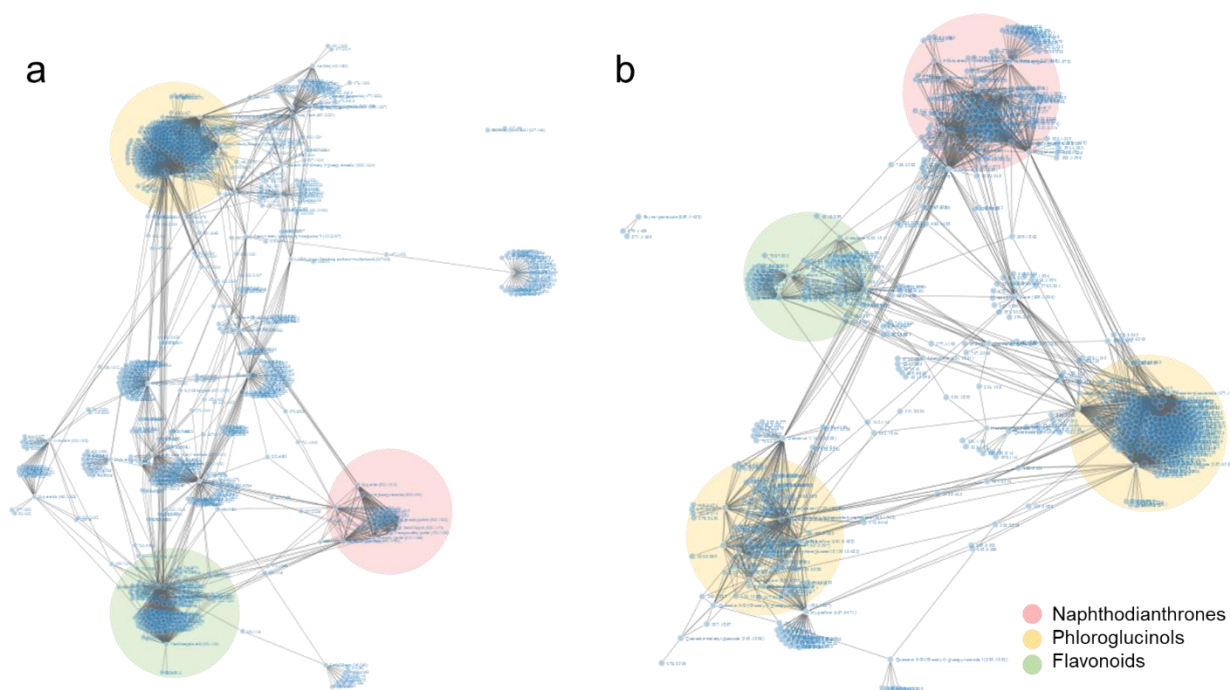


Figure 4.10 Correlation network of major identified compounds (Table 4.1) with all features ($r > 0.4$) for a) leaf and b) flower data from *H. perforatum* genotypes.

Monomeric anthraquinones

Hypericin (**3**) consists of two anthraquinone units. In most of the literature, it is assumed that the dimerization occurs between emodin (**20**) and emodin anthrone (**21**)^{27,55,77,96} catalyzed by a phenolic oxidative coupling protein (POCP).^{27,67} The correlation between emodin (**20**) and hypericin (**3**) was already reported using different *Hypericum* species.^{77,153} Recently, emodin (**20**) was also found in non-hypericin-containing *Hypericum* species, which raised doubts if emodin (**20**) is the precursor.^{39,101} The group of Kimáková et al. (2018)¹⁰¹ and our group⁶⁷ reported the co-occurrence of skyrin derivatives (**29**, **30**) and hypericin (**3**), which could be a hint to a different biosynthetic pathway (see chapter 3). By analyzing different pistil phenotypes, with and without dark glands, emodin (**20**) was not detectable, but the closely related endocrocin (**28**). In the present correlation analysis both compounds **20** and **28** correlated with the hypericin (**3**) and pseudohypericin (**4**) content in the flowers. This result is in line with several studies, which stated a correlation to emodin (**20**) and our previous detection of endocrocin (**28**) (chapter 3.3.2). Whether **28** is a by-product or an intermediate of the hypericin biosynthesis needs to be clarified. The abundance of both monomeric anthraquinones may hint to their common precursor athrochryson. As already discussed in chapter 3, this compound might play a central role as precursor for dianthrones. The reason for the exclusive and significant correlation in the flower data likely is due to the generally higher concentration of naphthodianthrones in flowers in comparison to leaves.¹³⁹

Furthermore, two correlating features were detected with m/z 431.1 and composition $C_{21}H_{19}O_{10}^-$ at Rt 5.56 min and 6.65 min. Despite the same elemental composition, the MS fragmentation

pattern differs strongly. For feature 431.1/399, the aglycone with m/z 269 is obtained as base peak with the molecular composition $C_{15}H_9O_5^-$ after cleavage of glucose ($C_6H_{10}O_5$). Two compounds with this formula are described to occur in *H. perforatum*, the anthraquinone emodin (**20**) and the flavonoid apigenin. In general, for the identification of anthraquinones in plants by MS, care must be taken because flavonoids often possess identical molecular formulas and similar fragmentation patterns. However, within the MS³ experiment of the aglycone with m/z 269, just the fragment m/z 225 which is typical for emodin (**20**), and not an additionally fragment m/z 149 typical for apigenin, could be detected. So the feature 431.1/399 was identified as emodin-8-*O*-glucoside (**34**). The MS² fragment m/z 311 corresponds to the loss of $C_4H_8O_4$, indicating a cleavage within the sugar moiety. This cleavage is characteristic for C-glycosides and was already described for puerarin (8-*C*-glycosyl-isoflavone) and fideloside (8-*C*-glycosyl-taxifolin).^{178,179} However, the comparison of emodin-1-*O*-glucoside and emodin-8-*O*-glucoside, also the 8-*O*-glucoside showed the fragment m/z 311, probably due to a stabilization of neighbouring atoms.¹⁸⁰ These findings indicate that emodin exists as glycoside in *H. perforatum*. In general, glycosides often represent storage compounds in plants. Glycosylation can enable transport of this lipophilic compound outside of the glands. That the abundance of emodin (**20**) is not restricted to the dark glands was reported by Kusari et al. (2015)⁹⁶ and Kucharíková et al. (2016)³⁹. They detected the compound also in gland free leaf tissue, whereas Zobayed et al. (2006)⁵⁵ were not able to identify emodin outside the dark glands.

In comparison to that, feature 431.1/338 initially loses $C_6H_{10}O_4$ so that $C_{15}H_9O_6^-$ (m/z 285) remains as base peak. The molecular formula of that fragment corresponds to kaempferol or to a methylated emodin like catenarin. Both substances were measured as reference compounds and their fragmentation was compared to the unidentified aglycon, but none of the compounds can be clearly assigned to the MS³ spectrum of 431.1/338.

Table 4.2 Correlating features to hypericin (**3**) and pseudohypericin (**4**) in the data sets (Ds) based on leaf data “L”(correlation coefficient > 0.4) and flower data “F” (correlation coefficient > 0.5). Mass fragmentations were obtained with ncd = 35 eV (or ncd = 45 eV marked with *).

Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	RDB	Fragment ions (Intensity)	Identification	Ds
269/453	269.0455	7.52	$C_{15}H_9O_5$	-0.359	11.5	269.0452 (100), 241.0503 (7), 169.1233 (4), 225.0557 (42) MS ³ [269]→ 181.0658 (100), 197.0607 (10), 197.1912 (10), 210.0320 (55), 225.0556 (59)	emodin (20)	F
285/366	285.0407	6.06	$C_{15}H_9O_6$	0.943	11.5	285.0401 (100), 267.1605 (5), 229.0504 (5), 151.0040 (5)	kaempferol ^a	F
301/343	301.0350	5.72	$C_{15}H_9O_7$	-1.182	11.5	178.9988 (100), 151.0040 (80), 273.0407 (13), 257.0453 (11)	quercetin (5) ^a	LF
313/275	313.0345	4.57	$C_{16}H_9O_7$	-2.862	12.5	269.0450 (100) MS ³ [269] → 269.0460 (100), 225.0548 (63)	endocrocin (28) ^a	F
329.1/242	329.0873	4.03	$C_{14}H_{17}O_9$	-1.414	6.5	291.0772 (23), 269.0651 (7), 221.0454 (11), 191.0349 (100), 167.0350 (77)	n.a.	F
355.3/534	355.3214	8.81	$C_{22}H_{43}O_3$	-1.009	1.5	309.3161 (100)	n.a.	F

Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	RDB	Fragment ions (Intensity)	Identification	Ds
371.2/334	371.2069	5.56	C ₁₉ H ₃₁ O ₇	-0.597	4.5	251.0927 (17), 209.1647 (38), 159.0301 (21), 143.0351 (10), 113.0247 (23), 101.0246 (11)	n.a.	L
391.2/306	391.1974	5.10	C ₁₈ H ₃₁ O ₉	0.074	3.5	331.1756 (100), 179.0561 (15)	n.a.	L
393.1/310	393.1193	5.17	C ₁₉ H ₂₁ O ₉	-3.779	9.5	273.0764 (9), 255.0660 (62), 231.0660 (100)	n.a.	F
407.2/334	407.1840	5.56	C ₂₅ H ₂₇ O ₅	-2.407	12.5	375.0505 (100), 285.0405 (84)	n.a.	L
415.2/474	413.2702	7.91	C ₂₆ H ₃₇ O ₄	1.275	8.5	369.2793 (18), 369.2424 (12), 276.1365 (53), 275.1289 (32), 233.0819 (100), 221.0819 (19), 208.0741 (38)	n.a.	F
419.1/267	419.0977	4.46	C ₂₀ H ₁₉ O ₁₀	-1.694	11.5	375.1079 (100), 237.0556 (19), 213.0555 (51), MS ³ [375] → 267.0663 (4), 255.0660 (5), 237.0554 (54), 213.0554 (100)	n.a.	LF
431.1/338	431.0980	5.56	C ₂₁ H ₁₉ O ₁₀	-0.812	12.5	371.2076 (13), 285.0402 (100) MS ³ [285] → 267.0297 (38), 257.0450 (100), 241.0503 (37), 229.0504 (55), 213.0558 (22), 197.0608 (18), 163.0038 (24) MS ³ [371] → 209.1547 (40), 161.0457 (100)	n.a.	LF
431.1/399	431.0987	6.65	C ₂₁ H ₁₉ O ₁₀	0.835	12.5	311.0560 (48), 282.0534 (12), 269.0454 (100) MS ³ [269] → 269.0454 (100), 225.0555 (50) MS ³ [311(35)] → 282.0534 (100), 268.0377 (19), 240.0436 (7)	emodin-8-O-glucoside (34) ^b	LF
435.1/308	435.1305	5.13	C ₂₁ H ₂₃ O ₁₀	1.792	10.5	341.0668 (34), 315.0868 (74), 302.0335 (87), 273.0763 (100)	n.a.	F
447.1/320	447.0929	5.33	C ₂₁ H ₁₉ O ₁₁	-0.929	12.5	301.0350 (100), 300.0281 (35)	quercitrin (8)	F
463.1/269	463.1247	4.49	C ₂₂ H ₂₃ O ₁₁	0.141	11.5	431.0971 (100), 317.0658 (9), 299.0553 (78), 285.0398 (8) MS ³ [431] → 285.0401 (100) MS ³ [299] → 284.0312 (100)	n.a.	LF
471.3/543	471.3120	9.04	C ₂₉ H ₄₃ O ₅	-2.308	8.5	471.3120 (100), 453.3012 (14), 383.2594 (9)	n.a.	F
503.1/517	503.0775	8.56	C ₃₀ H ₁₅ O ₈	0.555	23.5	503.0773 (100), 459.0874 (11), 461.0668 (4) *MS ² [503] → 503.0773 (100), 487.0459 (6), 461.0665 (11), 459.0874 (32), 447.0874 (1), 431.0925 (3), 433.0718 (3)	hypericin (3) ^b	LF
505.1/503	505.0924	8.40	C ₃₀ H ₁₇ O ₈	-1.071	22.5	*MS ² [505] → 505.0930 (100), 489.0613 (5), 477.0981 (5), 463.0894 (11), 462.1091 (8), 461.1034 (40), 436.0552 (2), 433.1082 (2)	protohypericin (18) ^b	LF
509.1/474	509.1239	7.88	C ₃₀ H ₂₁ O ₈	-0.611	20.5	491.1131 (6), 254.0584 (100), MS ³ [491] → 491.3166 (37), 473.4017 (38), 447.4200 (100), 421.2768 (35) MS ³ [254] → 254.0586 (85), 226.0638 (100)	emodin dianthrone (24) ^b	LF
511.3/543	511.3428	9.06	C ₃₂ H ₄₇ O ₅	-0.172	9.5	436.3301 (11), 435.3272 (100), 377.2851 (36)	n.a.	F
517.1/527	517.0559	8.75	C ₃₀ H ₁₃ O ₉	-0.655	24.5	517.0566 (33), 499.0459 (100), 489.0616 (7), 487.0461 (4), 473.0668 (4), 475.0444 (2), 455.0559 (2) MS ² [499] → 499.0457 (100), 455.0560 (6)	35	LF

Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	RDB	Fragment ions (Intensity)	Identification	Ds
519.1/467	519.0718	7.75	C ₃₀ H ₁₅ O ₉	-0.8	23.5	519.0722 (100), 503.0408 (7), 487.0458 (13) *MS ² [519]→ 519.0722 (100), 503.0410 (22), 487.0460 (44), 475.0456 (13)	pseudohypericin (4) ^b	LF
519.1/493	519.0713	8.20	C ₃₀ H ₁₅ O ₉	-0.845	23.5	519.0707 (20), 503.0392 (100), 487.0466 (1)	n.a.	LF
521.1/459	521.0876	7.65	C ₃₀ H ₁₇ O ₉	-0.432	22.5	*MS ² [521]→ 521.0878 (100), 503.0776 (2), 493.0927 (4), 489.0616 (2), 479.0805 (11), 477.0980 (37), 449.1040 (1),	protopseudo-hypericin (27) ^b	LF
525.1/424	525.1191	7.05	C ₃₀ H ₂₁ O ₉	-0.055	20.5	507.1083 (8), 270.0534 (100), 254.0586 (21) MS ³ [507]→ 507.1082 (18), 489.0976 (100), 479.1145 (9), 461.1036 (3) MS ³ [270]→ 270.0532 (100), 242.0582 (85), 241.0505 (87), 214.0635 (3) MS ³ [507]→ 254.0586 (94), 226.0636 (100)	hydroxyemodin dianthrone (26) ^b	LF
535.1/420	535.0666	6.99	C ₃₀ H ₁₅ O ₁₀	-0.822	23.5	535.0663 (100), 517.0558 (5), 503.0403 (10), 475.0452 (7) *MS ³ [503]→ 503.0402 (100), 461.0314 (15), 459.0491 (31)	36	F
535.1/435	535.0672	7.21	C ₃₀ H ₁₅ O ₁₀	0.318	23.5	535.0664 (55), 519.0366 (30), 518.0632 (95), 517.0563 (21), 504.0437 (28), 503.0398 (100), 493.0562 (81), 489.0609 (30), 290.2121 (7)	n.a.	L
547.1/408	547.0685	6.80	C ₃₁ H ₁₅ O ₁₀	2.578	24.5	529.0568 (19), 515.2674 (4), 503.0769 (100)	oxyskyrin ^b	F
553.1/431	553.0787	7.14	C ₃₀ H ₁₇ O ₁₁	1.999	22.5	553.0775 (13), 535.0666 (20), 536.0730 (21), 525.0817 (11), 509.0864 (100), 485.0865 (79), 465.0964 (13)	n.a.	F
569.1/372	569.07251	6.15	C ₃₀ H ₁₇ O ₁₂	-0.069	22.5	569.0709 (11), 552.0684 (21), 551.0614 (21), 541.0795 (11), 525.0817 (100), 501.0820 (87), 481.0926 (11)	n.a.	F
569.4/543	567.3696	9.05	C ₃₅ H ₅₁ O ₆	0.824	10.5	471.3116 (100), 497.3272 (13), 453.3007 (15), 415.3218 (12)	n.a.	F
669.1/463	669.1246	7.7	C ₃₅ H ₂₅ O ₁₄	-0.521	23.5	519.0720 (100)	S-skyrin-6-O-β-arabinofuranoside (30) ^a	L
699.1/459	699.1350	7.63	C ₃₆ H ₂₇ O ₁₅	-0.834	23.5	519.0727 (100) *MS ³ [519]→ 519.0718 (100), 491.0771 (25), 475.0820 (3), 463.0823 (1), 447.0871 (4)	S-skyrin-6-O-β-glucopyranoside (29) ^a	LF

^a identified with standard compound

^b identified based on MS/MS

* fragmentation energy ncd=45

*Naphthodianthrone*s

The dianthronic structure is reached after the dimerization of two anthraquinone units. After the initial C-C bond, the formation of one further C-C bond generates protohypericin (**18**) and protopseudohypericin (**27**). Finally, the last ring closure between the units takes place and the endproducts hypericin (**3**) and pseudohypericin (**4**) are obtained (Figure 3.4). These four compounds, summarized in the following discussion as hypericins, highly correlated with each other in flower and leaf data. For protohypericin (**18**), hypericin (**3**), and pseudohypericin (**4**), this

was shown by several studies.^{39,153} All hypericins are localized in the dark glands as evidenced by MS-imaging techniques.^{96,117}

In addition to the hypericins, features 517.1/527 and 535.1/420 hint, based on their chemical composition, to a close structural relation. Further the features 509.1/474 and 525.1/424 correlated which could correspond to emodin dianthrone (**24**) or penicilliposin (**23**) and hydroxyemodin dianthrone (**26**) and hydroxyenicilliposin (**25**). MS/MS experiments were performed and compared to the fragmentation pattern of the known hypericins to gain structural information of the analogs (Table 4.3).

Table 4.3 MS² fragments of naphthodianthrone and unidentified related features with relative intensities colored in yellow.

Fragments	Compound / Feature (nce)							
	hypericin (3) (nce=45)	proto- hypericin (18) (nce=45)	pseudo- hypericin (4) (nce=45)	Proto- pseudo- hypericin (27) (nce=45)	517.1/527 (35) (nce=35)	535.1/420 (36) (nce=35)	525.1/424 (nce=35)	509.1/474 (nce=35)
[M-H] ⁺	503 (100)	505 (100)	519 (100)	521 (100)	517 (32)	535 (100)	525 (0)	509 (0)
[M-H-H ₂ O] ⁺	-	-	501 (10)	503 (2)	499 (100)	517 (12)	507 (8)	491 (6)
[M-H-CO] ⁺	475 (1)	477 (6)	491 (7)	493 (4)	489 (5)	-	-	-
[M-H-2CO] ⁺	447 (2)	449 (1)	-	465 (1)	-	-	-	-
[M-H-CO ₂] ⁺	459 (33)	461 (40)	475 (14)	477 (75)	473 (3)	491 (3)	-	-
[M-H-CO ₂ -H ₂ O] ⁺	-	-	-	-	455 (1)	-	-	-
[M-H-CO-H ₂ O] ⁺	-	-	-	-	-	489 (5)	-	-
[M-H-CO ₂ -CO] ⁺	431 (3)	433 (2)	-	449 (2)	-	-	-	-
[M-H-CH ₃ -OH] ⁺	-	-	487 (44)	489 (3)	-	503 (45)	-	-
[M-H-CH ₂ CO] ⁺	461 (12)	-	477 (5)	479 (12)	475 (1)	-	-	-
[M-H-CH ₂ CO-CO] ⁺	433 (3)	-	-	-	-	-	-	-
[M-H-CH ₂ CO-2CO] ⁺	405 (0.3)	407 (2)	-	423 (1)	-	-	-	-
[M-H-CH ₄] ⁺	487 (7)	489 (6)	503 (23)	505 (1)	-	-	-	-
[M-H-CH ₃] ⁺	-	-	-	-	502 (2)	-	-	-
[M-H-CO ₂ -CH ₃] ⁺	444 (1)	-	-	-	-	-	-	-
[M-H-CH ₂ O] ⁺	-	-	489 (5)	491 (2)	487 (4)	505 (11)	-	-
[M-H-C ₃ H ₆] ⁺	-	-	477 (8)	-	-	-	-	-
[M-H-C ₄ H ₅ O ₂] ⁺	-	-	-	-	-	450 (3)	-	-
[M-H-2H ₂ O] ⁺	-	-	-	-	-	499 (3)	-	-
[M-H-2CH ₂ O] ⁺	-	-	-	-	-	475 -	-	-
[M-H-CHO] ⁺	-	-	-	-	-	506 (20)	-	-
[M-H-C ₁₅ H ₉ O ₄] ⁺	-	-	-	-	-	-	270 (100)	-
[M-H-C ₁₅ H ₉ O ₅] ⁺	-	-	-	-	-	-	254 (22)	-
[M-H-C ₁₅ H ₁₁ O ₄] ⁺	-	-	-	-	-	-	-	254 (100)

The fragmentation of the four known hypericins **3**, **4**, **18**, **27** only could be realized with high collision-induced dissociation energies. Although high normalized collision energy (nce = 45) was applied, the fragmentation rate of the molecular ion peak is low, which shows the high stability of the polyaromatic compounds. Additionally, the eliminations that occur are unspecific. Nevertheless, Hecka et al. (2009) could show dependencies between the fragmentation pathway and the nature of substituents.¹⁸¹ To reach a comparable grade of fragmentation of the four correlating new features, a nce of 35 was sufficient.

Feature 517.1/527 corresponds to a molecular ion with the chemical composition $C_{30}H_{13}O_9^-$, which is consistent with pseudohypericin bearing an additional bond. Two possible structures are shown in the Figure 4.11. On the one hand, the loss of two protons from the core structure could lead to a highly conjugated system with two keto groups instead of two hydroxy groups. On the other hand, a structural difference could be at the side chains at position 11 and 11'. It could be realized by an aldehyde at position 11 or a ring closure between the side chains. However, the molecular change of feature 517.1/527 results in the loss of stability of the molecular ion, since the base peak m/z 499 is already formed with the nce of 35. The preferred H_2O loss indicates a change in the hydroxymethyl group. Furthermore, the compound shows the cleavages also observed for pseudohypericins (**4**). However, a striking difference is the formation of the radical m/z 502 after the cleavage of a CH_3 group, whereas the loss of methane was observed for all hypericins. This leads to the conclusion that the methyl group at position 11' is highly influenced, but based on this data the structure can not be determined. For one of the suggested structures the fragmentation pathway is shown in Figure 4.11. It should be noted that for all naphthodianthrones several mesomeric structures exist due to the keto-enol tautomerism.

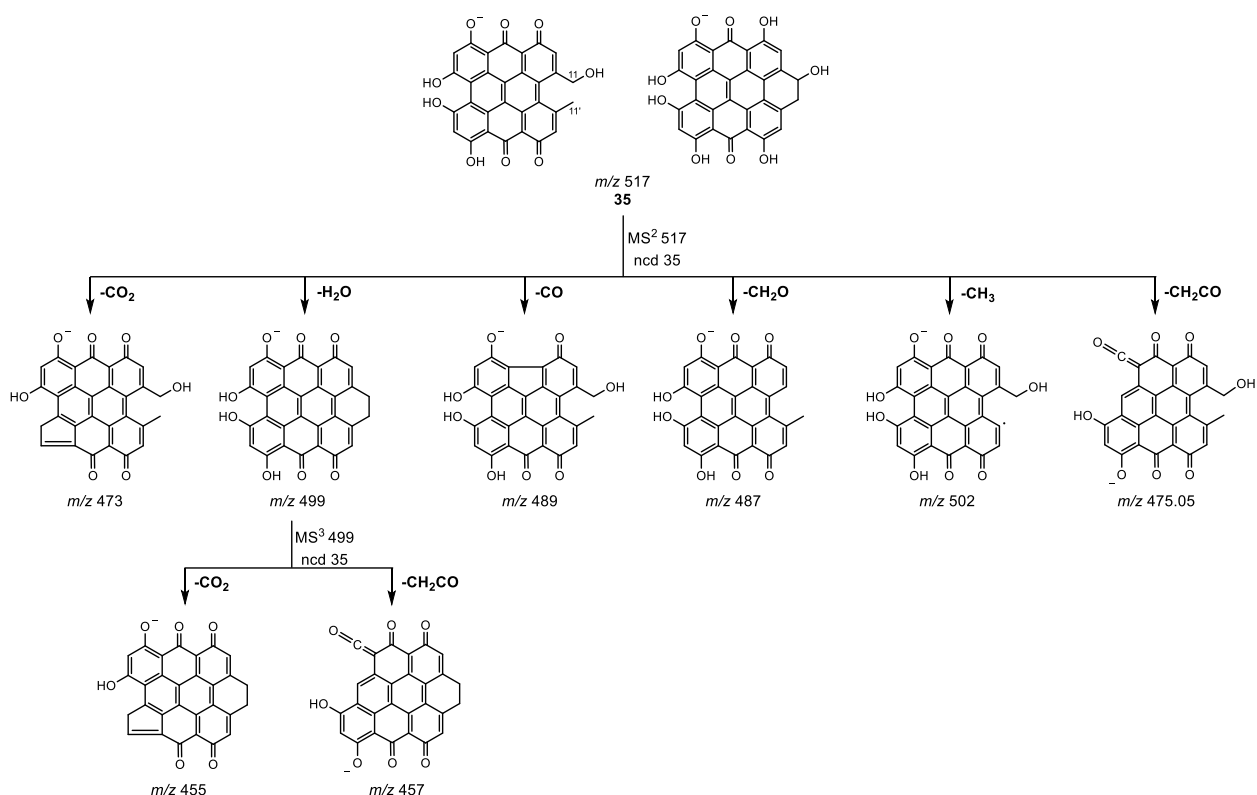


Figure 4.11 Proposed structures for the MS^2 fragmentation of the ion m/z 517 and MS^3 fragmentation of m/z 499.

The fragmentation behavior of feature 535.1/420 (Figure 4.12) is similar to pseudohypericins **4**, **27**, and is in accordance with the proposed structure **36**. The losses of H_2O and CH_2O , which could also be observed in pseudohypericins, suggests hydroxylation at position 11. Compound with m/z 535 shows cleavages of two H_2O and two CH_2O , pointing to a second OH group at position 11'. The loss of $C_4H_5O_2$ is the only radical cleavage observed for this molecule. A possible fragmentation product with m/z 450 is shown in Figure 4.12. If the system is conjugated in this

way, the radical can be well stabilized, and various mesomeric structures are possible. However, the presence of a methyl group instead of the CH_2OH group at position 11 or 11', would not allow this so easily. Therefore the $\text{C}_4\text{H}_5\text{O}_2$ loss cannot be observed in pseudohypericins **4** and **27**. Fragment 503 ($\text{C}_{29}\text{H}_{11}\text{O}_9^-$) is formed after the loss of CH_3OH , known from pseudohypericin (**4**). The loss of the side chain favors a ring closure potentially resulting in a five-membered ring or a six-membered ring with integrated heteroatom. Since Piperopoulos et al. (1997) already described the formation of the five-membered ring moiety for hypericin and protohypericin after CH_3 and for pseudohypericins after CH_3OH -cleavage, the five-ring substructure is considered probable.¹⁸² The MS^3 fragmentation of ion 503 showed losses of CO_2 and CH_2CO , also typical for hypericins. The absence of the $\text{C}_4\text{H}_5\text{O}_2$ loss in MS^3 of m/z 503 indicate a stabilization of the side chains 11 and 11' like through a ring closure. The elimination of 29 Da corresponds to the loss of CHO . This splitting cannot yet be explained. The two correlating features 535.1/420 and 517.1/527 could represent two side products which structures are putatively assigned by MS/MS fragmentation.

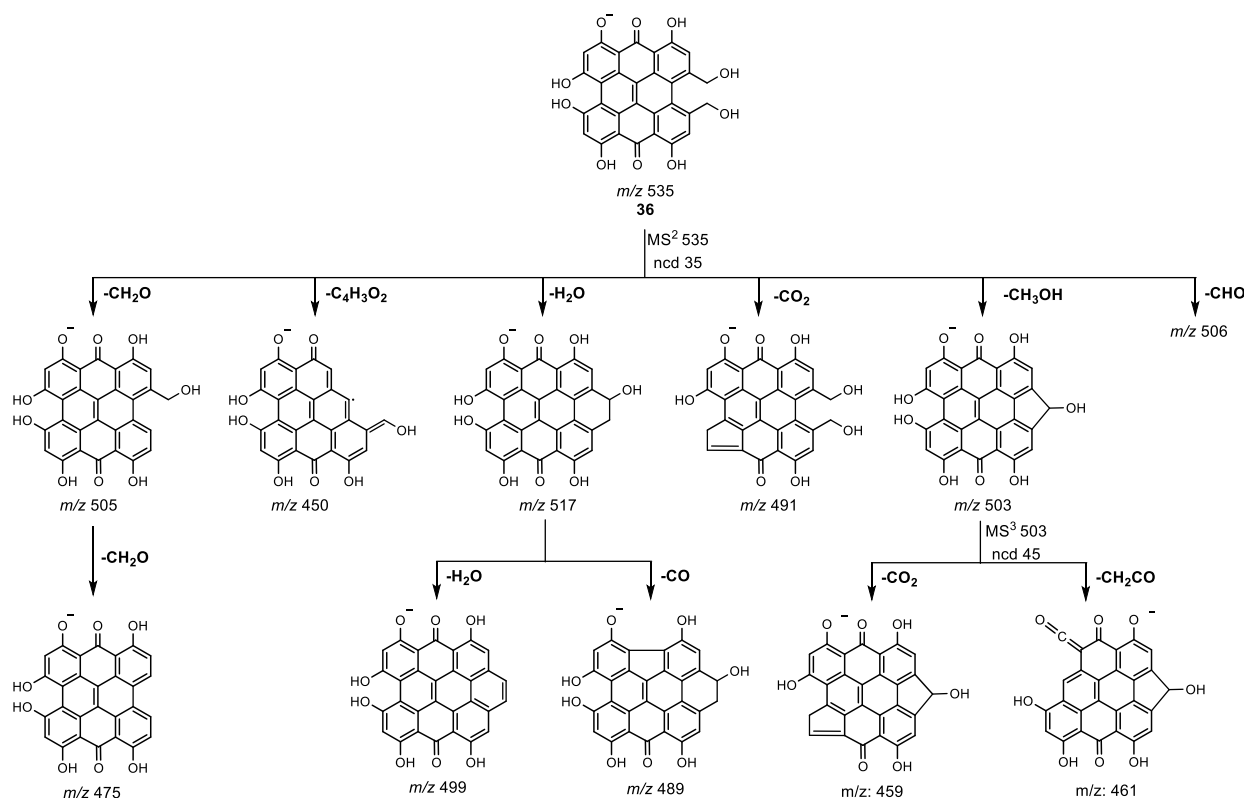


Figure 4.12 Proposed structures for the MS^2 fragmentation of the ion m/z 535 (**36**) and MS^3 fragmentation of m/z 503.

The fragmentation of the two features 525.1/424 ($\text{C}_{30}\text{H}_{21}\text{O}_9^-$) and 509.1/474 ($\text{C}_{30}\text{H}_{21}\text{O}_8^-$) differs significantly from hypericins. Instead of unspecific neutral losses, these two structures split into subunits (Figure 4.13). Ion m/z 509 shows only the fragment m/z 254 corresponding to the equal halves of the molecule. Contrary, ion m/z 525 forms m/z 254 and as base peak m/z 270. The fragment m/z 254 of both compounds behave in MS^3 experiments in the same manner, suggesting the occurrence of identical subunits. The $[\text{M}-\text{H}]^-$ ion at m/z 525 contains one oxygen more than m/z 509, comparable to pseudohypericin (**4**) and hypericin (**3**). So it is assumed that feature

525.1/424 has a hydroxymethyl group at position 11. This additional oxygen is also reflected by the formation of subunit fragment at m/z 270. According to the molecular formula, the subunits of both compounds are connected by one single bond. The question where the first bond between the subunits is formed during hypericin biosynthesis is still open. These two intermediates and their fragmentation behavior could answer this question. The easy cleavage of subunits points out to a weak bond between the halves. The bond between 5 and 5' is formed by two sp^2 -hybridized carbons (-1288 kJ cal.MOE), whereas the bond between 10-10' is between sp^3 -carbons (-1273 kJ cal.MOE), which leads to different bond strength. The bond energy between sp^2 -carbons is higher because of the additional π -character in comparison to sp^3 -hybridized carbon bonds with higher σ -character. The easy cleavage of both halves indicates the lower energy bond at position 10-10'. The fragmentation behavior is comparable to this of sennosides in rhubarb, which disintegrate into their subunits after C10-C10' (sp^3) cleavage.¹⁸⁰ In contrast to that, skyrin, which subunits are connected at position C5-C5', shows more hypericin like neutral losses.

In MS^2 spectra of features 525.1/424 and 509.1/474 a water loss is observed (Figure 4.13). If these ions are further fragmented in MS^3 , water and CO are eliminated. Because these cleavages take place independently of the side chain, the water loss can be explained by forming an additional bond between the subunits. Furthermore, the dehydrated products in the MS^3 experiments do not fragment into the subunits anymore, which shows that the structures are stabilized. In fact, they split off another water unit so that a further bond is formed between the subunits.

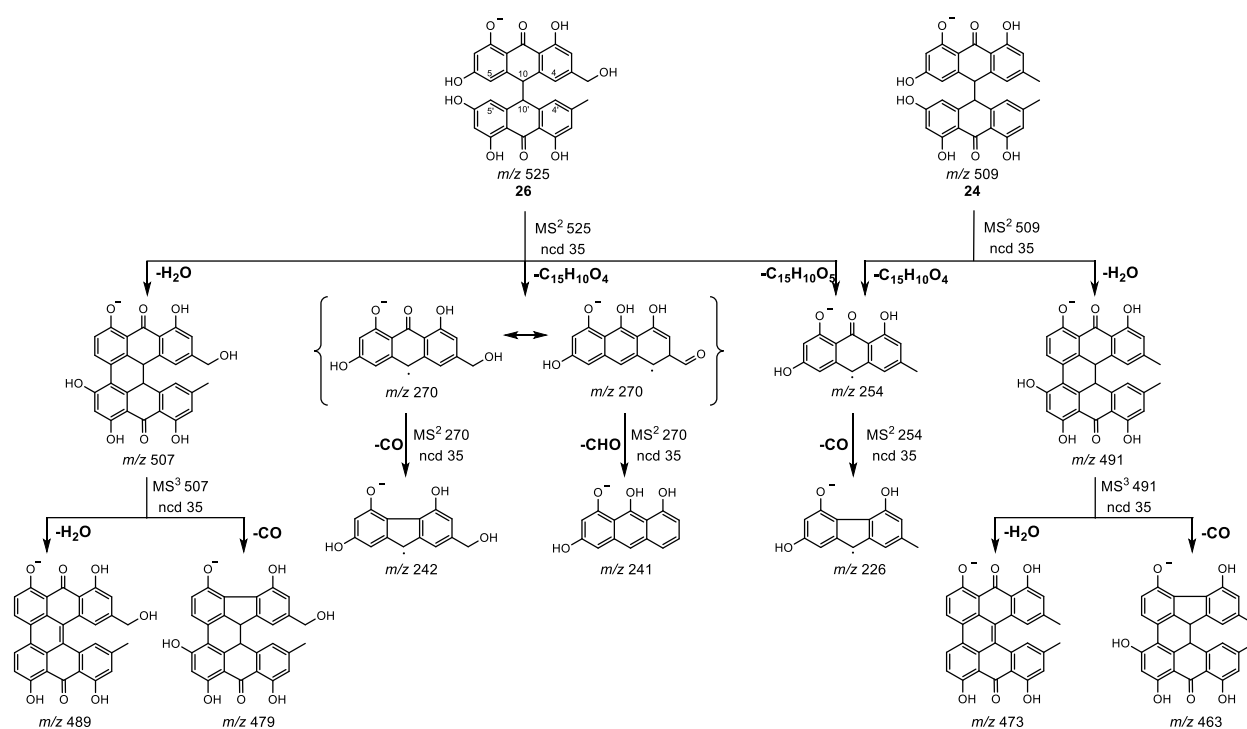


Figure 4.13 Proposed structures for the MS^2 fragmentation of the ion m/z 525 (**26**) and m/z 509 (**24**) (keto-form) and their MS^3 fragmentation after H_2O loss.

The results obtained by MS-fragmentation studies for these two features suggest the structures emodin dianthrone (**24**) and hydroxyemodin dianthrone (**26**) as possible precursors of the protohypericins. Furthermore, the cleavage in two monomeric subunits makes an initial connection between C10 and C10' of the subunits more likely. These findings would indicate that skyrin or skyrin derivatives are not directly involved in the hypericin formation, as recently discussed.^{67,101,183} However, definitive proof cannot be obtained from MS-data only.

Bisanthrones

The correlating features 699.1/459 (**29**) and 669.1/463 (**30**) were assigned as already described skyrin-glycosides.^{52,67,101,118,184} For structure elucidation, the compounds were isolated and identified via NMR and CD spectroscopy. The findings are in line with the results of Wirz et al. (2000)¹¹⁸ so that the features were identified as *S*-skyrin-6-*O*- β -glucopyranoside (**29**) and *S*-skyrin-6-*O*- β -arabinofuranoside (**30**). The correlation to the dark gland occurrence (chapter 3.3.2) raised more interest in their role in the hypericin biosynthesis.^{67,101} Kimakova et al. (2018)¹⁰¹ and Revuru et al. (2020)¹⁸³ suggested a possible pathway where skyrin serves as direct intermediate to the protohypericins. However, our MS/MS study indicates the ions **24** and **26** may be direct precursors of the protohypericins **4**, **27** and their fragmentation behavior better fits a C10-C10' connection, though a C5-C5' cannot be ruled out either. Nevertheless, based on this evidence, skyrin derivatives might be just minor compounds or byproducts co-located in the dark glands.

Feature 553.1/431 corresponds to the $[M-H]^-$ $C_{30}H_{17}O_{11}^-$, which is in accordance with the molecular formula of oxyskyrin, described from *H. perforatum*.¹⁸⁴ Whether some of the unidentified features, like 535.1/420, 553.1/431, 569.1/372, are also dianthrones remains to be clarified.

Structurally non-related compounds

The correlation analysis was used to find hypericin (**3**) biosynthesis-related compounds. But also structurally non related compounds correlate to hypericins (Table 4.2). In particular, flavonoids like quercetin (**5**), quercitrin (**8**), and kaempferol (285/366) are related. The positive correlations between the amounts of quercitrin (**8**) and emodin (**20**), hypericin (**3**) and pseudohypericin (**4**) were previously reported.⁷⁷ In our previous analysis of the pistil tissue (chapter 3.3.2), the dark glands occurrence is associated with the amount of quercetin (**5**) and quercitrin (**8**).⁶⁷ The investigation of the leaf and flower data confirm these findings. In addition, the unidentified features 435.1/308, 419.1/267, 463.1/269 and 431.1/338, correlate to the hypericin content. They are suggested to be flavonoid glycosides due to their over-proportionally high oxygen content in the molecular composition and the comparatively early retention time, indicating a high polarity.

It was reported that flavonoids, e.g. quercetin (**5**) are located in the dark glands like hypericin (**3**) in the *H. perforatum* leaf.^{96,120} Hypericins are phototoxic, and it is believed that they defend the plant against predators.^{20,22,185} Insects not specialized in feeding *H. perforatum* avoid the dark gland tissue.²⁰ The light-activated compounds generate reactive oxygen species, which initiate cell

death after intake.¹⁸⁵ On the one hand, it can be assumed that the plant protects itself by compartmentalization of these substances in the dark glands.⁵⁵ And on the other hand, the accumulation of antioxidative substances, such as **5** and **8**, could be another survival strategy to prevent ROS induced damage.^{47,120} Further, a better solubility of hypericin was reported in the presence of flavonoids,¹⁷¹ therefore possibly their co-location is necessary to avoid naphthodianthrone precipitation and an associated loss of function.

4.3.5.2 Hyperforin correlated metabolites

The percentual composition of major phloroglucinols reveals differences between the GTs and between the organs (Appendix 4.7). In leaf data, GTs 3, 21, and 29 are characterized by a high percentage of feature 331.1/487 (**17**). This compound is known so far from *H. punctatum*⁶⁵ and was also detected in our previous investigations in chapter 2. Interestingly, **17** is almost absent in the corresponding flower data, but instead they contain in contrast to other GTs, a high proportion of features 413.2/508 and 413.3/476 based on the MS/MS experiments tentatively assigned as maculatoquinone (**37**) and deprenyl hyperpolyphyllirin (**38**). Besides, GT 3 contains as major compound hyperpolyphyllirin (**39**). Compound **39** was first identified by Porzel et al. (2014) from *H. polyphyllum*.⁵²

A higher proportion of hyperforin (**40**) characterizes nine GTs (45, 51, 53, 61, 76, 85, 90, 94, 97). Interestingly, this trend is comparable for flower data.

The correlation network in Figure 4.10 shows that phloroglucinols highly correlate with each other for leaf data, whereas for flower data, two separate groups were observed. This was visualized with correlation heatmaps (Figure 4.14) of the major phloroglucinols characterized in Table 4.1. A positive correlation between hyperforin and other phloroglucinols was observed in leaf data except for compounds **15**, **17**, **37**, **38**, **39**, and **40** where no relation exists. In contrast to that, the flower data heatmap (Figure 4.14b) is divided into two groups. In the upper left part of Figure 4.14b, the correlation is displayed between hyperforin (**1**), adhyperforin (**2**), and possible degradation products furohyperforin (**32**), P59, and P56. In the lower right part, the less prenylated phloroglucinols correlate with each other.

To further investigate this behavior, the correlation heatmap of the selected phloroglucinols was extended to the whole flower feature table. The unidentified correlating features were evaluated, and, if possible, their fragmentation behavior was analyzed. Promising features with a positive correlation to hyperforin are shown in Appendix 4.8. Features that correlate to the less prenylated phloroglucinols are summarized in Appendix 4.9. Correlating features belong mainly to the PPAPs. Exceptions are quercitrin (**8**) which correlate to the hyperforin (**1**) content and astilbin (**19**), which is related to the less prenylated PPAPs. PPAPs ionizing well, having structural similarities. They often differ only in the number and kind of prenylations. Therefore mass losses during MS² fragmentation were evaluated to determine structural similarities between the correlating compounds (Appendix 4.10).

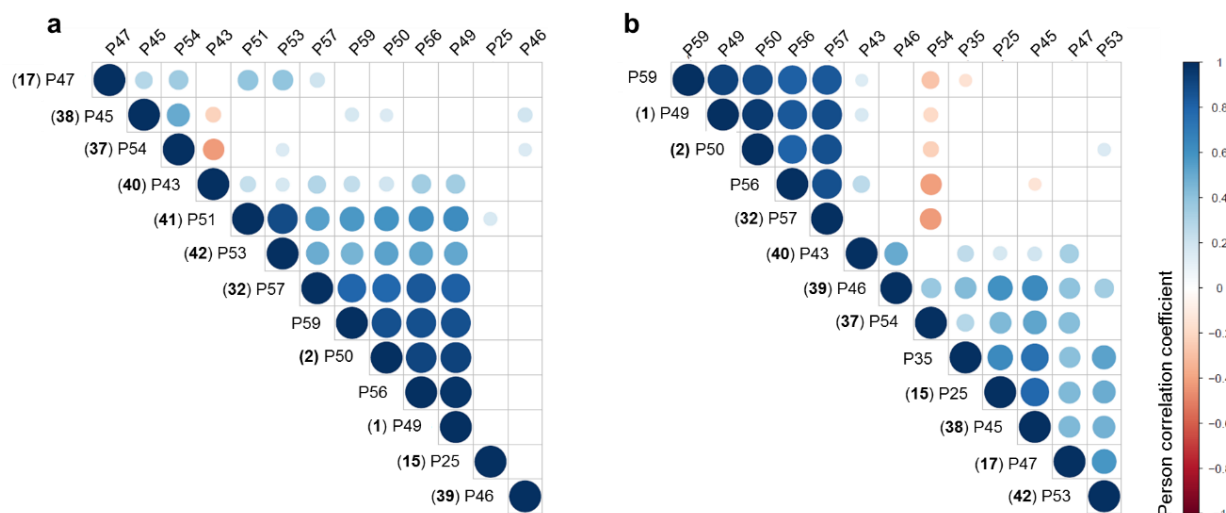


Figure 4.14 Correlation heatmap (Pearson) of a) leaf data and b) flower data ($p < 0.01$) with major polyprenylated acylphloroglucinols: hyperforin (**1**), adhyperforin (**2**), dimethylallyl-phlorisobutyrophenone (**15**), 3-geranyl-methylpropanoyl-phloroglucinol (**17**), furohyperforin (**32**), maculatoquinone (**37**), deprenylhyperpolyphyllirin (**38**), hyperpolyphyllirin (**39**), hyperfirin (**40**), 2-*O*-geranyl-methylpropanoyl-phloroglucinol (**41**), 2-*O*-geranyl-1-methylbutanoyl-phloroglucinol (**42**), 331.2/425 (P35), 567.4/543 (P59), 485.3/522 (P56).

Thus, the hyperforin (**1**) correlated features 551.4/557, 565.4/563, 567.4/470, 567.4/525 were putatively assigned as furohyperforin isomers and the corresponding hydroperoxides, described by Fuzatti et al. (2001)¹⁴¹ (Figure 4.15). Characteristic fragmentations of these compounds are the losses of C_6H_8O , CO, and CO_2 . Those furohyperforin derivatives were detected after degradation of hyperforin (**1**), and adhyperforin (**2**),^{140-142,186} which why a strong correlation to hyperforin (**1**) is as expected. Exemplary structures of numerous isomers are shown in Figure 4.15.

The highly prenylated compounds hyperforin (**1**), adhyperforin (**2**), furohyperforin isomer I (**32**), and the feature 549.4/562 share the loss of one (C_5H_9) and two prenylated side chains ($C_{10}H_{18}$). More specific is the cleavage of C_6H_{11} , which may correspond to a fragmentation of the geranyl side chain after ring closure. Besides, these hyperforin related compounds show the loss of $C_{11}H_{20}$, which may result from the subsequent losses of C_5H_9 and C_6H_{11} .

Features correlating with less prenylated compounds are often characterized by the loss of an entire geranyl side chain. In case of a C-linkage, C_9H_{16} is cleaved off. If the geranyl sidechain is connected via oxygen, $C_{10}H_{17}$ is removed. This is the case for features 331.2/425, 331.2/500, and 345.2/507. In addition, feature 331.2/500 and 345.2/507 show a CO_2 loss. This cleavage can occur if two meta substituted hydroxy groups exist and the common ortho position does not bear a side chain.¹⁸⁷ Therefore, 331.2/500 is tentatively assigned as 2-*O*-geranyl-methylpropanoyl-phloroglucinol (**41**) and 345.2/507 as 2-*O*-geranyl-methylbutanoyl-phloroglucinol (**42**). The feature 331.2/425, which does not show the CO_2 loss, could therefore be the 4-*O*-geranyl-methylpropanoyl isomer (**43**). The cleavage of C_4H_6O corresponds to the loss of isobuteryl- and C_5H_8O to 2-methylbuteryl (“ad”-forms) and is especially visible in the smaller compounds.

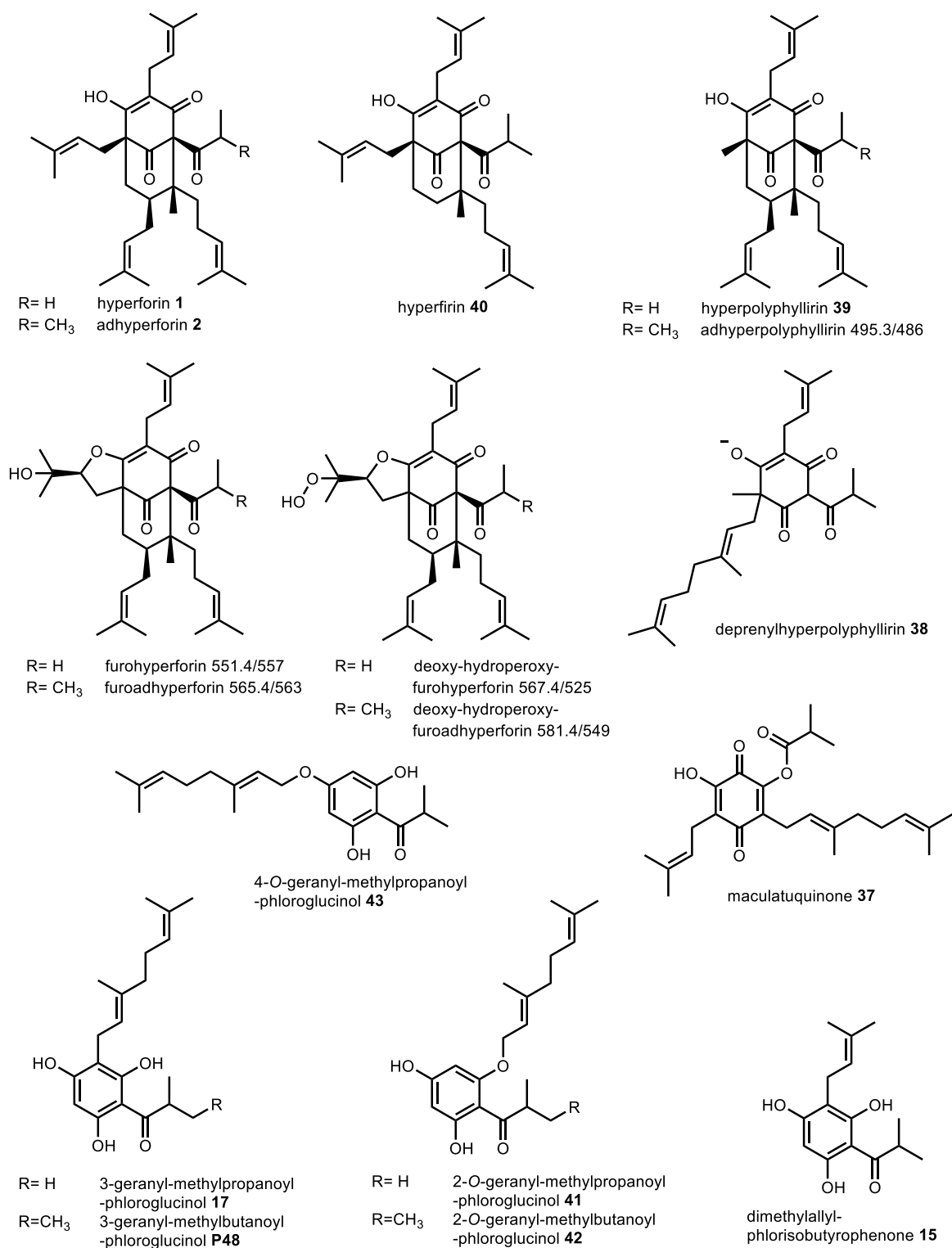


Figure 4.15 Structures of tentatively assigned polyprenylated acylphloroglucinols in *H. perforatum*.

4.3.5.3 Correlation behavior of polyphenols

H. perforatum is rich in polyphenols. In our data set, 15 flavonoids and two organic acids could be tentatively assigned. The flavonoid composition of flower and leaf data is displayed in Figure 4.16. Despite the occurrence of GTs with reduced rutin content, the variation within the organs is

low. However, when comparing the composition of flowers and leaves, significant differences become apparent. Astilbin (**19**) and the acetylated quercetin-glycosides (P12, P13, P23, P24) are abundant in leaves, whereas their amount in flowers is negligible. In contrast, the percentages of quercetin (**5**), miquelianin (**50**, Figure 6.7), and biapigenin (**33**) are much higher in flowers.



Figure 4.16 Percentual mean intensity of major flavonoids per GT in a) leaf data and b) flower data.

In the correlation networks (Figure 4.10), just one small group with strong connections within the polyphenols was observed. The remaining polyphenols were spread over the whole network and are sometimes correlated to other compound classes (chapters 4.3.5.1, 4.3.5.2.). For better visualization, a correlation heatmap of flavonoids was created, which showed comparable results for flower (not shown) and leaf data (Figure 4.17). The organ independent correlation was identified between procyanidin (P5) and its putative precursors catechin (**14**) and epicatechin (**14a**). Due to their biosynthetic relation a correlation is expected.¹⁸⁸

Further correlations were observed between acetylated quercetin glycosides (P12, P13, P23, P24). The presence of the two acetyl quercetin-3-*O*-hexosides P12 and P23, with different retention times, suggest the acetylation at different positions of the sugar. Each acetyl quercetin-3-*O*-hexoside correlates with one acetyl rutin (P12/P13 and P23/P24). This may indicate that the acetyl-quercetin-3-*O*-hexosides contain acetylated glucose, and after rhamnosyl transfer, they give acetyl-rutin. The non-acetylated compounds **7/9** and **6** do not correlate to their corresponding acetylated forms. It can be assumed that the acetylation occurs enzymatically. To confirm this, the combination with transcriptomic data of our project partner can help to identify the responsible

enzymes in the future. Although the abundance of acetylated quercetin glycosides was previously described,^{18,171} little is known about these compounds and their biological function.

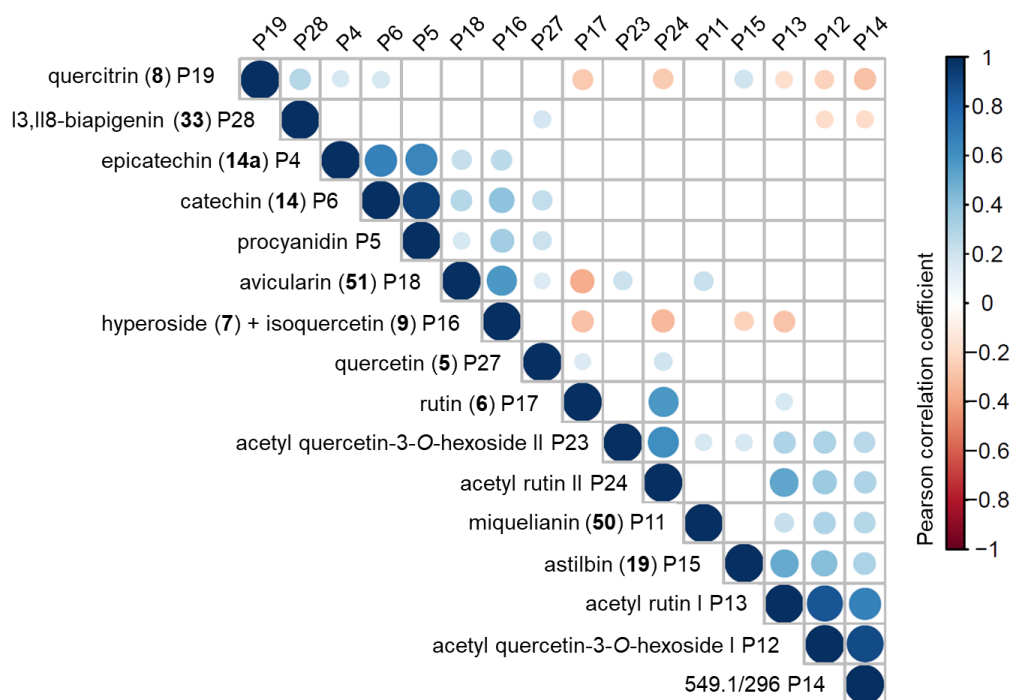


Figure 4.17 Correlation heatmap of major flavonoids (peak numbers according to Table 4.1) ($p < 0.01$).

4.4 Conclusion

The intraspecific variance of *H. perforatum* GTs from North America and Europe was analyzed with UHPLC-MS in an untargeted fashion. Main differences were observed in the content of flavonoids and polyprenylated phlorogolucinols. Especially acetylated flavonoids are varying between the GTs. More than 20% of the GT were not in line with the European pharmacopoeia because of the absence of hyperforin or rutin. This could lead to changed pharmaceutical effects of extracts derived from different GTs, and are possibly the reason for conflicting bioactivity study results. Both anomalies were found in leaf and flower material. The naphthodianthrone composition was similar for all GTs and comparable in leaves and flowers, whereas polyprenylated phloroglucinol and flavonoid abundancies were more organ specific. The metabolic result reflects the genetic diversity of the sampling sites.⁶⁴

Comprehensive correlation analysis to hypericin confirmed the relation to the biosynthetically connected compounds and skyrin derivatives. Furthermore, extensive MS/MS analysis suggested two new byproducts (**35**, **36**) of the hypericin biosynthesis and indicate that the identified precursors are emodin dianthrone. By combination of the metabolite profiles with transcriptomic analysis done by our project partners (IPK Gatersleben), the elucidation of gene-metabolite relationships will be possible. It offers a good opportunity to get insights into the not yet fully understood hypericin (**3**) and hyperforin (**1**) biosynthesis.

5 Chemical comparison of two leaf phenotypes

Abstract

Hypericum perforatum is one of the best known medicinal plants in the world, but the production of comparable extracts is still challenging. In this study, it is shown how intense the genetic variation influences the phenotype and the chemical profiles of the plant. A gland counting tool for fast screening of the leaf phenotypes was developed, and the differences in the chemical constitutions were characterized by UHPLC-ESI-MS. Significant changes can be found in all compound classes, especially the pharmaceutical important constituents hypericin and hyperforin are affected. Another highly different compound was isolated and identified as 1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid. It is the first sulfonated xanthone found in *H. perforatum*.

5.1 Introduction

Hypericum perforatum L., commonly known as St. John's wort, is a widely distributed herbaceous plant used as traditional medicine. It is applied against mild to moderate depression but also antiviral, antimicrobial, and wound healing activities are reported.¹² Especially the neurological effects increased the attention throughout the world, to analyze the responsible constituents.¹⁸⁹ The species contains a large variety of secondary metabolites including mainly flavonoids, naphthodianthrones, phloroglucinols, and xanthenes.^{18,52,53} Genetic infidelity is the major cause of the chemical inconsistency but also the abiotic and biotic environment plays a role.^{64,190} These variations highly influence the success of plant-derived medicines, particularly full plant preparations, which are usually used for St. John's Wort applications.

This study's objective is the investigation of one *H. perforatum* accession, where after asexual proliferation, two phenotypes were distinguishable by their leaf size and form (bL= broad leaf, nL= narrow leaf). The leaf size, as well as the number of dark and translucent glands was determined, and the chemical profiles were analyzed with UHPLC-ESI-MS. The differential compounds were assigned.

5.2 Experimental

5.2.1 Plant material

Cuttings of the *Hypericum perforatum* genotype HyPR-01 were provided by the IPK Gatersleben and cultivated at IPB in the greenhouse, field, and phytochamber (long day conditions: 16 h light, 400 $\mu\text{mol m}^{-2}$, 70% humidity) at 21°C. The cuttings developed two different phenotypes: broader oval leaves (bL) and small narrow leaves (nL).

5.2.2 Leaf phenotyping

The characterization of the leaf phenotypes was performed with three plants per phenotype grown in the phytochamber. Each replicate, represented by ten adult leaves, was immediately photographed after harvest using a Nikon SMZ1000 light-microscope with an up and down light source. ImageJ (Version 1.51j8) was used for processing. In the first step, the picture was separated into three grayscale channels. The blue channel image was binarized with a thresholding method by Huang et al. (1995) to calculate the leaf size with the Shape Filter tool and the feature `getContainedPoints()`.¹⁹¹ A gland counting tool was developed using the green channel grayscale image, which has the highest contrast of the secondary leaf structures to the mesophyll. For the dark glands, the minimal and for the translucent glands, the maximal pixel value indicated a hit. By applying a circular filter, pixels belonging to the same gland were combined.

5.2.3 Extraction and sample preparation

Leaves were harvested from all cultivation sites in 2017. For the analysis of different plant parts, bL plants of the field were separated in different organs before freezing in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$. After lyophilization plant material was pulverized with a cooled mixer mill (MM400, Retsch GmbH, Germany) applying a 30 Hz frequency for 30 seconds. After repeated drying, the plant powders were stored in darkness at room temperature. For UHPLC-MS analysis, the plant material (bL genotype organs: 1.0 mg ml^{-1} ; leaf samples of different cultivation sides: 1.5 mg ml^{-1}) was extracted with LC-MS grade methanol containing $8\text{ }\mu\text{g ml}^{-1}$ umbelliferone (HPLC grade Sigma) as an internal standard by brief mixing on a vortex followed by ultrasonication for 15 min. Following centrifugation (10 min, $14,000\text{ min}^{-1}$) the supernatant was applied to UHPLC-MS/PDA.

5.2.4 High-resolution UHPLC-MS analysis and data processing

Negative ion high-resolution ESI mass spectra were obtained from an Orbitrap Elite mass spectrometer with the in chapter 2.2.7 described method. The data were evaluated by the Xcalibur software 2.2 (Thermo Fisher) and processed in R with XCMS package (version 1.52.0, bioconductor.org/). The raw output files (*.raw) were converted into standard format mzData files (*.mzML) utilizing proteowizard (proteowizard.sourceforge.net/). Peak picking was performed in XCMS with centWave parameters: ppm = 10, peakwidth = c(5,20), snthr = 10, and prefilter = c(3,1000). After peak grouping (minfrac = 1, bw = 5, mzwid = 0.2), retention time correction was performed using LOESS correction, and peak grouping was repeated. Missing values were filled with fillPeaks function. The extraction of the most significant different features (m/z retention time pair) was done with the diffreport() function. The identification of known compounds was based on the exact mass of detected ions and fragmentation pattern.

5.2.5 NMR analysis

The NMR spectra were recorded on a Varian VNMR5 600 system at $+25\text{ }^{\circ}\text{C}$ operating at a proton NMR frequency of 599.83 MHz using a 5 mm inverse detection cryoprobe. 2D NMR spectra were measured using standard CHEMPACK 8.2 pulse sequences (^1H - ^{13}C gHSQCAD and ^1H - ^{13}C gHMBCAD) implemented in Varian VNMRJ 4.2 spectrometer software.

5.2.6 Isolation procedure

The aerial parts of the bL phenotype, cultivated in the phytochamber, were cut, frozen in liquid nitrogen, and stored at $-53\text{ }^{\circ}\text{C}$. The crushed material (90.2 g) was sonicated with MeOH and further extracted under shaking at room temperature with 95% aqueous MeOH overnight. After drying under vacuum, the extract (9.34 g) was partitioned via liquid-liquid extraction between CHCl_3 and water. The dark red aqueous phase (6.36 g) was aliquoted and 1.04 g was applied to Diaion column chromatography. The elution was performed with water, MeOH, EtOAc, acetone, and acidified

acetone. The last fraction (11.6 mg) was further partitioned between EtOAc:water yielding the compound 1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (**47**).

1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (**47**): orange powder; negative ion ESI-HRMS m/z 337.0019 [M-H]⁻ (calcd for C₁₄H₉O₈S⁻, 337.0024); MS² see table Table 5.1; ¹H-NMR (600 MHz, Methanol-*d*₄) δ 7.76 (*dd*, $J = 8.0, 1.5$ Hz, 1H, H-8), 7.43 (*dd*, $J = 8.0, 1.5$ Hz, 1H, H-6), 7.37 (*t*, $J = 8.0$ Hz, 1H, H-7), 6.24 (*s*, 1H, H-2), 4.04 (*s*, 3H, 11). ¹³C-NMR (600 MHz, Methanol-*d*₄, measured with HSQC and HMBC) 164.1 (C-1/3), 165.3 (C-1/3), 150.6 (C-5), 148.0 (C-10a), 125.6 (C-7), 122.3 (C-8a), 118.9 (C-6), 117.2 (C-8), 109.9 (C-4), 104.1 (C-9a), 99.7 (C-2), 57.7 (C-11).

5.3 Results and discussion

5.3.1 Leaf phenotyping

Although the *H. perforatum* plants were produced by asexual proliferation, two phenotypes were distinguishable by their leaf size and form (Figure 5.1.a,b). Ploidy analysis showed that the plants with small narrow leaves (nL) are diploid, whereas the plants with broader oval leaves (bL) are pentaploid (Figure 5.1.b). To quantify the morphological differences of the leaf phenotypes an ImageJ Plugin was developed to automatically determine leaf size and the number of dark and translucent glands of microscopic pictures. The bL phenotype exhibited significantly higher leaf areas compared to the nL plants (Figure 5.1a). Nevertheless, the nL contain significantly more translucent glands, while the number of dark glands was higher in the bL (Figure 5.1c,d). To generate comparable data, the number of glands in relation to the leaf area was calculated. The so-called density was at least two times higher in the nL phenotype for both gland types (Figure 3.1d,e).

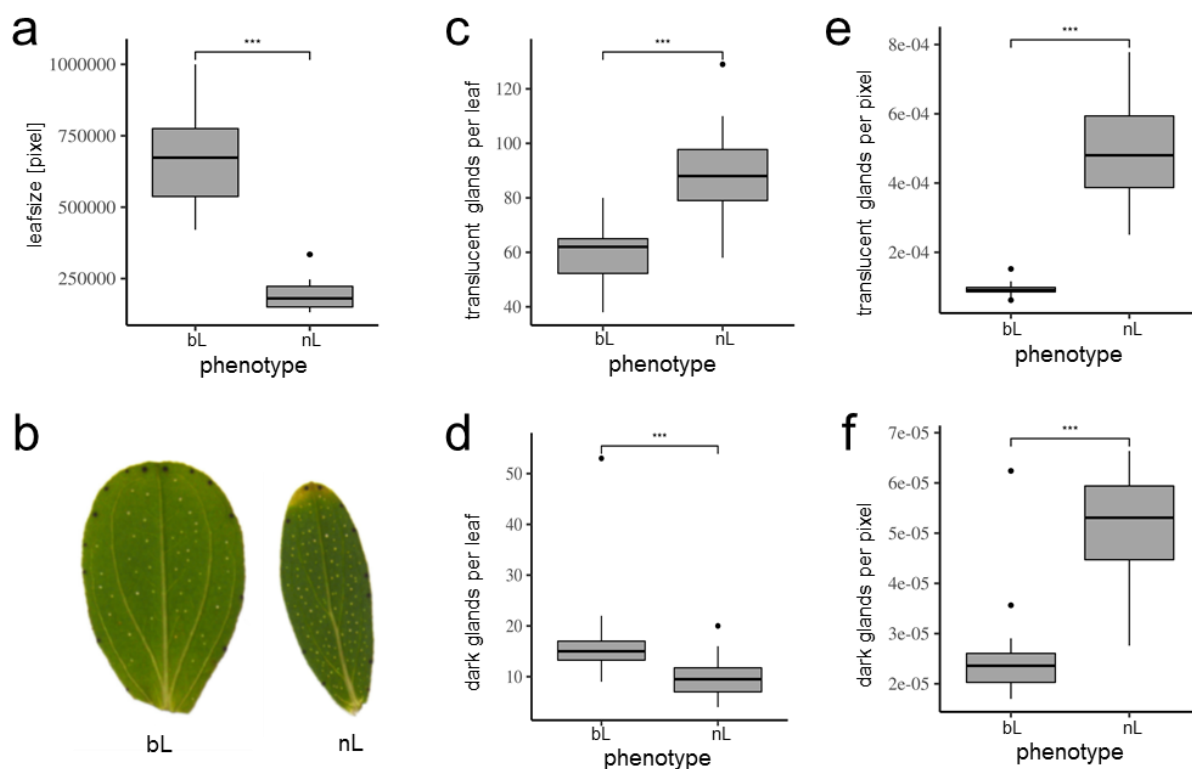


Figure 5.1 Comparison of morphological parameters from nL and bL phenotype ($n=3 \times 10$ leaves), Determination of significance level by Wilcoxon rank sum test (***) $p < 0.001$). a: leaf area; b: microscopic pictures c,d: number of translucent and dark glands per leaf; e,f: number of translucent and dark glands per pixel.

5.3.2 Differential metabolites

To investigate if the phenotypes differ in their chemical composition, untargeted UHPLC-ESI(-)-HRMS was performed. The comparison of both metabolite profiles displays major differences over the whole retention time range. The most significant different peaks with p -values smaller than 0.01 in a Welch's two-sample t -test are summarized in Table 5.1. The compounds were assigned based on the molecular formula and their fragmentation pattern. All common compound classes are affected, but just a few hits are unique in one of the phenotypes.

Chlorogenic acid (**10**) is the only compound that is detectable in the nL plants only. Its isomer neochlorogenic acid ($R_t = 1.81$ min), which differs in the binding site between the quinic and the cinnamic acid moieties, however, can be found in the same intensity range in both phenotypes. Expectedly, the plants with a higher ploidy level showed a higher chemical variance, and some unique features were found. A trihydroxy-xanthone- C -hexoside was assigned to the bL plants. The fragmentation pattern corresponds to lancerin (**45**) and the isomer neolancerin described by Li et al. (2015).¹⁹² Further, two phloroglucinol derivatives were uniquely detectable in the bL phenotype. The m/z 501 ($[M-H]^-$ $C_{33}H_{41}O_4^-$) likely corresponds to 7-epiclusianone, which was already described by Porzel et al. (2014) in other *Hypericum* species.⁵² Also the 3-geranyl-1-(2'-methylpropanoyl)-phloroglucinol (**17**), isolated from *H. punctatum* (see 2.2.3), which showed antimicrobial properties against gram-positive bacteria,⁶⁵ was only detectable in bL material. Additionally two unique compounds for bL with the chemical formulas $C_{23}H_{24}O_{13}$ and $C_{25}H_{42}O_9$ are not yet assigned.

Table 5.1 UHPLC-MS-characteristics of significantly different compounds of two *H. perforatum* phenotypes (bL and nL) (Welch two sample t-test $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***)

M-H] [m/z]	Rt [min]	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (intensity)	Compound	detected replicates		p
						bL	nL	
353.0873	4.27	C ₁₆ H ₁₇ O ₉	-1.205	191.0562 (100), 179.0352 (4)	chlorogenic acid (10)	0	3	**
421.0771	5.50	C ₁₉ H ₁₇ O ₁₁	-1.436	301.0355 (100), 331.0461 (82), 403.0674 (19)	mangiferin (46)	3	3	**
431.1915	5.69	C ₂₀ H ₃₁ O ₁₀	-1.995	385.1869 (100), 285.0405 (80), 284.0329 (21)	n.a.	3	3	**
371.0978	5.71	C ₁₆ H ₁₉ O ₁₀	-1.266	249.0616 (100), 353.0878 (6), 231.0511 (6), 121.0298 (4)	n.a.	3	3	**
507.1136	5.81	C ₂₃ H ₂₃ O ₁₃	-1.132	243.0329 (100)	n.a.	3	0	**
451.1032	5.85	C ₂₄ H ₁₉ O ₉	-0.256	341.0667 (100), 285.0399 (1)	cinchonain Isomer	3	3	***
405.0819	6.12	C ₁₉ H ₁₇ O ₁₀	-1.629	285.0406 (100), 315.0511 (24), 387.0723 (5)	lancerin Isomer (44)	3	0	**
503.1762	6.25	C ₂₂ H ₃₁ O ₁₃	-0.286	195.0689 (100), 195.0526 (12), 195.0444 (5), 151.0784 (5)	n.a.	3	3	***
449.1084	6.54	C ₂₁ H ₂₁ O ₁₁	-0.968	303.0511 (100), 285.0406 (76), 323.0774 (12), 151.0040 (23)	astilbin (19)	3	3	***
463.0876	6.54	C ₂₁ H ₁₉ O ₁₂	-1.359	301.0355 (100), 300.0279 (45)	hyperoside (7/9)	3	3	**
609.1451	6.65	C ₂₇ H ₂₉ O ₁₆	-0.013	301.0354 (100), 343.0461 (7), 271.0250 (5), 255.0302 (3), 300.0279 (39)	rutin (6)	3	3	**
433.1135	7.03	C ₂₁ H ₂₁ O ₁₀	-0.670	269.0456 (100), 259.0613 (12), 178.9988 (6), 151.0040 (4), 287.0562 (44), 286.0484 (10)	dihydrokaempferol-3-rhamnoside (45)	3	3	**
451.1026	7.27	C ₂₄ H ₁₉ O ₉	-0.921	341.0667 (100), 299.0560 (2)	cinchonain Isomer	3	3	***
417.2123	7.46	C ₂₀ H ₃₃ O ₉	-1.692	371.2074 (100)	n.a.	3	3	***
337.0019	7.58	C ₁₄ H ₉ O ₈ S	-1.339	321.9790 (100), 337.0025 (15), 257.0457 (70), 242.0222 (4)	1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (47)	3	3	***
301.0350	7.67	C ₁₅ H ₉ O ₇	-0.584	178.9988 (100), 151.0040 (83), 273.0407 (13), 257.0457 (10)	quercetin (5)	3	3	**
331.1910	10.36	C ₂₀ H ₂₇ O ₄	-0.974	194.0586 (100), 166.0637 (40), 151.0040 (10)	3-geranyl-1-(2'-methylpropanoyl)-phloroglucinol (17)	3	0	**
519.0717	11.63	C ₃₀ H ₁₅ O ₉	-0.877	521.0883 (100), 384.9357 (9), 316.9482 (3), 452.9222 (4), 477.0987 (7)	pseudohypericin (4)	3	3	*
485.2753	11.96	C ₂₅ H ₄₁ O ₉	-0.054	485.2791 (100), 439.2702 (30), 349.2386 (9), 319.2280 (19), 277.2175 (61)	n.a.	3	0	**
501.3005	12.25	C ₃₃ H ₄₁ O ₄	-0.824	432.2306 (100), 432.1801 (3), 417.2072 (4), 271.1342 (3), 327.1968 (2), 287.2015 (2)	unidentified phloroglucinol	3	0	**
345.2068	12.29	C ₂₁ H ₂₉ O ₄	-0.964	208.0743 (100), 301.2176 (63), 152.0119 (19), 261.1499 (6)	2-O-geranyl-1-methylbutanoyl-phloroglucinol (42)	3	3	***
567.3686	12.36	C ₃₅ H ₅₁ O ₆	-1.414	523.3796 (100), 539.3744 (52), 549.3588 (74), 535.3433 (20), 415.3220 (18), 385.2387 (13), 358.2525 (14), 355.2281 (14)	unidentified phloroglucinol	3	3	**
481.3322	12.65	C ₃₁ H ₄₅ O ₄	-0.754	437.3426 (100), 411.2907 (22), 276.1369 (28), 233.0821 (58), 207.0665 (27)	hyperpolyphyllerin (39)	3	3	**
535.3790	13.03	C ₃₅ H ₅₁ O ₄	-0.529	466.3096 (100), 397.2390 (28), 383.2234 (51), 315.1606 (42), 313.1814 (30)	hyperforin (1)	3	3	***
549.3949	13.12	C ₃₆ H ₅₃ O ₄	0.340	480.3251 (100), 411.2546 (21), 397.2389 (39), 329.1762 (33), 313.1813 (27)	adhyperforin (2)	3	3	**
503.0769	13.71	C ₃₀ H ₁₅ O ₈	1.649	503.0773 (100), 459.0874 (7), 461.0668 (2)	hypericin (3)	3	3	***

grey background highlights higher intensities in comparison to the other phenotype

It is well known that hyperforin (**1**) and related phloroglucinols are located in the translucent glands.^{30,97,117} In accordance with this, increased intensities of hyperforin (**1**), adhyperforin (**2**), and the unidentified phloroglucinols ($C_{21}H_{30}O_4$, $C_{35}H_{52}O_6$) were found in the nL phenotype as these plants exhibit a higher translucent gland density. An exception is hyperpolyphyllerin (**39**) with increased intensities in the bL phenotype.

The naphthodianthrone hypericin (**3**) is a supposed agent for photodynamic cancer therapy^{22,23}. It is stored in the dark glands of the plant.^{55,96,117,183} The nL phenotype showed a higher number of these glands per leaf area. Hence it is not surprising that the hypericin (**3**) content is increased in this material. The correlation between the dark gland number and the hypericin (**3**) content was already reported from Zobayed et al. (2006)⁵⁵ and Çirak et al. (2007)¹⁹³, whereas Kirakosyan et al. (2003)¹⁹⁴ were not able to confirm this relation.

The flavonoid profiles varied considerably between both phenotypes. The nL phenotype contained a higher amount of astilbin (**19**), dihydrokaempferol-3-rhamnoside (**45**), and quercetin (**5**), whereas the quercetin-glycosides (**6**, **7/9**) intensities are decreased. Further, the analysis of Kusari et al. (2015) stated that quercetin is located all over the *Hypericum* leaf but especially the dark glands are surrounded by this compound.⁹⁶ Rizzo et al. (2019)⁶⁷ (see chapter 3.3.2) and the correlation studies in chapter 4.3.5.1 showed the coherence between dark glands and the amount of quercetin. Our findings underline this hypothesis because in the dark gland rich phenotype the quercetin content is significantly increased.

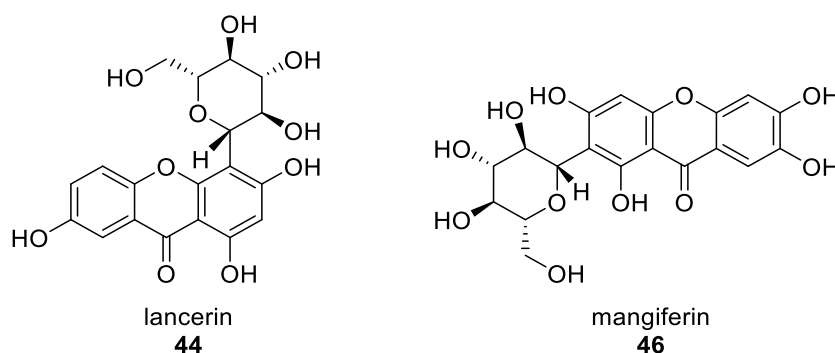


Figure 5.2 Assigned xanthone-glycosides enhanced in broad leaf phenotype.

Interestingly, all xanthone related structures are increased in the bL phenotype. Mangiferin (**46**) and the lancerin isomer (**44**) are xanthone-C-glycosides (Figure 5.2). The sugar moiety increases their polarity and it can be speculated that therefore they are not restricted to secretory cells. Additionally, a sulfonated xanthone (**47**) was enhanced in bL *H. perforatum*. Compound **47** showed in MS/MS spectra a characteristic loss of m/z 80, indicating the cleavage of SO_3 . To isolate this constituent of interest, a screening of different plant organs was performed to identify which material is rich in the compound. It was found that the leaves have an enhanced content in comparison to flowers, fruits, and stems. Especially the bL phenotype cultivated in the phytochamber was rich in the target constituent (data not shown). Therefore it was used as starting material for isolation by liquid-liquid extraction and flash chromatography. The structure was

elucidated with MS/MS analysis, $^1\text{H-NMR}$, and 2D $^1\text{H-}^{13}\text{C-NMR}$ techniques to be 1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (**47**, Figure 5.3). It was already described by Hong et al. (2004), isolated from *H. sampsonii*.¹⁹⁵ As far as we know, *Hypericum* is the only genus containing sulfonated xanthenes. In general, compounds with sulfonic acid moieties are rare in plants. Just a few species are known for sulfonic acid-containing flavonoids, but the importance for the plant is still unknown.¹⁹⁶⁻¹⁹⁹

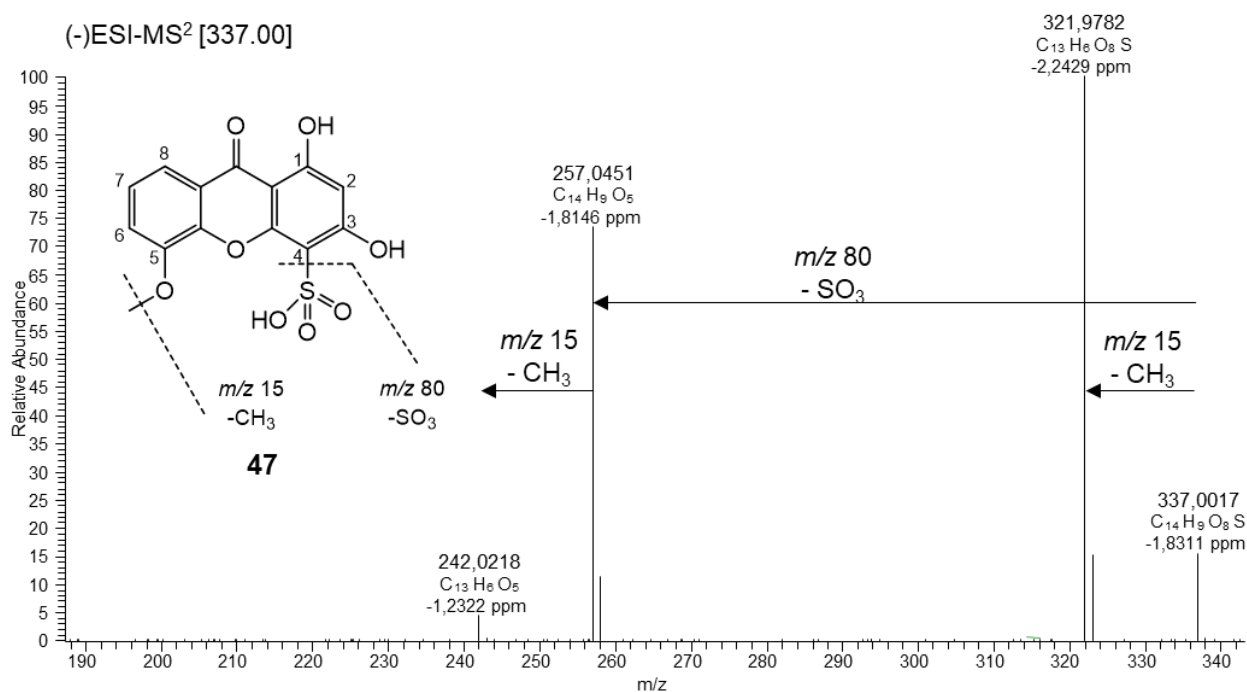


Figure 5.3 ESI-MS² spectrum of 1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (**47**).

5.4 Conclusion

The morphological characteristics of *H. perforatum* plants are highly variable. The change of ploidy and phenotype is associated with the change of the chemical composition. The developed gland counting tool is qualified to compare leaves of different cultivars efficiently. A high gland density was associated with higher contents of the gland located compounds connected to these secretory structures. Thus, the gland counting tool could be used to predict the content of valuable compounds located in translucent or dark glands. A sulfonated xanthone was found for the first time in *H. perforatum*.

6 Phytochemical analysis of different *Hypericum* species

Abstract

The phytochemical analysis of 21 *Hypericum* species was performed with TLC, LC-MS, and ¹H-NMR to determine their interspecific variance. Differences and similarities between the species were evaluated, and characteristic compounds of *Hypericum* were identified. Major differences between species were detected in the profile of high abundant polyprenylated acylphloroglucinols. Based on the spectral data, hierarchical cluster analysis was performed and spectral similarities between different taxonomical sections were discussed. The antibacterial properties against gram-positive (*Bacillus subtilis*) and gram-negative (*Aliivibrio fischeri*) bacteria were evaluated and the activity correlated to the phytochemical features. To reveal additional information about structurally important features which all active constituents have in common, metabolomics workflow was developed for 2D-NMR methods (HSQC, HMBC). The methods combined with activity correlation analysis (ACorA) were used to identify a putative antibacterial phloroglucinol from *H. canariense* and *H. reflexum*.

6.1 Introduction

Natural product research continues to contribute to the discovery of new medicinal drugs.^{2,200} However, more than 80% of the 300000 vascular plants were never analyzed.²⁰⁰ The genus *Hypericum* represents a promising source, because many of the species are used ethnomedicinally. In the meantime, several biologically active compounds were isolated. It is assumed that just 40% of the *Hypericum* species are chemically investigated.⁹

To study the chemical constituents of a species, different research strategies are applied. The classical natural product workflow uses plant material from the wild to isolate the main compounds or via bioactivity guided isolation the active compounds. These time-consuming methods highly depend on the focus of the researcher and normally can not give an overview of the chemical fingerprint of a species. That's why comparative analysis are more often applied to determine similarities and dissimilarities within the chemical compositions of different species. So the results can be used for authenticity validation⁵⁹, to determine taxonomic relations¹⁴⁸, and to identify new plant material to utilize for the classical isolation approach²⁰¹.

The number of comparative studies of different *Hypericum* species increased in the last years. Most relevant studies are combined in Table 6.1. The majority of the studies compared species with respect to the major compounds of the best-investigated species *H. perforatum*. But these studies could show that normally the analyzed species have not much in common with *H. perforatum*.^{148,153,202-205} The occurrence of dark gland structures and the associated naphthodianthrone are sometimes used as a taxonomic marker.⁹ Further the potential of hypericin as anticancer drug raised the interest, which why some studies addressed the naphthodianthrone exclusively.^{25,105,183,206} Just a few studies compared *Hypericum* species in an untargeted fashion to reveal additional information about the chemical constituents, which were not investigated so far.^{52,201}

Hypericum is a highly divers genus, with more than 470 species, which occur as flowering herbs, shrubs, and a few trees. In the present study, 21 species were investigated. A selection of morphologically different species are shown in the Figure 6.1. In Table 6.1 is assembled which species were already included in other studies. Seven of the 21 analyzed species were not chemically investigated in a comparative study.

Some *Hypericum* species are known for their antimicrobial effects.^{65,207} Because of the urgent need of new antibiotics, the analyzed species were tested against gram-negative and gram-positive bacteria.

Table 6.1 Chemical profiling studies that included more than four *Hypericum* species. General parameters of the study are given. Species that were also analyzed in the present study are marked with “x”. Species that were not included in the selected literature are highlighted in grey.

Studies																	
	Kimakova et al. (2018) ¹⁰¹	Fobofou Thesis (2016) ²⁰¹	Porzel et al. (2014) ⁵²	Kosuth et al. (2011) ²⁵	Sołák et al. (2016) ²⁷	Urnek et al. (1999) ²⁰⁵	Kitanov (2001) ²⁰⁶	Saroglou et al. (2007) ²⁰⁸	Kitanov et al. (1998) ²⁰⁹	Smelcerovic Spitteller (2006) ¹⁴⁴	Crockett et al. (2005) ¹⁴⁸	Lazzara et al. (2020) ²⁰³	Nigutová et al. (2017) ¹⁵³	Kucharikova et al. (2016) ³⁹	Cirak et al. (2016) ²⁰²	Revuru et al. (2020) ¹⁸³	
No of species	17	17	6	14	5	6	36	6	36	6	74	6	17	17	8	5	
Plant origin	in vitro	mix	field	in vitro	in vitro	wild	wild	wild	mix	wild	wild	mix	In vitro	in vitro	wild	in vitro	
target	major	non	non	Naphtho-dianthrones	Naphtho-dianthrones	major	Naphtho-dianthrones	volatiles	mangiferin	major	major	Hypericin hyperforin	major	major	major	Naphtho-dianthrones	
Method	LC-MS	NMR LC-MS	NMR LC-MS	LC-MS	LC-PDA	LC-PDA	TLC	GC-MS	TLC	LC-MS	LC-PDA	LC-DAD	LC-MS	DESI-MS	LC-PDA	Maldi-MS	
Included species	<i>H. androsaenum</i>	x	x	x	x	x			x					x	x		
	<i>H. attenuatum</i>								x								
	<i>H. barbatum</i>								x	x							
	<i>H. calycinum</i>		x						x			x	x				
	<i>H. canariense</i>	x			x									x	x		
	<i>H. coris</i>																
	<i>H. curvisepalum</i>																
	<i>H. elodes</i>																
	<i>H. grandifolium</i>																
	<i>H. hirsutum</i>								x	x	x	x	x				
	<i>H. hookerianum</i>																
	<i>H. humifusum</i>	x			x		x							x	x		
	<i>H. inodorum</i>			x													
	<i>H. lagarocladum</i>																
	<i>H. maculatum</i>	x			x		x	x	x	x	x		x	x			
	<i>H. orientale</i>				x							x					
	<i>H. perforatum</i>	x	x	x	x	x	x	x	x	x	x	x	x	x			
	<i>H. polyphyllum</i>	x	x						x							x	
	<i>H. pulchrum</i>	x	x		x									x	x		
	<i>H. reflexum</i>																
	<i>H. tetrapterum</i>	x	x	x	x		x	x		x	x	x	x	x	x		

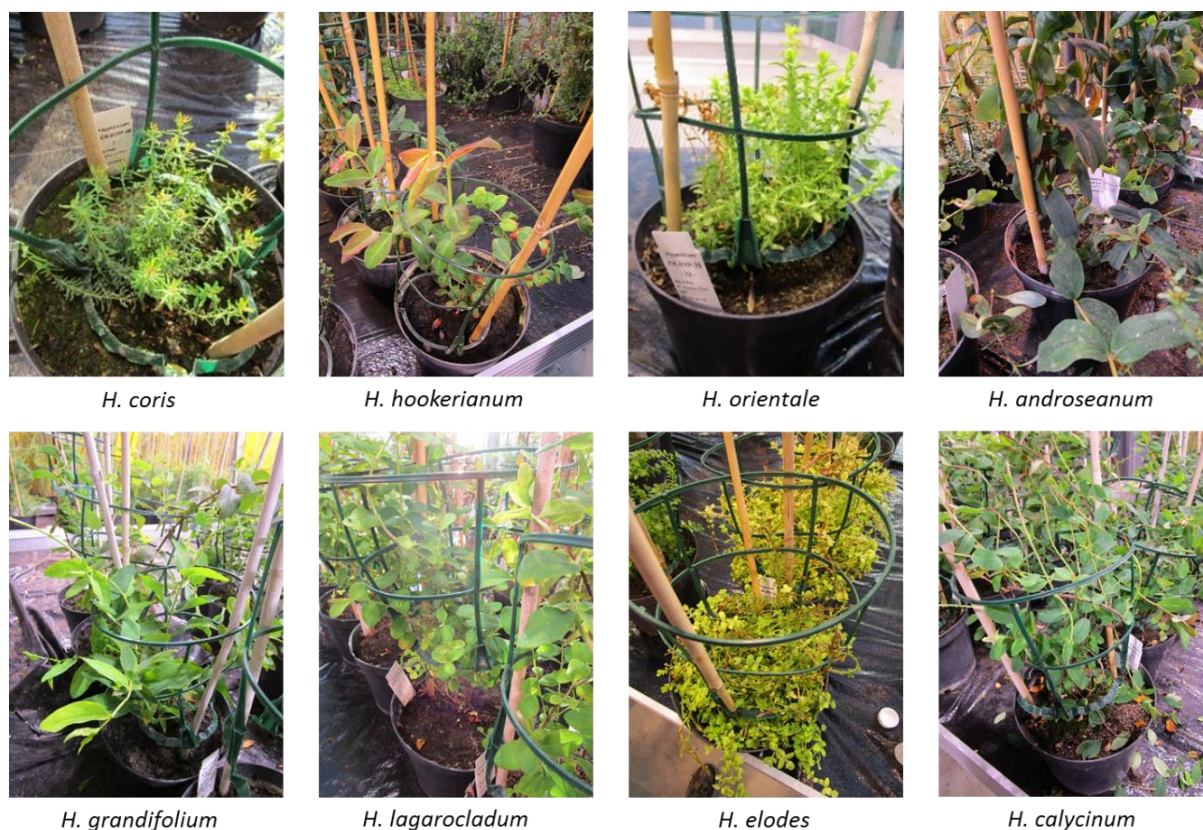


Figure 6.1 Morphologically different *Hypericum* species utilized in the present study.

6.2 Experimental

6.2.1 Plant material and preparation

In total 29 genotypes of 21 different *Hypericum* species were cultivated by Dr. Rizzo from IPK Gatersleben. Seeds were provided by genebank in IPK and Kew Gardens in London. Cultivating conditions in the greenhouses are as described in chapter 2.2.1. The plant specification is provided in Table 6.2. In the following experiments, the sample name is compiled by the origin (IPK/KEW), a genotype identification number, and the replicate identifier (a,b,c). Depending on the germination success, each genotype was represented with up to 6 plants. For investigation one to three biological replicates were generated.

In November 2018 leaves of all samples were collected into 20 ml grinder polyvials (Zinsser Analytic) and immediately frozen in liquid nitrogen. After lyophilization (144 h) two stainless steel balls were added two each sample and powdered with the Cryogenic Plant Grinder (Labman), with the method described in chapter 4.2.2. Equal parts of the powdered material were mixed to produce a pooled quality control sample (QC) for method development and quality criteria during the measurements.

Table 6.2 Specification of used plant material and naming scheme.

Genotype	Species	Section	Biological replicates
IPK_2	<i>H. attenuatum</i> Fisch. ex Choisy	Hypericum	3
IPK_3	<i>H. grandifolium</i> Choisy	Androsaemum	3
IPK_4	<i>H. calycinum</i> L.	Ascyreia	3
IPK_5	<i>H. inodorum</i> Mill.	Androsaemum	3
IPK_10	<i>H. polyphyllum</i> Boiss. & Balansa	Olympia	3
IPK_24	<i>H. perforatum</i> L.	Hypericum	1
IPK_27	<i>H. androsaemum</i> L.	Androsaemum	3
IPK_28	<i>H. maculatum</i> Crantz	Hypericum	3
IPK_33	<i>H. coris</i> L.	Coridium	3
IPK_34	<i>H. hirsutum</i> L.	Taeniocarpium	3
IPK_35	<i>H. orientale</i> L.	Crossophyllum	3
KEW_02	<i>H. hookerianum</i> L.	Ascyreia	2
KEW_04	<i>H. canariense</i> L.	Webbia	3
KEW_05	<i>H. reflexum</i> L.f.	Adenosepalum	1
KEW_09	<i>H. perforatum</i> L.	Hypericum	3
KEW_10	<i>H. perforatum</i> L.	Hypericum	2
KEW_12	<i>H. perforatum</i> L.	Hypericum	3
KEW_15	<i>H. tetrapterum</i> Fr.	Hypericum	2
KEW_19	<i>H. humifusum</i> L.	Oligostema	3
KEW_22	<i>H. elodes</i> L.	Elodes	3
KEW_23	<i>H. barbatum</i> Jacq.	Drosocarpium	3
KEW_26	<i>H. perforatum</i> L.	Hypericum	3
KEW_27	<i>H. pulchrum</i> L.	Taeniocarpium	3
KEW_28	<i>H. pulchrum</i> L.	Taeniocarpium	3
KEW_33	<i>H. curvisepalum</i> N. Robson	Ascyreia	3
KEW_37	<i>H. lagarocladum</i> N. Robson	Ascyreia	3
KEW_44	<i>H. androsaemum</i> L.	Androsaemum	3
KEW_45	<i>H. tetrapterum</i> Fr.	Hypericum	2
KEW_46	<i>H. coris</i> L.	Coridium	2

6.2.2 Thin Layer Chromatography

For thin-layer chromatography (TLC), 100 mg powdered plant material of each sample was extracted with 3.5 ml distilled MeOH. After 30 s rapidly mixing with vortex genie 2 (Scientific Industries), samples were sonicated (15 min) and centrifuged (15 min, 4000 rpm, Megafuge 1.0R Heraeus). The supernatant was removed and dried under nitrogen. The extraction was repeated with the plant material pellet.

For each GT, the extract of one biological replicate (a) was resolved with 1 ml MeOH, and finally, 1 μ l was supplied to the TLC plate. Used material and method are according to section 2.3.1.

6.2.3 UHPLC-MS/PDA analysis

Extracts of all biological replicates were produced according to chapter 4.2.2. For each biological replicate, the extraction was performed in triplicates. The extraction concentration was 2 mg plant material per ml, and the final concentration for injection was adjusted to 1 mg ml⁻¹. Measurements were performed in three randomized batches with the previously used batch structure (chapter 4.2.3) and randomized complete block design. Parameter settings for the analysis with ultra high performance liquid chromatography coupled to a high-resolution ESI mass spectrometer were described in chapter 4.2.3.

6.2.4 UHPLC-MS data processing

Processing of spectra was according to chapter 4.2.4. Peak picking was performed with `findChromPeaks(ppm=15, peakwidth=c(7,25), sntresh=10, noise=1000, prefilter=c(3,1000))`. Retention time correction was applied with `method = loess`. The peaks were grouped within the genotypes with `groupChromPeaks(minFraction=1, bw=3)`. The resulting data matrix contained 5416 features.

Batch correction was applied to overcome batch effects (see 4.3.1.3). Different methods for batch correction available in the R package `BatchCorrMetabolomics` (version 0.1.14) were tested.¹³⁶ The best result was achieved by applying a linear model regression on the study samples.

6.2.5 NMR data acquisition

40 mg of each plant sample was directly extracted with 1 ml methanol-d₄ (99.8%, Deutero), containing 0.935 mM HMDS (hexamethyldisiloxane) as internal standard. After 30 s vigorous mixing with vortex, 15 min ultrasonic bath extraction was applied. Samples were centrifuged for 10 min at 14 000 rpm (5415 C Eppendorf) to separate the plant powder. 0.73 ml of the supernatant were transferred to Deu-Quant-5-7 NMR tubes (Deutero).

Spectra were recorded on an Agilent VNMRS 600 NMR spectrometer at 25 °C equipped with a 5 mm inverse detection cryoprobe using standard CHEMPACK 8.1 pulse sequences `s2pul`, `gHSQCAD` and `gHMBCAD` implemented in the Varian VNMRJ 4.2A software. Signals were referenced to internal HMDS at ¹H- δ 0.062 ppm and ¹³C- δ 1.96 ppm, respectively.

Quantitative ¹H-NMR (1Hq) spectra were measured with: pulse angle = 90°, relaxation delay (d1) + acquisition time (at) = 30.0 s, number of scans (nt) = 128, digital resolution = 0.37 Hz/point, spectral width = 13 ppm. Conventional ¹H-NMR spectra were acquired with: pulse angle = 30°, d1 + at = 3.0 s, nt = 40, digital resolution = 0.95 Hz/point, spectral width = 13 ppm. HSQC and HMBC spectra were acquired with: nt = 16, number of increments (ni) = 128, complex data points in F2 (fn) = 4k, d1 + at = 1.25 s. HSQC were measured with the spectral width of 13 ppm in F2 (¹H) and 175 ppm in F1 (¹³C). The experiment was optimized for ¹J_{CH} = 146 Hz with DEPT-like editing and ¹³C-decoupling during acquisition time. HMBC was measured with the spectral width of 13 ppm in F2 (¹H) and 230 ppm in F1 (¹³C). The HMBC experiment was optimized for a long-range coupling of 8 Hz and the ¹³C band selective HMBC of 4 Hz. For both experiments data acquisition with 50% Non Uniform Sampling (NUS) was applied.

6.2.6 NMR data analysis

¹H-NMR data processing was carried out with MestrelNova 12.0.4-220023. After automatic Fourier transformation with the standard VNMRJ window function and zero filling, phase correction and baseline correction (Bernstein polynomial fit) were applied. Spectra were reduced to integrated regions of equal width of 0.02 ppm. A binned data table (.csv) was generated.

2D-NMR experiments were processed with ACD/Labs 2018.2.5 with a group macro. Spectra of an experiment type were evaluated separately. HMBC was processed using a sine square window function in F2, forward linear prediction up to 256 K, and zero filling to 1,024 K. For HSQC a Gaussian window function in F2, forward linear prediction up to 256 K and zero filling to 1024 K was applied. A Gaussian window function in F1 was used for both methods before Fourier transformation to ESP files. Spectra were divided in 2D pieces, called pixels or buckets. The applied pixel width in F1 (^{13}C) dimension was 1.40 ppm and in F2 (^1H) 0.04 ppm.

Data analysis was performed on the generated data tables with R (3.5.1, available at <https://cran.r-project.org>). Binned ^1H -NMR data was freed from bins corresponding to spectral regions of residual methanol (δ 3.27–3.33 ppm) and water (δ 4.7–5.0 ppm). Bin values were normalized to the quantitative internal standard HMDS. For analyzing the aromatic region of the spectra, bins higher than 5 ppm were evaluated. The pixel data of 2D-NMR experiments undergo a noise-based drift correction. Data were normalized to the bins containing the HMDS signal. To reduce the data matrix, bins below the limit of detection (LOD) in all samples were removed. The LOD was determined as three times the standard deviation of the noise pixels, including all bins with ^1H - δ 10.0–11.0 ppm and -1.0– - 0.2 ppm.

6.2.7 Chemometric analysis

For principal component analysis (PCA), the `pcaMethods` (1.78.0)⁷⁰ package was utilized. Hierarchical cluster analysis (HCA) was performed with `hclust()` function (`method=ward.D`). It was plotted with `ggdendro` (version 0.1-20) and `ggplot` (version 3.2.1)²¹⁰ and `heatapply` (version 1.0.0)¹³⁷.

Activity correlation analysis was performed in R and adapted from Michels et al. (2020).²¹¹ Spearman's rank correlation coefficients were calculated between each pixel of the HSQC and HMBC measurements (normalized to extract amount) and the bioactivities against *A. fischeri* ($c_2 = 500 \mu\text{g ml}^{-1}$). A follow-up permutation test (3000 repeats) was used for calculating an empirical p-value for each correlation. It is computed by dividing the number of correlation values higher or equal than the measured ones by the number of all computed permutations. Pixels with an empirical p-value = 0.0 (none of the calculated permutations showed an equal or higher correlation to the bioactivities) were considered as a bioactive correlated signal.

6.2.8 Antibacterial Assays

The antibacterial assays against the gram-positive *Bacillus subtilis* and the gram-negative *Aliivibrio fischeri* were performed according to the procedures described in Dos Santos et al. (2019).⁷³ The activity against *B. subtilis* was determined in a turbidimetric assay after 16 h incubation ($\lambda = 612 \text{ nm}$). In the assay against *A. fischeri* the bioluminescence was detected after 24 h incubation. For both assays, the extracts described in chapter 6.2.2 were tested in two concentrations (50 and $500 \mu\text{g ml}^{-1}$).

6.3 Results and discussion

6.3.1 Chemophenetic analysis of different *Hypericum* species

6.3.1.1 TLC analysis

TLC analysis is an easy and fast screening method, as described in chapter 2. In pharmaceutical practice it is used for identification and quality control of *H. perforatum*. Therefore this analytical technique was also implemented in this study. One biological replicate per GT was utilized as a representative for the analysis (Figure 6.2). After spraying with Natural Product Reagent (NPR), the color of the spots characterizes the compound class. The red spots belong to the naphthodianthrone hypericin (**3**, $R_f = 0.75$) and pseudohypericin (**4**, $R_f = 0.70$), typical for *Hypericum* species (see chapter 2.3.1). *H. humifusum*, *H. reflexum*, *H. attenuatum*, *H. barbatum*, *H. tetrapterum*, *H. inodorum*, and *H. maculatum* also contain these red compounds which are stored by plants in dark glands. *H. humifusum* seems to be especially rich in naphthodianthrone, which is also indicated by the red colored raw extract. The occurrence of **3** and **4** is used as taxonomical marker because primitive sections lack these compounds whereas more advanced sections contain them.²⁰⁶

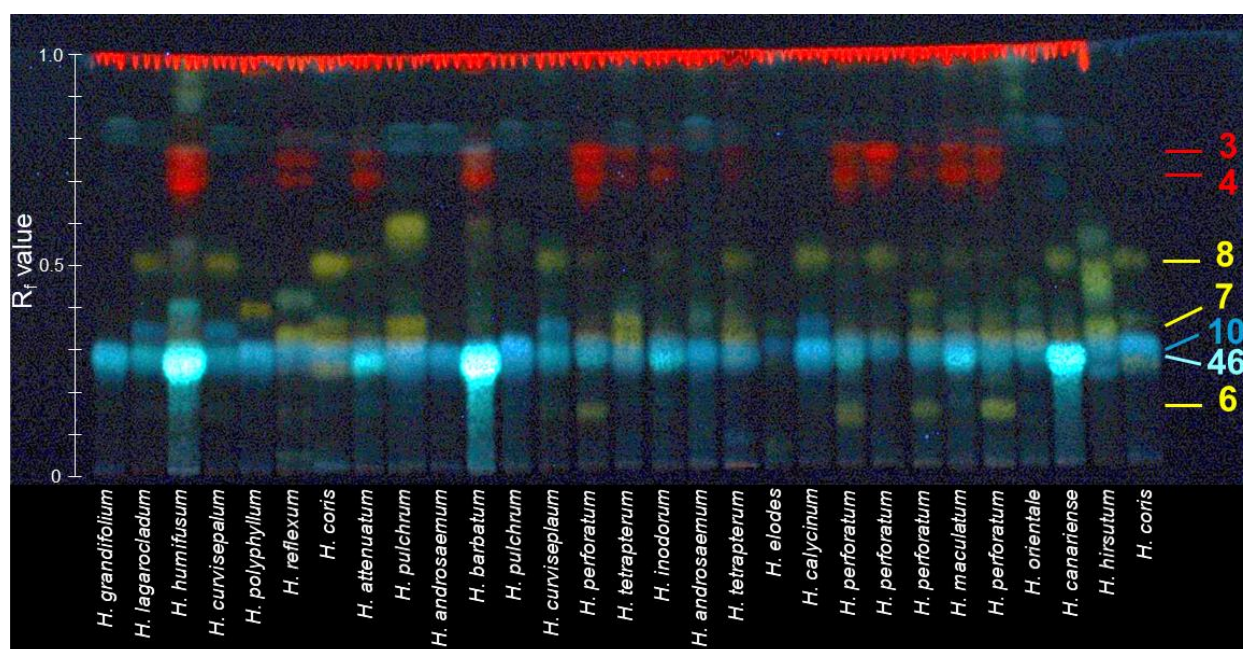


Figure 6.2 TLC fingerprints of *Hypericum* species included in the study. Developed with the mobile phase: ethyl acetate, formic acid, acetic acid, water, dichloromethane (100:10:10:11:25 v/v/v/v/v), sprayed with natural product reagent and visualized at 366 nm. Identified compounds: hypericin (**3**), pseudohypericin (**4**), rutin (**6**), hyperoside (**7**), quercitrin (**8**), chlorogenic acid (**10**), and mangiferin (**46**).

H. perforatum was the only species showing the yellow band at $R_f = 0.16$ corresponding to rutin (**6**). Nevertheless, the usability as a marker compound is questionable because one out of five analyzed *H. perforatum* GTs did not contain rutin (**6**) in detectable amounts. This is in line with the results of chapter 4, where rutin was absent in more than 20% of the analyzed 93 GTs. The

TLC profile of the rutin lacking GT (KEW_10) is hard to differentiate from the similar profiles of *H. maculatum*, *H. inodorum*, and *H. attenuatum* (Figure 6.2).

The majority of the species show yellow bands in the R_f range 0.3–0.6, which correspond to flavonoid monoglycosides. Especially, quercitrin (**8**) and hyperoside (**7**) occur in most of the species. Only *H. grandifolium* and *H. androsaemum* are poor in these compounds. Chlorogenic acid (**10**, blue $R_f = 0.29$) is ubiquitous in all species. Coeluting to that organic acid, a major light blue spot can be observed in *H. humifusum*, *H. barbatum*, and *H. canariense*. This spot was assigned as mangiferin (**46**) and fluoresces originally yellow in UV light (366 nm) but turns intense blue after spraying with NPR. Also *H. attenuatum*, *H. inodorum*, and *H. maculatum* contain this compound in detectable amounts. This suggests the potential use of the mangiferin occurrence to differentiate these species from *H. perforatum* by TLC. However, Kitanov et al. (1998) detected mangiferin in 26 of 36 evaluated species including also *H. perforatum*.²⁰⁹ Further, LC-MS analysis (chapter 2.3.3, Table 5.1) showed that mangiferin (**46**) is present in *H. perforatum*, so that it is not suitable as marker compound.

6.3.1.2 $^1\text{H-NMR}$ analysis

$^1\text{H-NMR}$ is a fast method to record the major primary and secondary metabolites for chemical fingerprinting. The obtained $^1\text{H-NMR}$ profiles are very similar within a species, while the interspecific differences are sometimes large (Appendix 6.1, Figure 6.3). The profiles differ not only in their chemical composition but also in the constituent concentrations. Despite a standardized amount of plant material is used for the analysis, different extract amounts resulted. The lowest amount of extract in relation to the applied plant material was obtained for KEW_22_a (*H. elodes*) with 15% and the highest for IPK_27_b (*H. androsaemum*) with 43%.

Regarding the chemical variance of the extracts from the 21 species, particularly large differences can be observed in ranges typical for phloroglucinol signals (Figure 6.3, yellow parts: 0.80–1.25 ppm; 1.4–1.70 ppm; 3.0–3.25 ppm; 5.0–5.25 ppm). Furthermore, changes in the aromatic range can be seen, i.e. above 5 ppm, characteristic for compound classes such as flavonoids, xanthenes, and phenolic acids.

Full spectrum

Multivariate data analysis was used to identify characteristic signals between the species. Therefore, the phase and baseline corrected spectra were normalized to the internal standard and binned (0.02 ppm). Bins containing solvent signals or whose value did not exceed the limit of detection were removed. The calculated PCAs are shown in Figure 6.4. In the scores plot of PC1 and PC2 (Figure 6.4a) the biological replicates of the GT cluster together. An exception is *H. pulchrum*, where the biological replicates are spread wider. Differences between GTs of the same species can be observed, but the interspecific variations are larger in general, recognizable by *H. perforatum* (5 GTs) or *H. coris* (2 GTs).

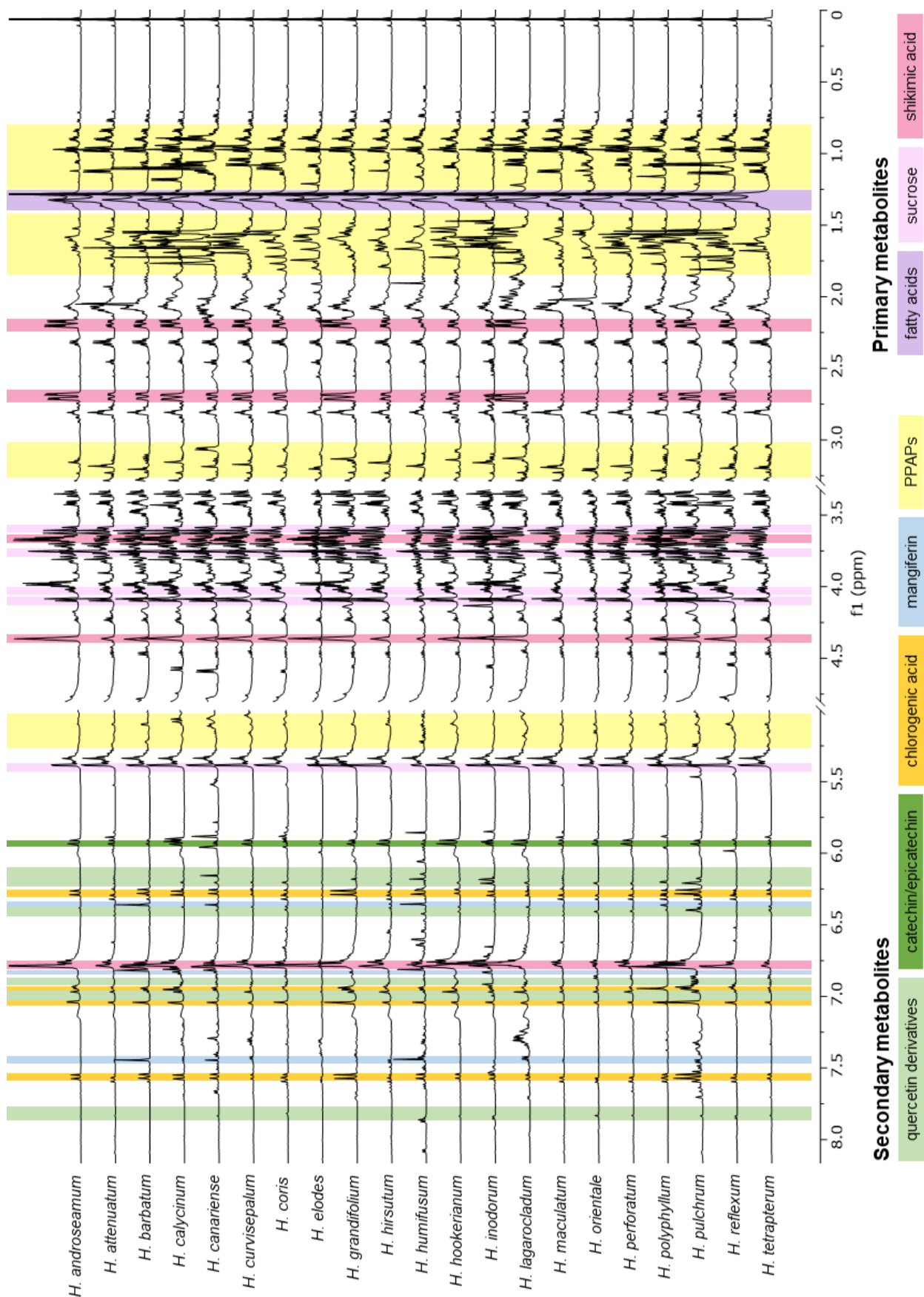


Figure 6.3 Stacked ¹H-NMR spectra of 21 *Hypericum* species. Discriminant regions from PCA loadings plot (Figure 6.4) are highlighted and colored by compound or compound class.

PC1 and PC2 separate two groups of species from the others (Figure 6.4a). Group A comprises *H. canariense* and *H. reflexum*. In the corresponding loadings plot (Figure 6.4c), it can be seen that especially bins typical for polyprenylated acylphloroglucinols (PPAPs) are responsible for the cluster. The greatest influence is exerted by bins with chemical shift δ 1.55 and 1.65 ppm. Most likely, they originate from the methyl groups of prenylated side chains. Furthermore, bins between δ 0.89–1.13 ppm are increased in group A. In this range alkyl signals of the acyl group of PPAPs are localized. Therefore, it can be concluded that *H. canariense* and *H. reflexum* are particularly rich in one or more phloroglucinol derivatives. Signals of unsaturated alkyl chains, which in plant extracts often belong to fatty acids (**13**, Figure 2.3), also appear to be elevated. However, they only have a lower influence on the separation of the two species towards negative values of PC2. Also as primary metabolites their content often depends on growing conditions rather than genetic differences.

Group B, consisting of *H. androsaemum*, *H. grandifolium*, *H. calycinum*, *H. inodorum*, *H. lagarocladum*, *H. curvisepalum*, and *H. pulchrum*, indicates several species which are shifted in the positive direction of PC1 and PC2. This variance is explained by several bins (Figure 6.4c) corresponding to shikimic acid (**11**), which is highly increased in these species. **11** can be easily identified in the spectra as one of the main constituents in these species (marked in pink, Figure 6.3).

The scores plot of PC3 (9.8%) and PC4 (8.0%) and the corresponding loadings plot are depicted in Figure 6.4b and Figure 6.4d, respectively. Especially *H. lagarocladum* (Group C) is separated from the other species in the negative direction of PC3. Responsible are higher bin values at 1.63, 1.65, 1.69, and 1.59 ppm known from prenylated side chains of phloroglucinols. This indicates an increased content of unique PPAPs in *H. lagarocladum*. In addition, also the bin at 0.95 ppm is enhanced. It includes a singlet, which is unique for that species. This enhanced singlet could be derived from a free methyl group bound to the phloroglucinol core structure like reported in hyperpolyphyllerin (**39**).⁵²

Furthermore, the loadings plot displays two discriminating bins (6.79 and 6.77 ppm) corresponding to the shikimic acid (**11**) signal H-3. The variance of this signal is not caused by different quantities of **11** but by a misalignment. The signal is slightly shifted downfield in group C and upfield in group E, possibly caused by different pHs of the extracts.

Group C, *H. canariense*, *H. androsaemum*, *H. reflexum*, *H. hookerianum*, and *H. pulchrum*, is shifted to positive PC3 values and shows enhanced signals especially in bins 1.55 and 1.07–1.31 ppm, which are typical chemical shifts for PPAP derivatives. Group D (*H. elodes* and *H. orientale*) clusters close to group C, but is further located in the positive direction of PC4. Enhanced PPAP bins 1.73–1.75 and 1.51–1.57 ppm explain the variance. Moreover, sucrose signals are decreased (Figure 6.4d). Especially rich in this primary metabolite are *H. barbatum*, *H. perforatum*, and *H. hirsutum*. However, sucrose contents vary significantly with growing conditions too.

Focussing on the whole $^1\text{H-NMR}$ spectra, the major differences between the species are differences in PPAPs and quantitative variations of primary metabolites like shikimic acid (**11**), sucrose (**12**), and fatty acids (**13**).

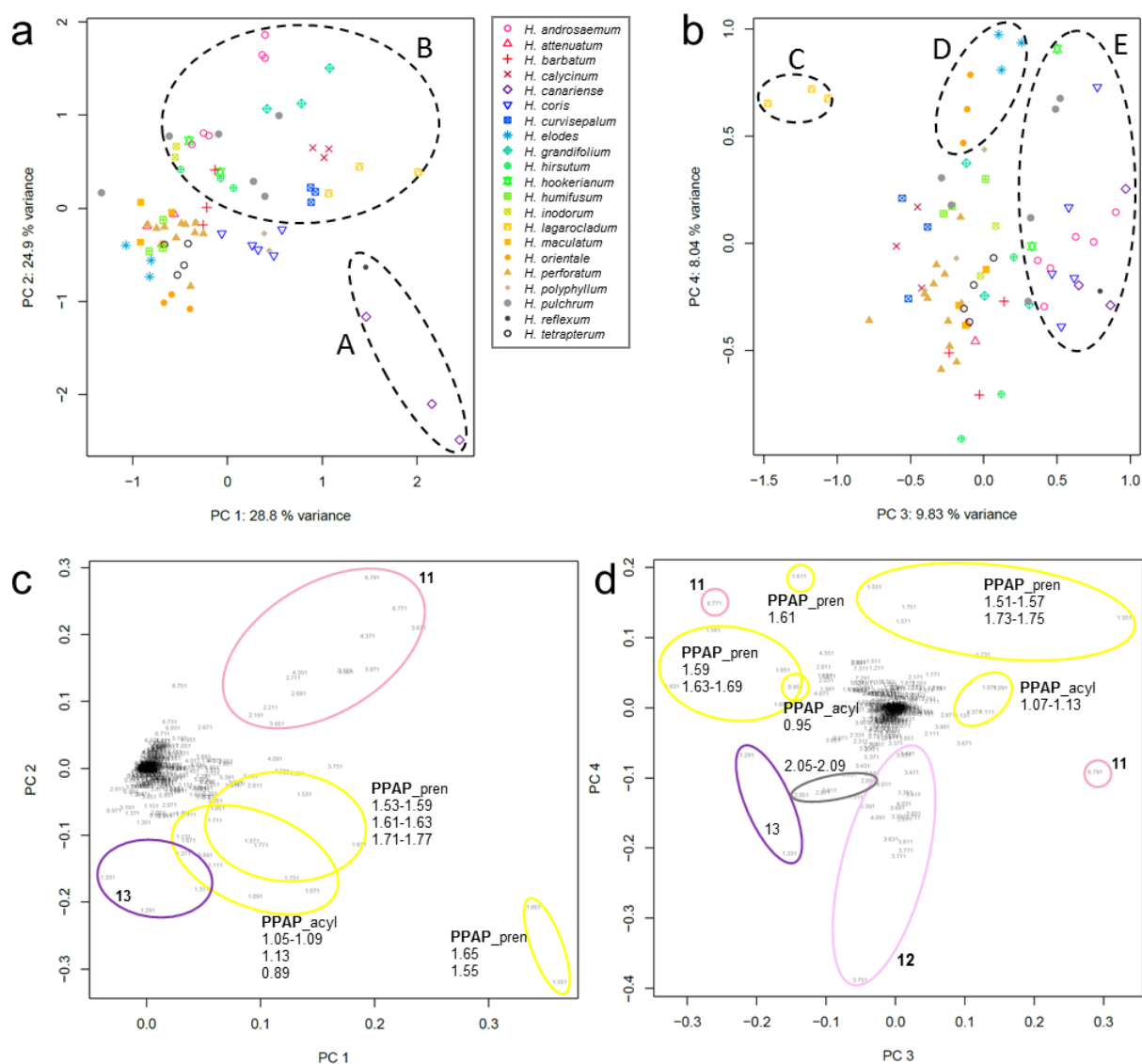


Figure 6.4 PCA of $^1\text{H-NMR}$ full spectra data: scores plots are colored by species and show a) PC1 and PC2 and b) PC3 and PC4. The corresponding loadings plots with assigned loadings colored by compound class (yellow = polyprenylated acylphloroglucinols (PPAPs), magenta = primary metabolites (shikimic acid (**11**), fatty acids (**13**), and sucrose (**12**)) c) PC1 and PC2, and d) PC3 and PC4. Ovals indicate discussed groups and are not necessarily discriminatory.

Aromatic range

The PCA analysis focused on the aromatic region of the spectra ($\delta_{\text{H}} > 5.0$ ppm) was performed to consider also less concentrated secondary metabolites. The scores plot of PC1 and PC2 confirm the results described for the whole spectra (Appendix 6.2). The species which cluster at positive values of PC1 are rich in shikimic acid (**11**). Whereas the variance in PC 2 is introduced by the misalignment of the H-3 signal of shikimic acid and thus is an artefact.

Figure 6.5 displays the clustering in higher PCs where still a grouping based on the species can be observed. PC 3 separates *H. inodorum* in the negative direction (group F, Figure 6.5a). Responsible are enhanced bin values for δ 6.71–6.75 ppm (Figure 6.5c), corresponding most likely to the H-6' of eriodictyol-7-*O*-glucoside (**48**, dd, $J = 2.16, 8.17$ Hz). *H. inodorum* is the only species which shows this signal in the NMR spectrum.

Group G clusters in the positive range of PC4 and comprises *H. pulchrum*, *H. humifusum*, and *H. barbatum* (Figure 6.5a). Several compounds are increased in these species, such as chlorogenic acid (**10**, H-2': δ 7.57 ppm and H-7': δ 7.03–7.05 ppm) and quercetin derivatives like hyperoside and quercitrin (**7**, **8**, H-5': δ 6.93–6.95 ppm). Further signals between δ 7.33 and 7.39 ppm are enhanced, however, could not be assigned. Group H (*H. reflexum* and *H. canariense*) is characterized by an enhanced region at δ 5.03–5.07 ppm. Possibly those signals correspond to the prenyl-CH of phloroglucinols. This would be in line with the previous results that these two species are particularly rich in phloroglucinols.

In PC7 and PC8 the separation of *H. barbatum* and *H. humifusum* (group K) from *H. pulchrum* (group I) is especially interesting (Figure 6.5b). The enhanced content of mangiferin (**46**, Table 6.3) in group K is the responsible differentiating feature (see Figure 6.5d; H-8: δ 7.34–7.45 ppm, H-5: δ 6.81 ppm, H-4: δ 6.35 ppm). These results are in line with the TLC analysis where these species and *H. canariense* show a big light blue spot coming from mangiferin (**46**). *H. canariense* is also in the PCA slightly shifted in the direction of group K (Figure 6.5b). The extract of *H. humifusum* appeared dark red, indicating a high concentration of naphthodianthrones, further, it showed the highest intensity of **3** and **4** in TLC analysis. In literature it is reported to contain up to 9% hypericins in dry extract.²¹² Nevertheless, no signals corresponding to hypericins or pseudohypericins could be observed in the ¹H-NMR spectrum.

Table 6.3 Resonance assignments of discriminant compounds (PCA Figure 6.5) from ¹H- and ¹H-¹³C-NMR spectra (600 MHz, methanol-d₄)

Compound	Assignments	¹ H, δ [ppm] multiplicity (J)	¹³ C δ [ppm] (HSQC)
Mangiferin (46)	CH, 1'	4.90 d (10.0 Hz)	75.3
	CH, 4	6.36 s	94.8
	CH, 5	6.82 s	103.4
	CH, 8	7.44 s	109.2
Eriodictyol-7- <i>O</i> -glucoside (48)	CH, 2	4.56 d (7.5 Hz)	82.8
	CH, 6	6.18 d (2.2 Hz)	98.0
	CH, 8	6.21 d (2.2 Hz)	96.9
	CH, 2'	6.83 d (2.1 Hz)	115.2
	CH, 6'	6.71 dd (2.1, 8.2 Hz)	120.0

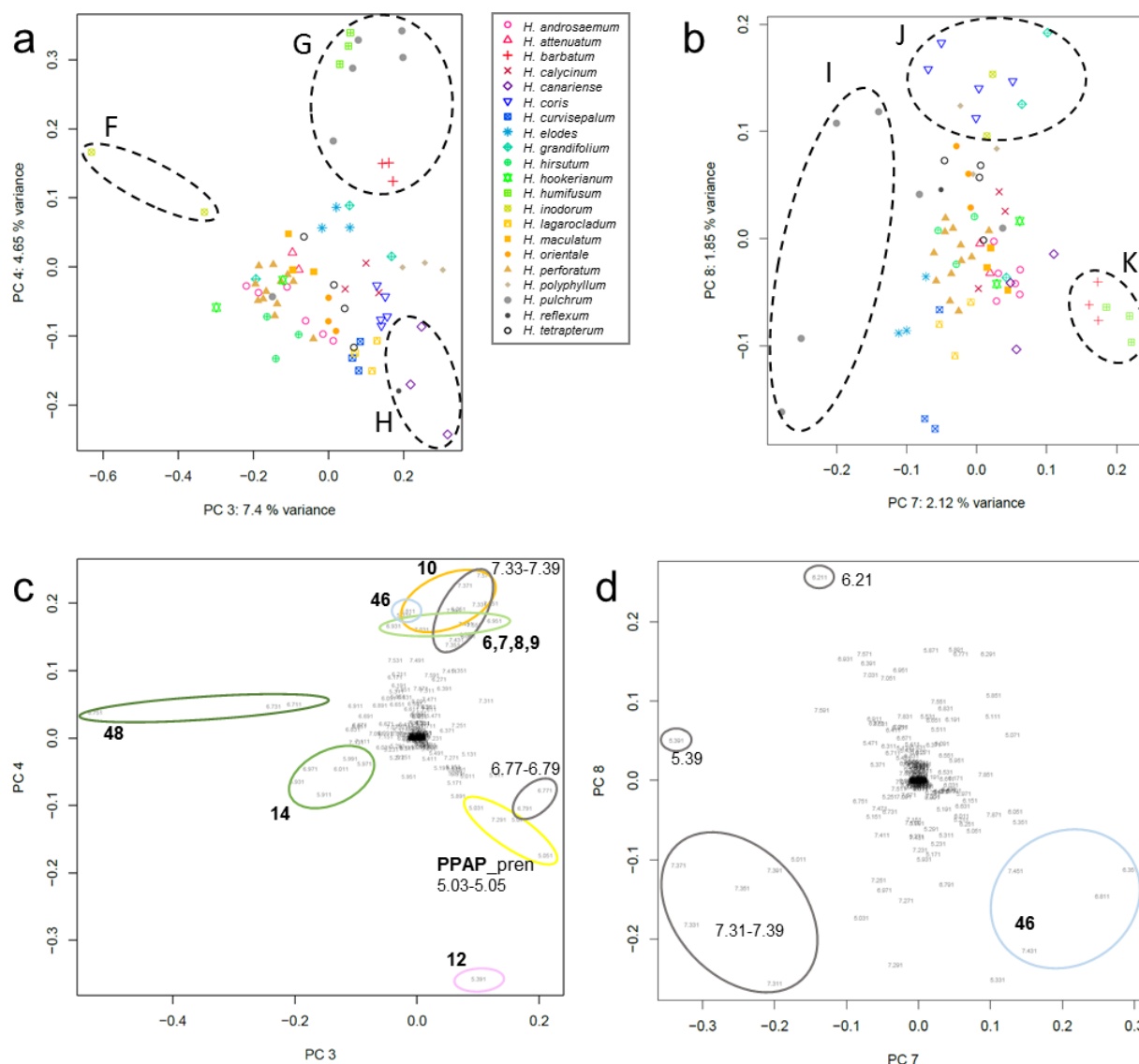


Figure 6.5 PCA of $^1\text{H-NMR}$ aromatic range ($\delta > 5.0$ ppm): scores plots are colored by species and show a) PC3 and PC4 and b) PC7 and PC8. The corresponding loadings plots with assigned loadings colored by compound class (yellow = polyprenylated acylphloroglucinols (PPAPs), blue = xanthenes (mangiferin (**46**)) magenta = primary metabolites (sucrose (**12**)), green = flavonoids (eriodictiol-7-*O*-glucoside (**48**)), catechin/epicatechin (**14**), chlorogenic acid (**10**), quercetin derivatives **6, 7, 8, 9**) c) PC3 and PC4, and d) PC7 and PC8.

6.3.1.3 UHPLC-HRMS

For the LC-MS analysis, each biological replicate was measured in triplicates. In contrast to the NMR investigation, the higher repetition was chosen to counteract the technical variance of LC-MS. The measurements were realized in three batches, with a batch-wise randomization to avoid inter-batch effects such as carryover. The partitioning in batches was necessary because the mass calibration of the MS system is losing accuracy after around 120 injections. Therefore a batch correction between the batches was required. Different methods, described in chapter 4.3.1.3, were tested and visually evaluated using the PCA and described parameters.¹³⁶ Since the samples were very different, QC correction was not satisfactory. By mixing all plant materials, the constituents

were diluted in the QC. Consequently, only a small fraction of the compounds were detectable in the QC and therefore, useable for correction. The best results were obtained by correcting with study samples, where all features were considered. The great chemical diversity of the *Hypericum* species is also underlined by a large number of features in the data set (5416 features).

Untargeted analysis

For each GT, peaks with the highest abundancies in the Full MS spectrum were fragmented and are listed in (Appendix 6.3). Based on the corresponding features a heatmap was generated (Figure 6.6). For all species unique features exist. These features belong almost exclusively to PPAPs, which highlight this class as suitable biomarker for species differentiation. Furthermore, PPAPs ionize really well in positive and negative mode in ESI-LC-MS analysis and are therefore suitable targets. Also in NMR the main differences were attributed to structural elements of PPAPs like prenylated side chains and acyl moieties (chapter 6.3.1.2).

The PCA based on the complete feature list confirms a good species differentiation by specific constituents (Figure 6.8, Table 6.4). In PC1 (35.8%) and PC2 (13.2%) (Figure 6.8a) 8 of the 21 species are separated from the others. *H. lagarocladum* and *H. curvisepalum* show the highest differences in PC1. *H. grandifolium* and *H. calycinum* are located in the same sector as *H. lagarocladum*. Responsible for the separation are the enhanced features 569.4/492 (P96, C₃₈H₄₉O₄) and 501.3/471 (P81, C₃₃H₄₁O₄) (Figure 6.8b). The chemical composition and the late retention time suggest PPAPs. Benzoylphloroglucinols derivatives with identical chemical composition were described for *H. sampsonii*.²¹³ They contain a characteristic unsubstituted phenyl ketone moiety. However, all detected fragment ions of these compounds are explainable by the continuous loss of prenylated sidechains and do not give insights about the acyl group. Bonkanka et al. (2008) observed these two ions in *H. grandifolium* in the positive ionization mode and detected a fragmentation ion at *m/z* 105 (C₆H₅CO⁺), which indicates this structural subunit.²¹⁴ Besides, all three species show increased hyperforin (**1**) signals.

H. androseaemum, *H. perforatum*, and *H. curvisepalum* cluster in the positive sector of PC1 and the negative sector of PC2 (Figure 6.5a). They are rich in hyperforin (**1**) and adhyperforin (**2**) (Figure 6.5b). The enhanced feature 567.4/543 (P114, C₃₅H₅₁O₆⁻) is in accordance to the degradation product deoxy-hydroperoxyfurohyperforin of **1**.

The shift of *H. pulchrum* and *H. elodes* to positive PC2 values is mainly due to the features: 447.3/451 (P65), 447.3/463 (P74, C₂₉H₃₅O₄⁻), and 433.2/507 (P105, C₂₈H₃₃O₄⁻). Those compounds, most likely PPAPs, differ by a methyl group equivalent to the difference between the acylisopropyl group and the acyl methylbuteryl group, as in hyperforin (**1**) and adhyperforin (**2**). Further, those species are characterized by the absence of hypericin related compounds as already shown in TLC. In literature the presence of hypericin (**3**) and pseudohypericin (**4**) is reported for the flowers of *H. elodes* exclusively.^{215,216}

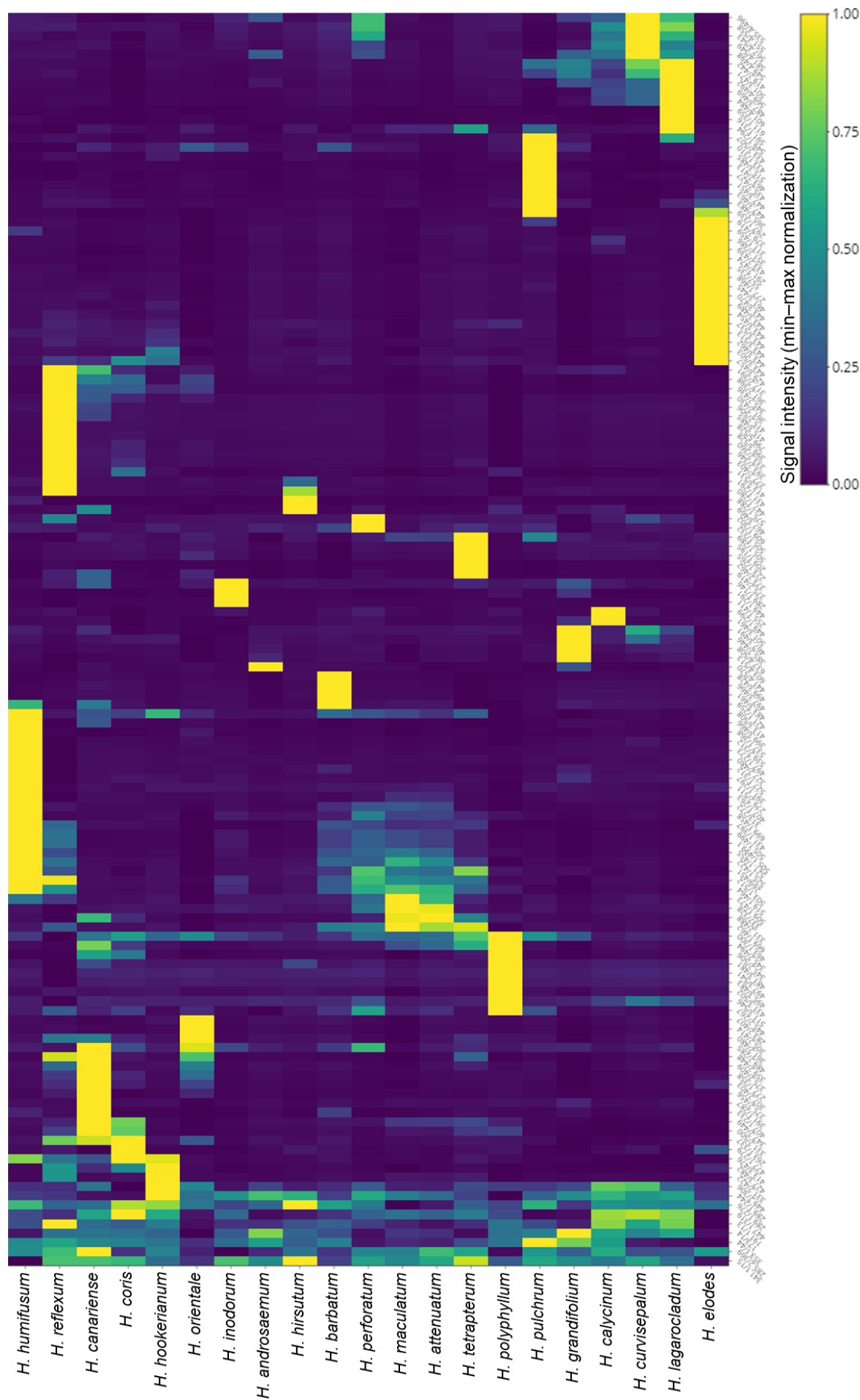


Figure 6.6 Heatmap of signal intensities from major compounds of 21 *Hypericum* species (min-max normalization for each feature) averaged for each species.

Although *H. pulchrum* and *H. elodes* are close in PC1 and PC2, *H. elodes* is clearly separated from the other species by PC4 (9.5%) (Figure 6.8c). The most prominent difference are enhanced compounds with m/z 443.3, represented by five features (P66, P80, P94, P102, P109). This leads to the conclusion that several isomers with the chemical composition $C_{27}H_{40}O_5$ occur in this species. The isomer P80 (443.3/470) reaches the highest values in PC4 and possesses the highest intensity in the chromatogram. Chinesin I (**49**), originally isolated from *H. chinense*, features the same chemical composition.²¹⁷ The molecule **49** (Figure 6.7) possesses five stereocenters, so that numerous stereoisomers are likely. The MS/MS analysis of the detected isomers vary slightly from each other but have two intense fragments in common. Fragment m/z 374 can be explained by the loss of a prenylated side chain $[M-H-C_5H_9]^-$, typical for PPAPs. The m/z 235 corresponds to the ion after additional loss of $C_9H_{15}O$, which would fit to a radical cleavage of the 1-methyl-3-(prop-1-en-2-yl)cyclopentan-1-ol group.

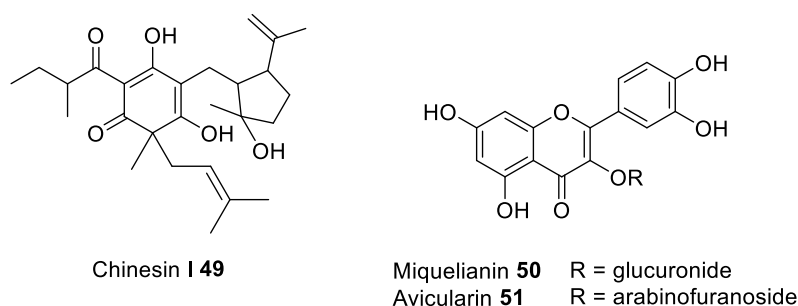


Figure 6.7 Molecular structure of chinesin I (**49**) and the quercetin-glycosides miquelianin (**50**) and avicularin (**51**).

H. coris, *H. polyphyllum*, *H. reflexum*, and *H. canariense* are shifted towards the positive values of PC3 and negative in PC4 (Figure 6.8c). These species contain some identical metabolites, mainly smaller PPAPs (m/z below 500) (Figure 6.8d). Some were also detectable in previous studies in *H. perforatum*, such as dimethylallyl phloroglucinol (**15**). Although these four species form a cluster in PC3 and PC4 they are nicely distributed in PC5 (6.7%) and PC6 (6.0%) (Figure 6.8e). PC6 separate *H. polyphyllum* from the other species, because of enhanced contents of hyperpolyphyllerin (**39**) and adhyperpolyphyllerin (P89) (Figure 6.8f). These PPAPs were already described as characteristic compounds for this species.^{52,201} In addition, also feature 345.2/506 (P103, $C_{21}H_{29}O_4^-$) is increased, based on MS measurements assigned as 2-*O*-geranyl-methylbutanoyl-phloroglucinol (**42**). This feature was also detected in *H. canariense* and *H. elodes*, which are shifted in the positive direction of PC5. The most discriminant feature 331.2/500 (P101, $C_{20}H_{27}O_4^-$) of *H. canariense* was tentatively assigned as 2-*O*-geranyl-methylpropanoyl-phloroglucinol (**41**). Both features were already found in small amounts in *H. perforatum* (chapter 4.3.5.2). The isomer 3-geranyl-methylbutanoyl-phloroglucinol (345.2/436, P58) is also enhanced in *H. canariense* and displays the most discriminant compound in *H. reflexum*. These two species have also an enhanced content of the not identified feature 359.2/448 (P63, $C_{22}H_{31}O_4^-$) in common (Figure 6.8f).

In summary it can be concluded that LC-MS is a suitable method for species specific marker identification in *Hypericum*, because mainly PPAPs differ. 1D ¹H-NMR is less useful, because PPAP signals are highly overlapping due to similarities of structural elements (prenyl chains, acyl-moiety) and only differ in number and connections of the prenyl moieties.

Table 6.4 Peaklist of major discriminant compounds of PCA analysis (extended table in Appendix 6.3)

No	Feature	[M-H] ⁻	Rt (min)	Molecular formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P27	463.1/300	463.0878	5.00	C ₂₁ H ₁₉ O ₁₂	-0.905	343.0462 (4), 301.0354 (100), 300.0285 (48)	hyperoside ^a (7) isoquercetin ^a (9)
P37	263.1/337	263.1287	5.62	C ₁₅ H ₁₉ O ₄	-0.541	194.0586 (100), 166.0639 (18), 151.0040 (6)	dimethylallyl-phlorisobutyrophenone ^a (15)
P40	277.1/358	277.1443	5.97	C ₁₆ H ₂₁ O ₄	-0.839	208.0742 (100), 180.0795 (32), 151.0040 (8)	1-(2-methylbutanone)-3-prenyl-phloroglucinol
P43	351.2/386	351.1606	6.43	C ₂₂ H ₂₃ O ₄	1.331	282.0900 (100), 254.0952 (10), 214.0274 (32), 186.0325 (8)	n.a.
P53	379.2/419	379.1914	6.98	C ₂₄ H ₂₇ O ₄	-0.297	310.1215 (53), 267.0667 (100), 255.0668 (10), 242.0589 (24), 241.0511 (2)	n.a.
P58	345.2/436	345.2063	7.27	C ₂₁ H ₂₉ O ₄	-2.296	276.1374 (61), 233.0824 (100), 221.0825 (15), 208.0746 (28)	n.a.
P59	433.2/439	433.2375	7.32	C ₂₈ H ₃₃ O ₄	-2.176	433.2375 (6), 365.1729 (6), 364.1690 (45), 321.1140 (100), 309.1141 (12), 296.1061 (16), 295.0985 (8), 253.0514 (12)	n.a.
P61	461.3/445	461.2904	7.42	C ₂₇ H ₄₁ O ₆	-1.045	392.2213 (26), 333.1716 (11), 235.0981 (100)	n.a.
P63	359.2/448	359.2224	7.47	C ₂₂ H ₃₁ O ₄	-1.01	359.2237 (10), 315.2338 (11), 290.1530 (42), 247.0980 (100), 235.0981 (14), 222.0902 (25),	n.a.
P65	447.3/451	447.2529	7.52	C ₂₉ H ₃₅ O ₄	-2.533	447.2529 (12), 403.2656 (8), 378.1848 (8), 310.1210 (34), 309.1141 (100), 267.0670 (70), 255.0670 (18), 242.0590 (43), 241.0515 (12)	n.a.
P66	443.3/454	443.2804	7.57	C ₂₇ H ₃₉ O ₅	0.141	399.2914 (25), 375.2147 (9), 374.2107 (100), 317.1764 (11), 303 (6), 248.1058 (10), 247.0981 (9), 236.1045 (9), 235.0981 (80), 222.0902 (17)	n.a.
P67	521.1/455	521.0884	7.58	C ₃₀ H ₁₇ O ₉	1.218	521.0890 (100), 477.0993 (16)	protopseudohypericin (18)
P70	311.1/459	311.1289	7.65	C ₁₉ H ₁₉ O ₄	0.089	267.1394 (55), 242.0587 (100)	n.a.
P71	483.3/461	483.3109	7.68	C ₃₀ H ₄₃ O ₅	-1.464	413.2708 (45), 303.1972 (100), 259.2074 (24), 193.1239 (7)	n.a.
P74	447.3/463	447.2529	7.72	C ₂₉ H ₃₅ O ₄	-2.533	447.2529 (14), 378.1848 (7), 310.1210 (12), 309.1141 (100), 267.0670 (22), 255.0670 (7), 242.0590 (11)	n.a.
P75	519.1/464	519.0772	7.73	C ₃₀ H ₁₅ O ₉	0.125	519.0733 (100), 503.0422 (8), 487.0470 (16), 475.0471 (6)	pseudohypericin (4)
P80	443.3/470	443.2794	7.83	C ₂₇ H ₃₉ O ₅	-2.047	236.1051 (26), 235.0983 (100), 145.1772 (6)	chinesin isomer (49)
P81	501.3/471	501.3007	7.85	C ₃₃ H ₄₁ O ₄	-0.585	501.3028 (17), 389.1770 (10), 363.1612 (100), 312.1140 (65), 309.1141 (14), 296.1060 (14), 253.0513 (22)	benzoylphloroglucinol
P82	291.2/474	291.1601	7.9	C ₁₇ H ₂₃ O ₄	-0.318	247.1707 (10), 223.0975 (6), 222.0899 (100), 166.0276 (2)	n.a.
P83	481.3/481	481.3319	7.91	C ₃₁ H ₄₅ O ₄	-0.879	412.2628	hyperpolyphyllirin (39)

No	Feature	[M-H] ⁻	Rt (min)	Molecular formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P84	413.3/476	413.2698	7.93	C ₂₆ H ₃₇ O ₄	0.066	369.2807 (8), 344.1999 (46), 301.1450 (100), 289.1451 (11), 276.1373 (15), 233.0824 (10)	Deprenylhyperpolyphyllirin (38)
P88	365.2/483	365.1757	8.05	C ₂₃ H ₂₅ O ₄	-0.418	321.1866 (99), 229.0502 (28), 228.0431 (100), 227.0357 (37), 200.0483 (7)	n.a.
P89	495.3/483	495.3488	8.05	C ₃₂ H ₄₇ O ₄	1.75	-	adhyperpolyphyllirin
P93	331.2/488	331.1913	8.13	C ₂₀ H ₂₇ O ₄	-0.491	331.1919 (35), 313.1813 (10), 287.2020 (100), 262.1214 (10), 261.1500 (16), 207.0666 (23), 194.0587 (8)	3-geranyl-methylpropanoyl-phloroglucinol ^a (17)
P94	443.3/489	443.2801	8.15	C ₂₇ H ₃₉ O ₅	-0.513	399.2914 (8), 374.2107 (35), 356.2001 (16), 317.1765 (43)	chinesin isomer
P96	569.4/492	569.3633	8.2	C ₃₈ H ₄₉ O ₄	-0.656	431.2244 (37), 309.1309 (14), 309.1143 (100), 309.0949 (11)	benzoylphloroglucinol
P98	535.4/496	535, 3783	8.27	C ₃₅ H ₅₁ O ₄	-1.799	467.3137 (31), 466.3106 (100), 397.2397 (36), 395.2608 (13), 384.2292 (14), 383.2242 (68), 315.1612 (54), 313.1819 (38)	hyperforin (1)
P99	549.4/498	549.3946	8.3	C ₃₆ H ₅₃ O ₄	-0.589	431.3289 (329), 480.3257 (100), 411.2550 (33), 398.2441 (15), 397.2393 (65), 395.2608 (13), 329.1765 (851), 313.1816 (42)	adhyperforin (2)
P101	331.2/500	331.1908	8.33	C ₂₀ H ₂₇ O ₄	-2.121	331.1921 (8), 287.2022 (64), 195.0662 (15), 194.0588 (100), 152.0120 (14)	2-O-geranyl-methylpropanoyl-phloroglucinol (41)
P103	345.2/506	345.2068	8.43	C ₂₁ H ₂₉ O ₄	-0.877	345.2078 (13), 301.2179 (59), 261.1502 (5), 209.0875 (17), 208.0745 (100), 152.0120 (21)	2-O-geranyl-1-methylbutanoyl-phloroglucinol (42)
P105	433.2/507	433.238	8.45	C ₂₈ H ₃₃ O ₄	-1.068	433.2394 (100), 389.2494 (54), 355 (1922 (61), 321.1138 (40), 309.1138 (38), 295.0982 (26), 287.2023 (27), 253.0512 (21), 145.0300 (19)	n.a.
P108	503.1/512	503.0778	8.53	C ₃₀ H ₁₅ O ₈	1.032	503.0785 (100), 461.0677 (5), 459.0886 (12)	hypericin (3)
P109	443.3/513	443.2803	8.55	C ₂₇ H ₃₉ O ₅	0.118	375.2141 (18), 374.2104 (100), 359.2239 (16), 359.1870 (32), 331.1564 (20), 2470980 (57), 236.1046 (25), 235.0977 (94.58), 234.0901 (62), 233.0825 (32)	chinesin isomer
P111	399.3/528	399.2542	8.8	C ₂₅ H ₃₅ O ₄	0.268	399.2551 (9), 355.2653 (21), 263.1279 (28), 262.1217 (100), 261.1141 (26), 219.0668 (84), 207.0668 (20), 194.0590 (42)	n.a.
P114	567.4/543	567.3694	9.05	C ₃₅ H ₅₁ O ₆	0.577	539.3746 (7), 497.3271 (7), 471.3118 (100), 453.3008 (7), 415.3221 (6)	deoxy-hydroperoxy-furohyperforin

^a identified with standard compound

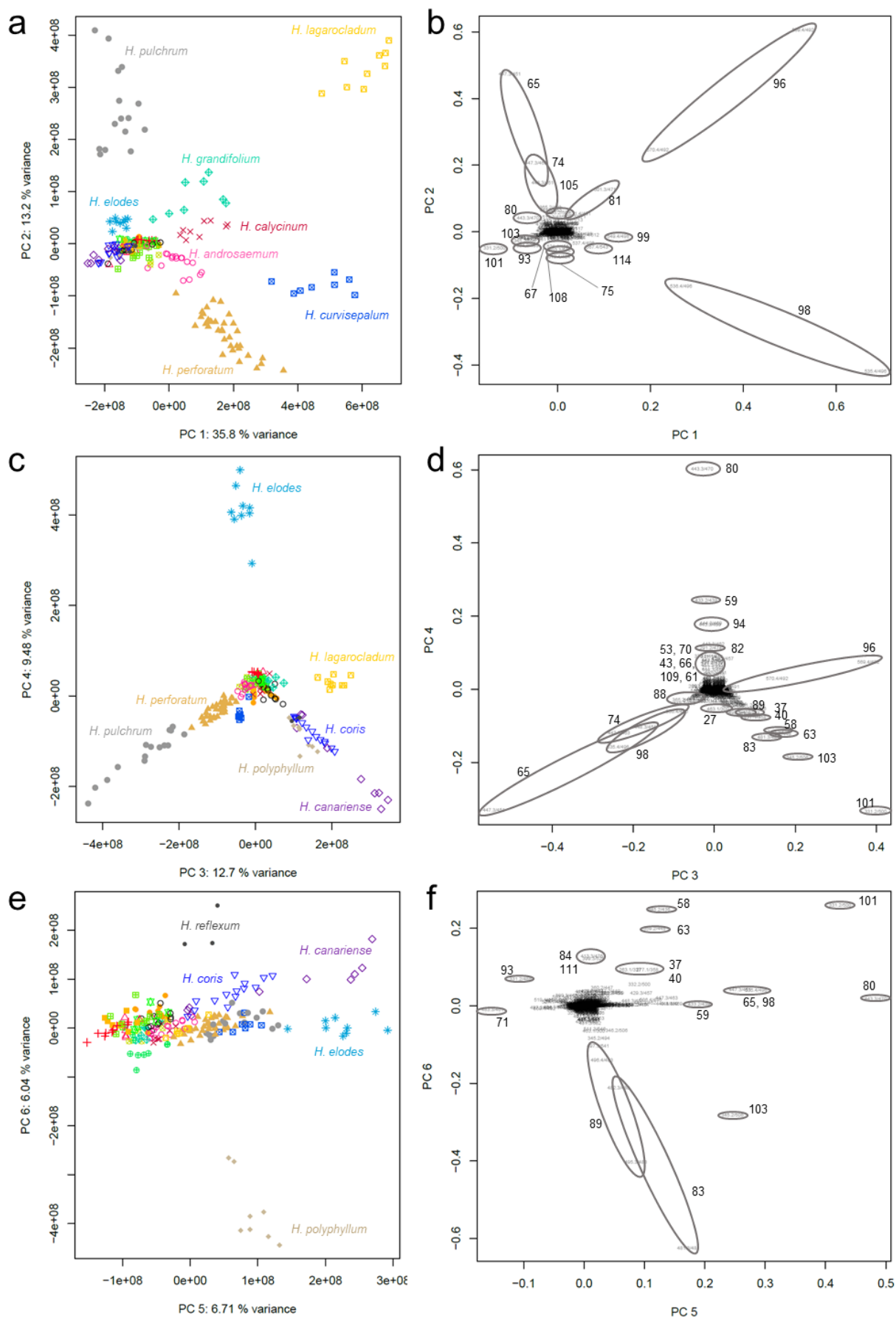


Figure 6.8 PCA of UHPLC-MS data of different *Hypericum* species: PC1 and PC2: a) scores plot, b) loadings plot; PC3 and PC4: c) scores plot, d) loadings plot; PC5 and PC6: e) scores plot, f) loadings plot. Loadings plot is shown with assigned peak numbers of discriminant features, shown in Table 6.4.

Comparison of *H. perforatum* metabolites

The untargeted analysis showed that the chemical fingerprints of the studied *Hypericum* species are highly diverse and have little in common. However, most research was focussed on the comparison of different species with *H. perforatum*. Those targeted studies are based on a relatively small number of secondary metabolites.^{148,202} In Figure 6.9 a heatmap of selected *H. perforatum* secondary metabolites is displayed. In addition to the already described major compounds also miquelianin (**50**) and avicularin (**51**) were included because quercetin glycosides were intensely considered in previous studies.²¹⁸

Notably, quercetin (**5**) and its glycosides (**7**, **8**, **9**) as well as chlorogenic acid (**10**) and catechin (**14**) are detectable within most species. In contrast to that, mangiferin (**46**), avicularin (**51**), and miquelianin (**50**) are more species-specific. The heatmap reveals a clustering of the hypericin (**3**) containing species and of hypericin missing species. Like in earlier studies described (chapter 4.3.5 and 3.3.2.2) hypericin (**3**) occurs together with related compounds. In the heatmap hypericin (**3**), pseudohypericin (**4**), protohypericin (**18**), emodin (**20**), endocrocin (**28**), skyrin-6-O- β -glucopyranoside (**29**), and skyrin-6-O- β -arabinofuranoside (**30**) form a cluster. They are observable in *H. humifusum*, *H. barbartum*, *H. inodorum*, *H. reflexum*, and species belonging to section *Hypericum* (*H. perforatum*, *H. maculatum*, *H. attenuatum*, and *H. tetrapterum*). This is in line with the TLC results (chapter 6.3.1.1). *H. humifusum* contains the highest amounts of hypericin-related metabolites (Figure 6.9). This species is an interesting target to gain more information about the hypericin biosynthesis. One abnormality of this species is the lack of skyrin-6-O- β -arabinofuranoside (**30**), which correlates in other species with **3** and **4**.

Interestingly, *H. perforatum* is not the only species with the co-occurrence of the PPAPs hyperforin (**1**) and adhyperforin (**2**) and naphthodianthrone. Although *H. inodorum* contains both valuable compound classes, as reported by Aziz et al. (2000),²¹⁹ the highest hyperforin concentration was found in *H. curvisepalum* and *H. lagarocladum*.

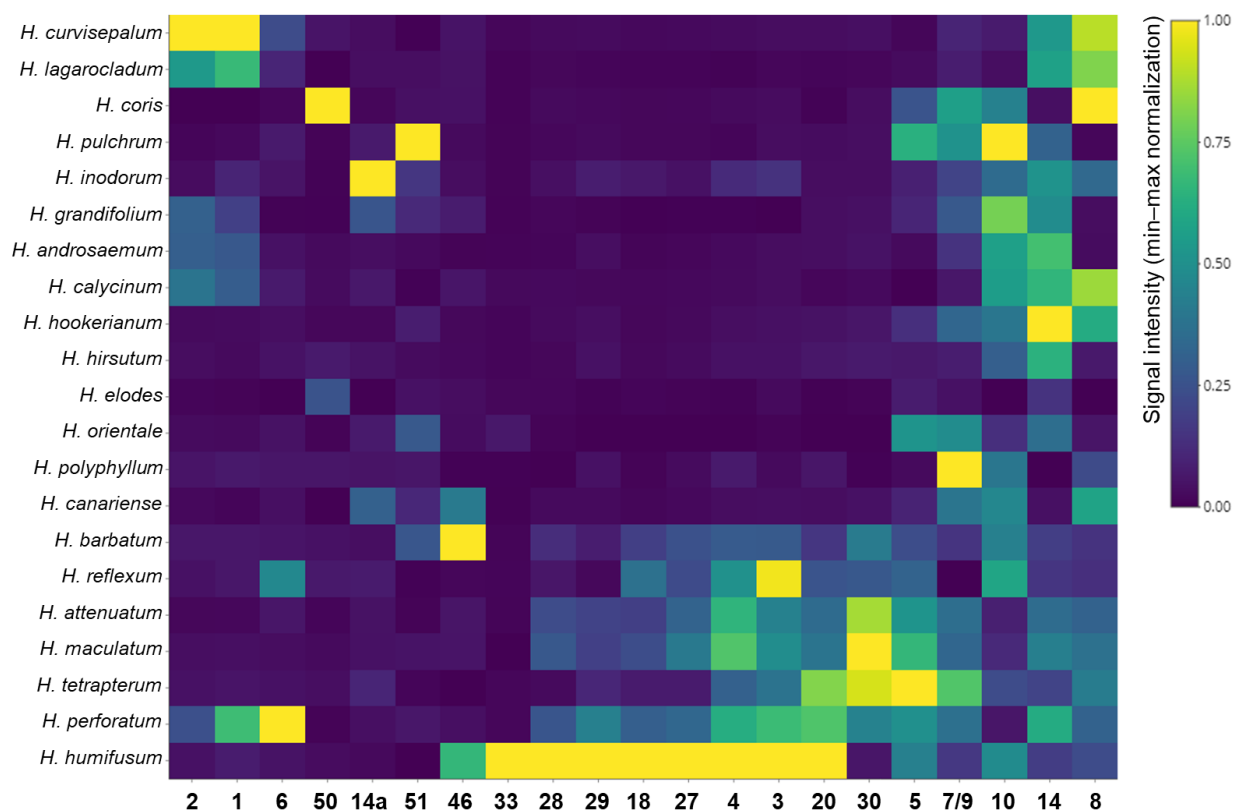


Figure 6.9 Heatmap of signal intensities from major *H. perforatum* compounds (min-max normalization for each feature) averaged for each species.

6.3.1 Phylogenetic correlation to metabolite profiles

The study includes 21 species of 12 taxonomic sections (Table 6.2). By hierarchical cluster analysis (HCA) it was investigated if species from identical section show similar molecular profiles and, therefore, cluster together. The HCA is more suited for this research question because it clusters the species by similarities and not as in the PCA by highest variance. The results based on LC-MS and $^1\text{H-NMR}$ data are shown in Figure 6.10a and 6.10b. The section *Hypericum*, represented by *H. perforatum*, *H. maculatum*, *H. attenuatum*, and *H. tetrapterum*, clusters together based on $^1\text{H-NMR}$ data. Based on LC-MS data, the cluster of *H. perforatum* exists in some distance to the cluster of the other three species. Both methods show small distances between *H. reflexum* (*Adenosepalum*), *H. canariense* (*Webbia*), *H. polyphyllum* (*Olympia*), and *H. coris* (*Coridium*). This trend was already observable in the PCAs derived from both types of analytical methods. Further, the species of the sections *Ascyreia*, *Androsaemum*, and *Taeniocarpium* are spread in the HCA but often appear close to each other. Their molecular similarity might be explainable by their evolutionary background, because the sections *Androsaemum* and *Taeniocarpium* developed from the *Ascyreia* section.²²⁰

The sections *Elodes*, *Drosocarpium*, and *Crossophyllum* are well separated by bigger distances to the other section, so that a quite unique molecular fingerprint for the corresponding species is assumed.

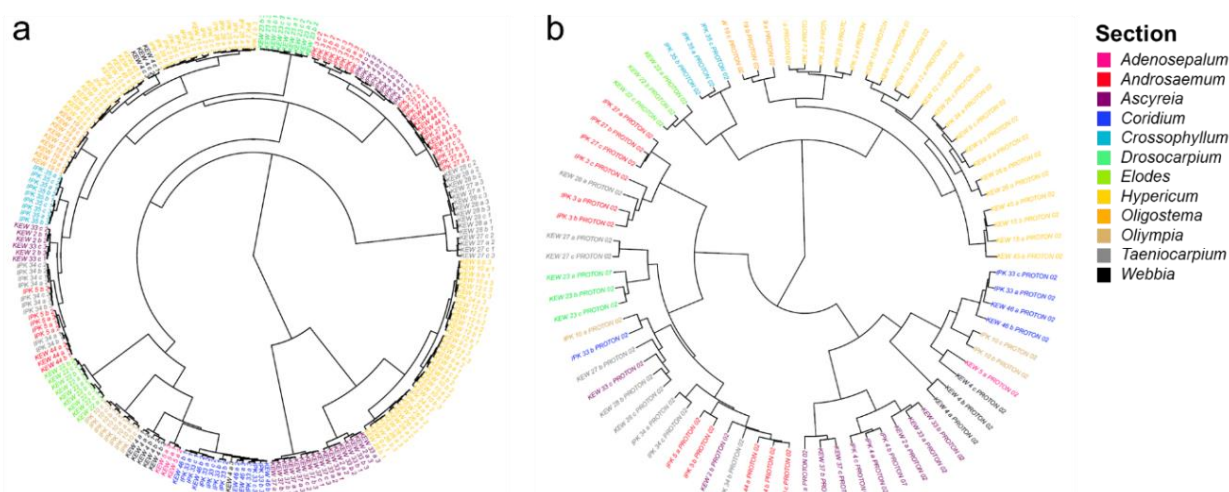


Figure 6.10 Hierarchical cluster analysis calculated with Ward's minimum variance method and colored by taxonomic section of a) UHPLC-MS data and b) $^1\text{H-NMR}$ data.

6.3.2 Antibacterial properties of *Hypericum* species

The lack of new antibiotics and the rising number of multidrug-resistant bacteria made it necessary to develop new antibacterial drugs. Some *Hypericum* species are known to have effects against microorganisms.^{28,65,207} Most of the species analyzed in our study were never tested for their antibacterial properties. Therefore, we included a screening against the gram-negative bacterium *Aliivibrio fischeri* and the gram-positive *Bacillus subtilis*. The antibacterial test was performed twice in triplicates for each biological replicate (Appendix 6.4). The averaged activities per species are compiled in Figure 6.11 for *B. subtilis* and Figure 6.12 for *A. fischeri*.

14 of the tested species extracts inhibited the growth of *B. subtilis* at the highest concentration tested (500 $\mu\text{g/ml}$). At ten times lower concentration only the species *H. canariense*, *H. elodes*, and *H. reflexum* extracts were able to prevent bacterial growth. None of the tested species were active against the gram-negative bacterium *A. fischeri* (Figure 6.12) at the lower concentration (50 $\mu\text{g/ml}$). However, *H. canariense*, *H. orientale*, *H. elodes*, and *H. reflexum* extracts were able to reduce the growth by more than 50%. The best result was achieved with *H. elodes* extract, which completely inhibited growth.

H. canariense, *H. elodes*, and *H. reflexum* showed promising activities against both bacteria, whereas *H. orientale* was explicitly active against the gram-negative *A. fischeri* and showed only moderate results against *B. subtilis*.

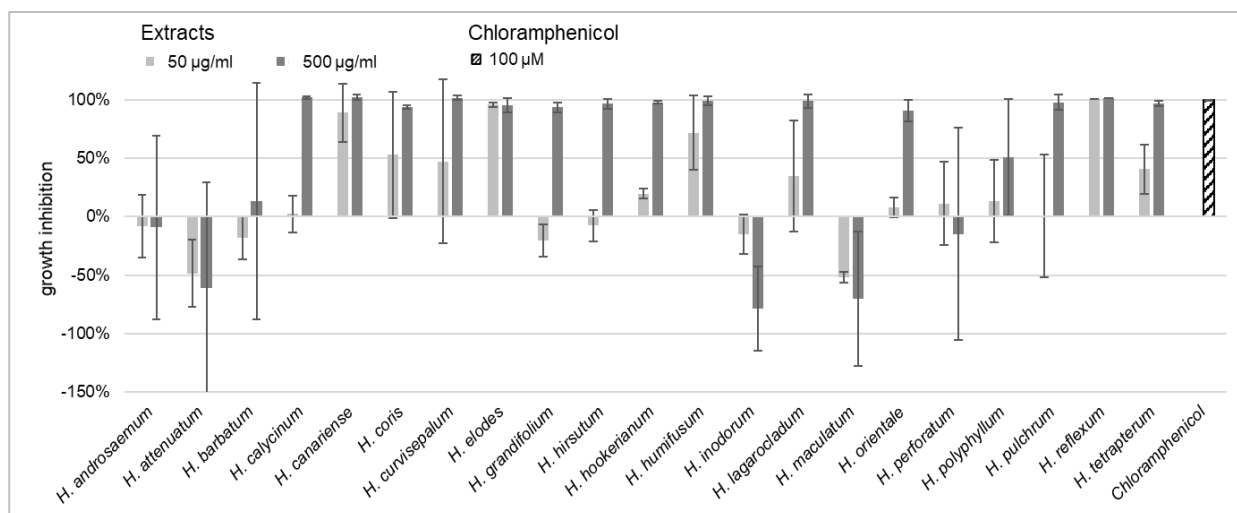


Figure 6.11 Activity of different *Hypericum* species methanolic crude extracts against *Bacillus subtilis* measured at two concentrations (50 µg/ml and 500 µg/ml). Shown is growth inhibition after 16 h in percent and the standard deviation of all average activities of biological replicates per species. Detailed experimental data are shown in Appendix 6.3. Chloramphenicol was used as positive control.

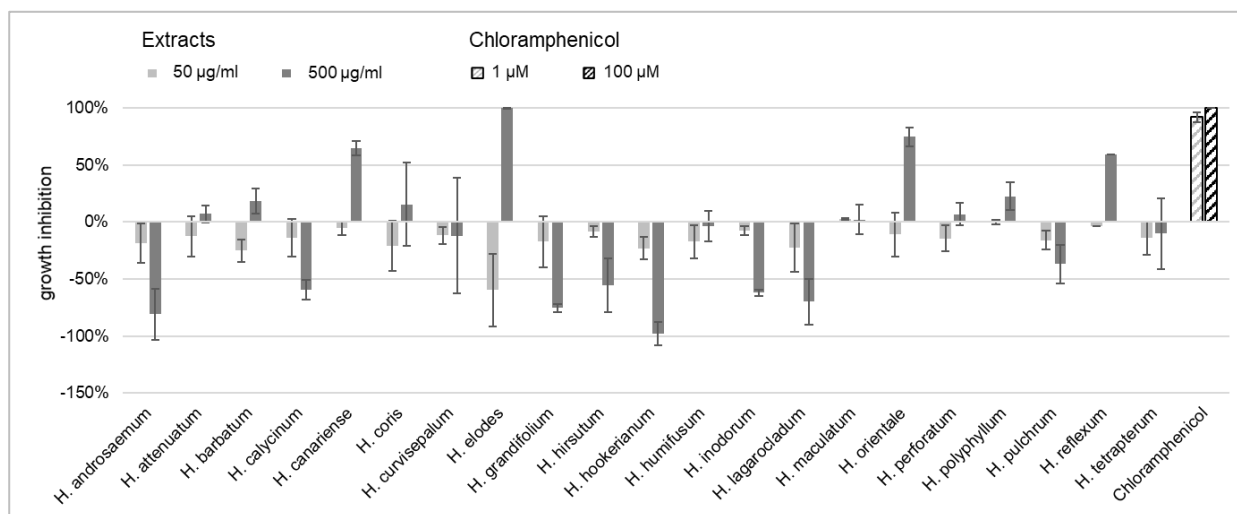


Figure 6.12 Activity of different *Hypericum* species methanolic crude extracts against *Alivibrio fischeri* measured at two concentrations (50 µg/ml and 500 µg/ml). Shown is growth inhibition after 24 h in percent and the standard deviation of all average activities of biological replicates per species. Detailed experimental data are shown in Appendix 6.3. Chloramphenicol was used as positive control.

6.3.3 Activity Correlation Analysis

Application of the method

The identification of active natural products is usually associated with a time-consuming activity guided isolation. In addition, the isolation requires large quantities of starting material. An alternative approach, is the activity correlation analysis (ACorA), where chemical profiles are correlated to the respective bioactivity data to identify the effective compounds.^{221,222}

The present study was conceptualized as high throughput metabolomics experiment, and therefore, the sampled material was not larger than required for profiling analysis. So the ACorA approach was used to identify compounds or a compound class responsible for the antibacterial effects of

the *Hypericum* species (chapter 6.3.2). Due to the overlap of the PPAP signals in conventional 1D NMR, 2D NMR experiments as well as a 1D Pure Shift methods (see chapter 7) were necessary to circumvent this resolution problem.

ACorA was already successfully applied to chemical fractions and extracts where mainly the quantities of the constituents differ and not the chemical composition.^{221,222} Therefore, the utilization of LC-MS seemed suitable. However, the major compositional differences between the *Hypericum* species (chapter 6.3.1) made classical correlation difficult, at least when each compound is classified by one unique identifier as it is the case for LC-MS analysis. In NMR each compound is characterized by multiple peaks that indicate structural elements of molecules. By correlating NMR data with bioactivities, the identification of active compound classes with characteristic chemical moieties was aimed for. 1D and 2D-NMR data for each biological replicate were correlated by Spearman rank correlation to their growth inhibition of *A. fischeri* (500 $\mu\text{g ml}^{-1}$).

For 1D-¹H-NMR the binning table used for multivariate data analysis (chapter 6.3.1.2) was utilized. The 2D-NMR (HMBC, HSQC) processing consists of multiple steps: 1) 2D binning (pixel analysis), 2) drift correction, 3) normalization, 4) noise filtration. The 2D-pixel matrix consists of more than 3000k values. Concerning the information that is included, many of the pixels cover areas of noise and non-crucial information. To reduce the data matrix, pixels with values below the detection limit were removed (~30% HSQC, ~10% HMBC).

Results of the ACorA approach

The ACorA results of all three NMR data sets showed similar results. Since the HMBC results are most meaningful, only these are presented here. The correlating pixels are depicted in Figure 6.13 with the overlaid HMBC spectrum of *H. canariense*. First, it can be recognized that most of the correlating pixels (highlighted in grey) are in reasonable chemical shift areas, so only very few false positives were recognized by the ACorA.

The highest number of neighboring correlating pixels correspond to highly shielded signals at δ_{H} 0.8–1.4 ppm, belonging to methyl groups. They show correlations to δ_{C} regions A (0–25 ppm), B (27–80 ppm), and C (175–220 ppm) (Figure 6.13). Region C is characteristic for ³J coupling to a carbonyl group. It is typical for the acyl side chain of PPAPs and was identified in three of the four active species. Only *H. orientale* is lacking these signals. Region A shows the correlation of the methyl groups within the isobutyryl-moiety (Figure 6.14a), whereas region B could also belong to a correlation within the 2-methylbutyryl-moiety (Figure 6.14b).

The correlating pixel regions labeled with D, E, and F (Figure 6.13) with δ_{H} 1.3–1.8 ppm are characteristic for prenylated side chains. The region D (δ_{C} 10–30 ppm) corresponds to the ³J correlation from neighboring methyl groups with each other. The strong HMBC correlations from the methyl groups to the neighboring double bond carbons are displayed in region E (tertiary carbon, δ_{C} 110–128 ppm) and region F (quarternary carbon, δ_{C} 130–150 ppm)(Figure 6.14c,d).

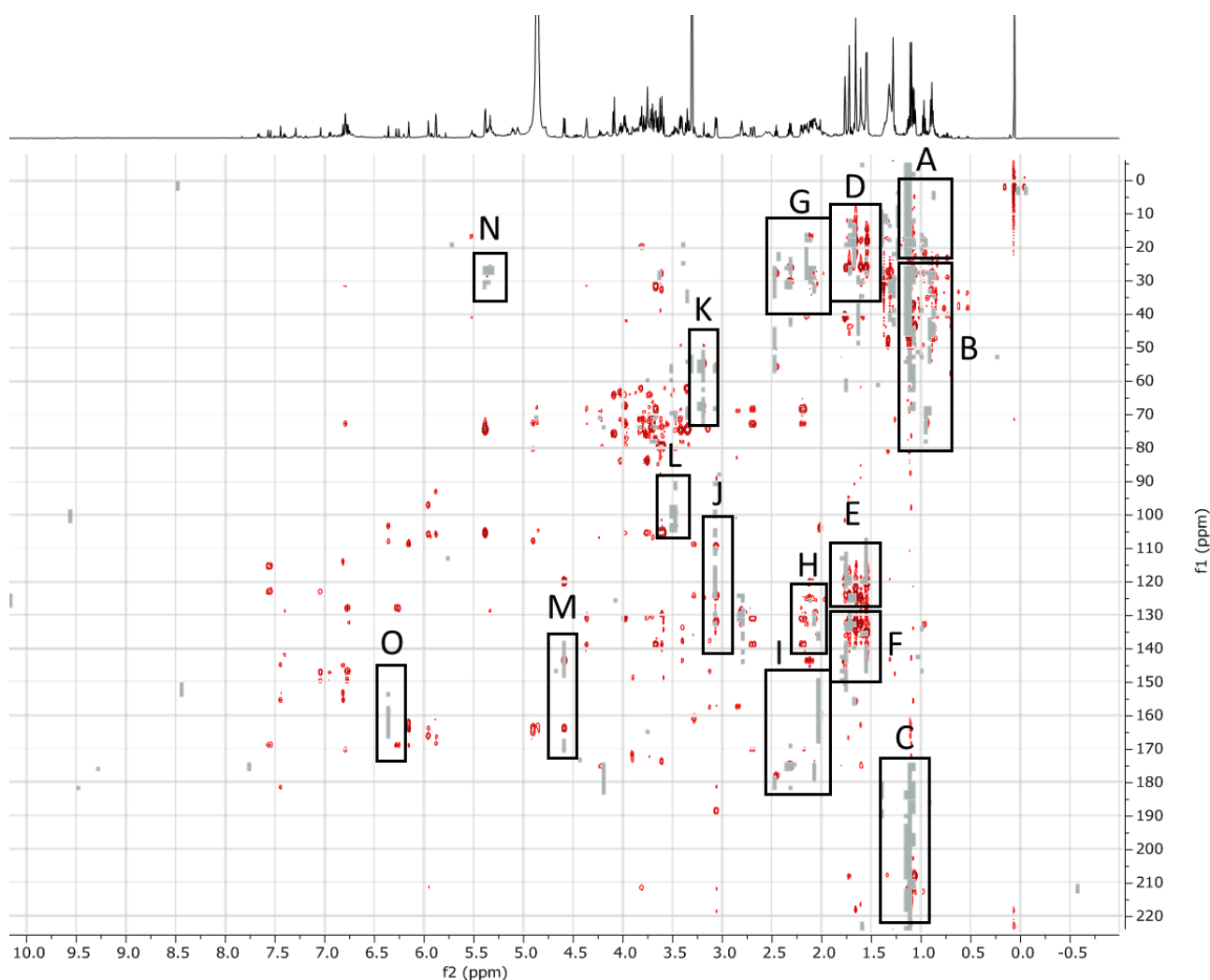


Figure 6.13 HMBC-spectrum of *H. canariense* (KEW_4_a) overlaid with pixels (grey), that correlate significantly to the antibacterial activity against *A. fischeri* (c2 500 μ g/ml), calculated with activity correlation analysis. Regions of black boxes are characteristic signals explained in Figure 6.14.

The correlating pixel regions at δ_{H} 1.9–2.6 ppm correspond to CH_2 signals. They correlate to the quaternary carbon of the prenyl double bond or indicate the direct connection to the core structure of phloroglucinols (region H: δ_{C} 120–145 ppm). Region G (δ_{C} 10–40 ppm) could correspond to geranyl side chains caused by two neighboring CH_2 groups. The correlation of the CH_2 protons to an oxygen substituted quaternary carbon with δ_{C} 150–180 ppm (region I) indicate a carbon bound prenyl chain (Figure 6.14c). The CH_2 group with δ_{H} 4.6–4.8 ppm (region M) is deshielded, indicating a neighbouring nucleophil. Further these protons show HMBC correlations to two quaternary carbons. Those signals are explainable by an O-prenylation. The quaternary carbon at δ_{C} 145 ppm is within the prenylchain whereas the δ_{C} 165 ppm belongs to the core ring structure (Figure 6.14d).

All the highlighted regions which correlate to the bioactivity hint towards PPAPs. In addition also regions K, J, L, N, and O were promising hits in ACorA. However, they can not be assigned to the classical PPAP signals yet. So possibly they correspond to a phloroglucinol derivative with other side chains. To identify an antibacterial compound with the help of the AcorA results the spectra

of *H. canariense* was used, because of the high concentration of PPAPs, recognizable by the dominant methyl signals with δ_{H} 1.5–1.8 ppm. The structure which was tentatively assigned based on MS fragmentation studies could be verified by NMR analysis as 2-*O*-geranyl-methylpropanoyl-phloroglucinol (**41**), as shown in Figure 6.15. The crucial couplings between H-11 and C-10 and C-3 revealed the binding of geranyl side chain at the oxygen bound to C-2.

The PPAP **41**, was not described for *H. canariense* yet, but it was already isolated from *H. punctatum*.⁶⁵ Sarkisian et al. (2012)⁶⁵ reported 1 good results for compound **41** inhibiting biofilm production of different *Staphylococcus* species. The isomer geranylated in 4-*O*-position, known from *H. densiflorum*, showed activity against MRSA.²²³ So compound **41** can be considered as antibacterial principle of *H. canariense* against *A. fischeri*. Furthermore, **41** occurred also in *H. reflexum* which exhibited comparable antibacterial effects as *H. canariense*.

However, not all ACorA hits contribute to compound **41** and it is not contained in *H. orientale* and *H. elodes*. Therefore, those species are interesting targets for further chemical investigation to identify the responsible bioactive compounds.

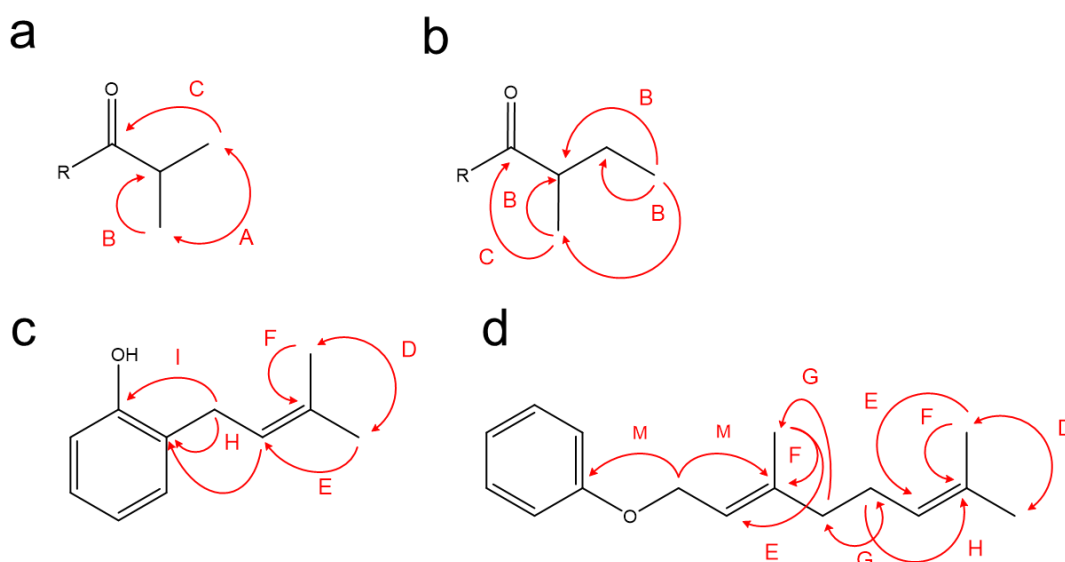
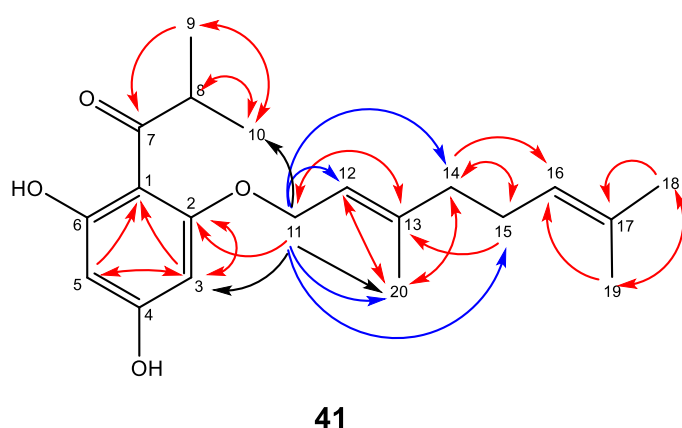


Figure 6.14 Possible HMBC correlations shown on simplified structures observed with ACorA highlighted as regions A to O in Figure 6.13.

**41**

Assignments	^1H , δ [ppm] multiplicity (J)	^{13}C δ [ppm] (HSQC/HMBC)
Cq, 1	-	105.5
Cq, 2	-	163.7
Cq, 3	5.95 d (2.3 Hz)	93.0
CH, 4	-	166.0
Cq, 5	5.88 d (2.3 Hz)	96.9
CH, 6	-	168.2
Cq, 7	-	211.3
CH, 8	3.81 m	40.4
CH ₃ , 9/10	1.10 d (6.7 Hz)	19.8
CH ₂ , 11	4.59 d (6.7 Hz)	66.4
CH, 12	5.52 t (6.9 Hz)	119.8
Cq, 13	-	143.4
CH ₂ , 14	2.11 m	40.4
CH ₂ , 15	2.15 m	27.4
CH, 16	4.79 m	120.1
Cq, 17	-	135.0
CH ₃ , 18	1.54 s	26.1
CH ₃ , 19	1.55 s	18.1
CH ₃ , 20	1.76	16.6

Figure 6.15 Structure of the putative bioactive compound **41** identified in *H. canariense* and the resonance assignments from ^1H - and ^1H - ^{13}C -NMR spectra (600 MHz, methanol- d_4). Couplings are presented with red arrows for HMBC. Selective TOCSY (blue) and selective ROESY (black) were measured irradiating proton 11.

6.4 Conclusion

The untargeted chemophenetic analysis reveals great variation among 21 *Hypericum* species. Both NMR and LC-MS analyses demonstrate that the highly concentrated polyphenylated phloroglucinols are mainly responsible for the discrimination of the species. LC-MS analysis proved to be more appropriate, since the signals of PPAPs often overlap in NMR. Overall, the investigated species possess only few similarities in their metabolite profiles. Some of the species showed antibacterial effects. The activity correlation analysis was applied to 2D-NMR data for the first time and several structural elements characteristic for polyphenylated phloroglucinols were identified. Based on the correlating regions the potentially antibacterial compound **41** was identified in *H. canariense* and *H. reflexum*. The combination of 2D-NMR with ACorA could be a new chance to effectively identify substructures responsible for the activity of a number of compounds. This means that an application of ACorA would be especially suitable for multispecies data sets, where similar compound classes are expected. Furthermore, this analysis will facilitate an easier selection of promising species and prior knowledge of the target molecules for isolation.

7 PSYCHE – a valuable experiment in plant NMR-metabolomics

This chapter presents the application of the homonuclear decoupling method PSYCHE for the comparable analysis of *Hypericum* species. It is published as Communication Article²²⁴:

Pauline Stark, Caroline Zab, Andrea Porzel, Katrin Franke, Paride Rizzo,
and Ludger A. Wessjohann.

Molecules (2020) 25, 1-13e, doi: 10.3390/molecules25215125

Abstract

¹H-NMR is a very reproducible spectroscopic method and, therefore, a powerful tool for metabolomic analysis of biological samples. However, due to the high complexity of natural samples, such as plant extracts, the evaluation of spectra is difficult because of signal overlap. The new NMR "Pure Shift" methods improve spectral resolution by suppressing homonuclear coupling and turning multiplets into singlets. The PSYCHE (Pure Shift Yielded by Chirp Excitation) and the Zangger-Sterk pulse sequence were tested. The parameters of the more suitable PSYCHE experiment were optimized, and the extracts of 21 *Hypericum* species were measured. Different evaluation criteria were used to compare the suitability of the PSYCHE experiment with conventional ¹H-NMR. The relationship between the integral of a signal and the related bin value established by linear regression demonstrates an equal representation of the integrals in binned PSYCHE spectra compared to conventional ¹H-NMR. Using multivariate data analysis based on both techniques reveals comparable results. The obtained data demonstrate that Pure Shift spectra can support the evaluation of conventional ¹H-NMR experiments.

7.1 Introduction

Metabolomics is understood as a quantitative and comprehensive analysis of metabolites in a complex biological specimen to describe the chemical phenotype.⁴¹ Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the key technologies used for that purpose. NMR is particularly convincing due to the simple sample preparation, the high reproducibility, and the non-destructive character.^{225,226} Originally, it was the method of choice for structure elucidation of pure compounds, however, NMR is increasingly used nowadays for the analysis of complex mixtures.⁴³ Especially the ability to quantify different metabolites at various concentration levels makes ¹H-NMR a valuable tool for metabolite profiling and fingerprinting of biofluids²²⁷, foods²²⁸, and natural product sources such as plants^{52,229}, fungi²³⁰, and corals^{231,232}.

Plant metabolomics is challenging because, in addition to the primary metabolites, each species includes a high number of secondary metabolites, which helps the organism to interact with its environment. It is assumed that the metabolome of the plant kingdom comprises over 200.000 metabolites.⁴¹ The genus *Hypericum* includes around 450 chemically diverse species. The best-known species is *H. perforatum* (St. John's Wort), which is commercially used against mild to moderate depression in the Western World.^{9,233} In general, *Hypericum* species are characterized by several secondary metabolite classes, such as naphthodianthrones, phloroglucinols, flavonoids, and xanthones.^{39,52,101,164} In the ¹H-NMR, those secondary metabolites and primary metabolites can be detected simultaneously.⁵² However, the high number of signals from complex mixtures in a ¹H-NMR spectrum is limited to a small spectral range, which leads to overlapping signals (chapter 6.3.1.2). It is further strengthened by proton-proton scalar couplings, which cause multiplet structures. Regions with overlapping signals complicate spectral analysis and identification.²³⁴⁻²³⁶ 2D NMR methods enhance the spectral resolution by spreading the overlapped signals in a second dimension, which is useful in structure elucidation. However, due to long acquisition times for most of the experiments, they were rarely applied in metabolomics investigations.^{225,237,238} Homonuclear broadband decoupling methods, also called “Pure Shift” methods, were stated by Aquilar et al. as possible “resolution of the resolution problem”.²³⁹ They offer enhanced resolution by removing the effect of proton-proton scalar couplings and turning multiplets into singlets.^{240,241} During the last decades different methods were evolved, such as 2D J-resolved NMR²⁴², slice selective decoupling (Zangger-Sterk)²³⁶, and PSYCHE (Pure Shift Yielded by Chirp Excitation) decoupling^{243,244}.

Besides all the theoretical benefits which Pure Shift methods promise for metabolomics experiments, a major bottleneck is still the low sensitivity, which (depending on the method) reaches around 1-20% of a conventional ¹H-NMR.²⁴⁰ However, recent studies show promising results that Pure Shift methods could be added to the metabolomics toolbox.^{232,234,245-247} Lopez et al. (2019) demonstrated that an untargeted metabolomics approach of *Physalis peruviana* fruits based on SAPHIRE-PSYCHE revealed good results in the STOCYSY and PLS analysis.²³⁴ Furthermore, Santacruz et al. (2020) used PSYCHE decoupling for the differentiation of coral

extracts, and Bo et al. (2019) adapted the method to honey and tea samples.^{232,245} All studies indicated, besides the additional structural information, advantages in the untargeted metabolomics data processing. Those studies handled samples with quantitative differences in metabolites, but just to a certain extent, they also differed in their chemical composition. The suitability of Pure Shift for highly diverse sample sets for which peak picking, and intelligent binning²⁴⁸ is not possible, has been examined in the study presented here. Until now, it is not yet clear to what extent binning has a positive or negative effect on the data processing of PSYCHE spectra. Uniform binning, the data reduction of the spectrum into small integral regions with the same size, often results in signals, especially multiplets of large widths, being unintentionally distributed over several bins and thus only represented inadequately by one bin. A better separation of signals could lead to better binning results.

First, the best suitable Pure Shift method was determined, and the parameters were optimized to analyze complex extracts of different *Hypericum* species, which are known to vary in their major secondary metabolites (chapter 6).⁵² Second, an optimized bin size was determined for PSYCHE spectra. Finally, the performance of the PSYCHE experiment within a multivariate data analysis was addressed in terms of quantifiability, metabolite identification, and applicability.

7.2 Materials and Methods

7.2.1 Plant material and sample preparation

For the study, the 21 *Hypericum* species (plant specification in Table 6.2.) data set described in chapter 6.2.1 was used. NMR sample were prepared in accordance to chapter 6.2.5.

For method development, a mixture of chlorogenic acid (Sigma Aldrich, 1.73 mg/ml) and rutin (Roth, 3.20 mg/ml) was solved in methanol-*d*4.

7.2.2 NMR data acquisition

Spectra were recorded on an Agilent VNMRS 600 NMR spectrometer at 25 °C equipped with a 5 mm inverse detection cryoprobe using standard CHEMPACK 8.1 pulse sequences s2pul, PS1D, and PSYCHE (parameter and pulse sequences see data repository, respectively) implemented in the Varian VNMRJ 4.2A software. Signals were referenced to internal HMDS at 0.062 ppm. Compound mixture and plant extract spectra were measured with a spectral width of 10 and 13 ppm, respectively. Quantitative ¹H-NMR (¹Hq) spectra were measured with: pulse angle = 90°, relaxation delay (d1) + acquisition time (at) = 30.0 s, number of scans (nt) = 128, digital resolution = 0.37 Hz/point. Conventional ¹H-NMR (¹H) spectra were acquired with: pulse angle = 30°, d1 + at = 3.0 s, nt = 40, digital resolution = 0.95 Hz/point. PS1D (Zangger-Sterk) spectrum were recorded with d1 + at = 3 s, nt = 4, digital resolution = 0.37 Hz/point, B1max = 0.1526 kHz, pulse width = 17 ms, and gradient = 1.5 G/cm. The PSYCHE spectra were acquired with d1 + at = 1.7 s; nt = 4; and digital resolution = 1.47 Hz/point for parameter adjustment and d1 + at = 3.0 s; nt = 16;

and digital resolution = 0.97 Hz/point for metabolomics measurements. The applied double saltire Chirp pulses were used with swept pulse flip angle = 10° , τ_{PS} = 15 ms, pulse width = 30 ms, and gradient = 1.0 G/cm.

7.2.3 NMR data processing and data analysis

Data processing was carried out with MestreNova 12.0.4-220023. After automatic Fourier transformation with the standard VNMRJ window function and zero filling, phase correction and baseline correction (Bernstein polynomial fit) were applied. Signal to noise ratios (SNR) of *H. perforatum* extracts were calculated with MestreNova, as the ratio of the intensity of signal δ 6.70–6.78 ppm to the standard deviation of the noise (δ 8.1–10.0 ppm). For the metabolomics experiments, spectra were reduced to integrated regions of equal width (0.01 ppm, 0.02 ppm, or 0.04 ppm). A binned data table (.csv) was generated.

Data analysis was performed on the binned data of conventional ^1H -NMR and PSYCHE spectra with R (version 3.5.1). Bins corresponding to spectral regions of residual methanol (δ 3.27–3.33 ppm) and water (δ 4.7–5.0 ppm) were removed. Bin values were normalized to the quantitative internal standard HMDS. For principal component analysis (PCA), the `pcaMethods` (1.78.0)⁷⁰ package was utilized. For the reduced data set, all bins below the limit of detection (LOD) in all samples were removed. The LOD was determined as three times the standard deviation of the noise (δ 10–11.0 ppm and -1.0– -0.2 ppm).

7.2.4 Data availability

Data are freely available in RADAR (<https://www.radar-service.eu/en/home>) (doi 10.22000/338). It includes raw data of NMR measurements, used pulse sequences, as well as the processed bin tables.

7.3 Results and discussion

7.3.1 Comparison of Pure Shift methods

Different Pure Shift methods are known. The most common experiments are the Zangger-Sterk²³⁶ and the PSYCHE²⁴⁹ method. Both techniques were tested with a compound mixture of chlorogenic acid (**10**) and rutin (**6**). The combination of these *Hypericum* constituents was chosen because the coupling constants of the signals cover a wide range from 2–16 Hz (Table 2.1). The first experiment was performed with the spectrometer control software (VnmrJ) default settings (pulse sequences given in data depository). Figure 7.1 shows the Pure Shift spectra in comparison to the conventional ^1H -NMR. The Zangger-Sterk method (Figure 7.1) reveals, despite the long measuring time of 50 min, a low signal to noise ratio (SNR) and can not compete with the sensitivity of the PSYCHE experiment. Additionally, the measuring time of the PSYCHE experiment is nearly ten times shorter than Zangger-Sterk. Therefore, the PSYCHE method was considered for metabolomics experiments, where acquisition times should be short as usually high

sample numbers and compound stability are important issues. The homonuclear decoupling was achieved for all signals in the PSYCHE experiment to an acceptable extent. The SNR is lower than in conventional $^1\text{H-NMR}$, as already described by Castanar (2017).²⁴⁰ Furthermore, artifacts (e.g., Figure 7.1: 7.65 ppm) were generated in the PSYCHE spectrum, so the parameters had to be optimized to obtain higher spectral quality.

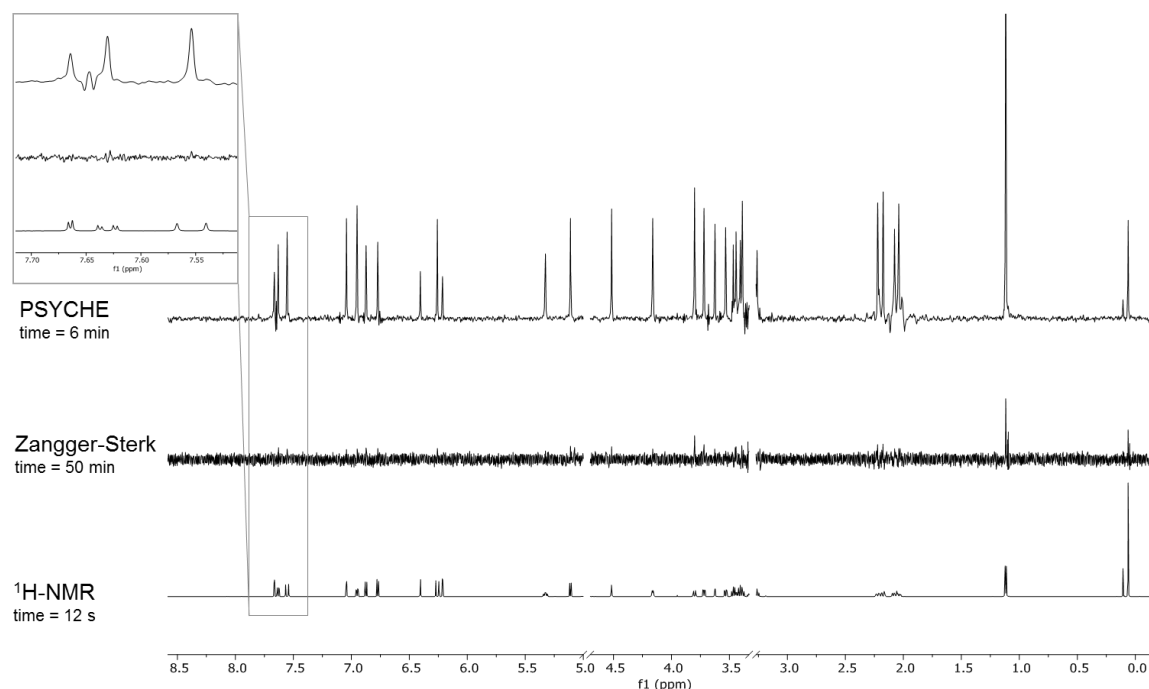


Figure 7.1 Comparison of $^1\text{H-NMR}$, Zangger-Sterk (PSD1²³⁶), and PSYCHE²⁴⁹ spectra acquired with the default parameter settings (number of scans = 4).

7.3.2 Parameter optimization of the PSYCHE experiment

The PSYCHE spectra show the decoupled signals and a variety of undesired signals, resulting from strong couplings, "chunking" sidebands, and other artifacts.²⁴⁴ Experimental parameters were improved to generate spectra with high spectral purity containing a good ratio between wanted and unwanted signals. Furthermore, the sensitivity, calculated as the quotient of SNR and measurement time, was considered as quality parameter. The influence of the essential parameters, swept pulse flip angle, and Pure Shift tau-delay ($\tau\text{PS}=1/(2\text{sw}1)$) on spectra quality was tested. The small pulse angle is the fundamental idea that enables the homodecoupling of the PSYCHE experiment.²⁴⁴ The impact was determined by changing the pulse angle (6° – 18°) gradually with constant τPS (30 ms). The obtained spectra are shown in Figure 7.2a. The utilization of pulse angles above 10° results in decoupling sidebands, observable at signal H-5' of **6** at 6.88 ppm. The calculated sensitivities are summarized in Figure 7.3. It becomes clear that big pulse angles accompany high SNR, and thus a better sensitivity is reached. Nevertheless, the use of small pulse angles is a reasonable compromise, since it leads to the acquisition of spectra with high homodecoupling efficiency.^{240,244,250}

In the next experiments, the τ PS was subsequently changed while the pulse angle was kept constant. With increasing τ PS, the experiment time is getting shorter. While a PSYCHE measurement with a τ PS of 5 ms takes 155 min, an experiment with 45 ms takes only 39 min with the same number of scans of 16. Furthermore, the spectral purity is highly influenced by the τ PS, observable in Figure 7.2b. The suitability of the parameter depends on the coupling constant J of the signals. As shown in Figure 7.2b, the largest J (signal H-7' of **10**, $J = 15.9$ Hz) is successfully decoupled with τ PS values up to 20 ms. These findings are in line with the recommendations of Foroozandeh et al. (2014), who stated that $sw1$ ($sw1=1/(2\tau PS)$) should be twice the highest J to decouple.²⁴⁹ For signals with large J values such as in the case of signal H-7' (**10**, $J = 15.9$ Hz), it can be clearly seen that the negative sidebands increase with increasing τ PS. These sidebands are caused by chunking and show up in the distance of $sw1$ or $1/(2\tau PS)$. On the other hand, a small τ PS leads to artifacts between signals with strong couplings, such as for **6** signals H-2' ($J = 2.2$ Hz) and H-6' ($J = 2.2$ Hz, 8.4 Hz). Thus, the choice of a certain τ PS value is always a compromise. Further, COSY type and strong coupling artifacts could be reduced by elongating the duration of the pulse. However, the pulse length of 30 ms, which is consistently reported in the literature, gave satisfactory results.^{234,245}

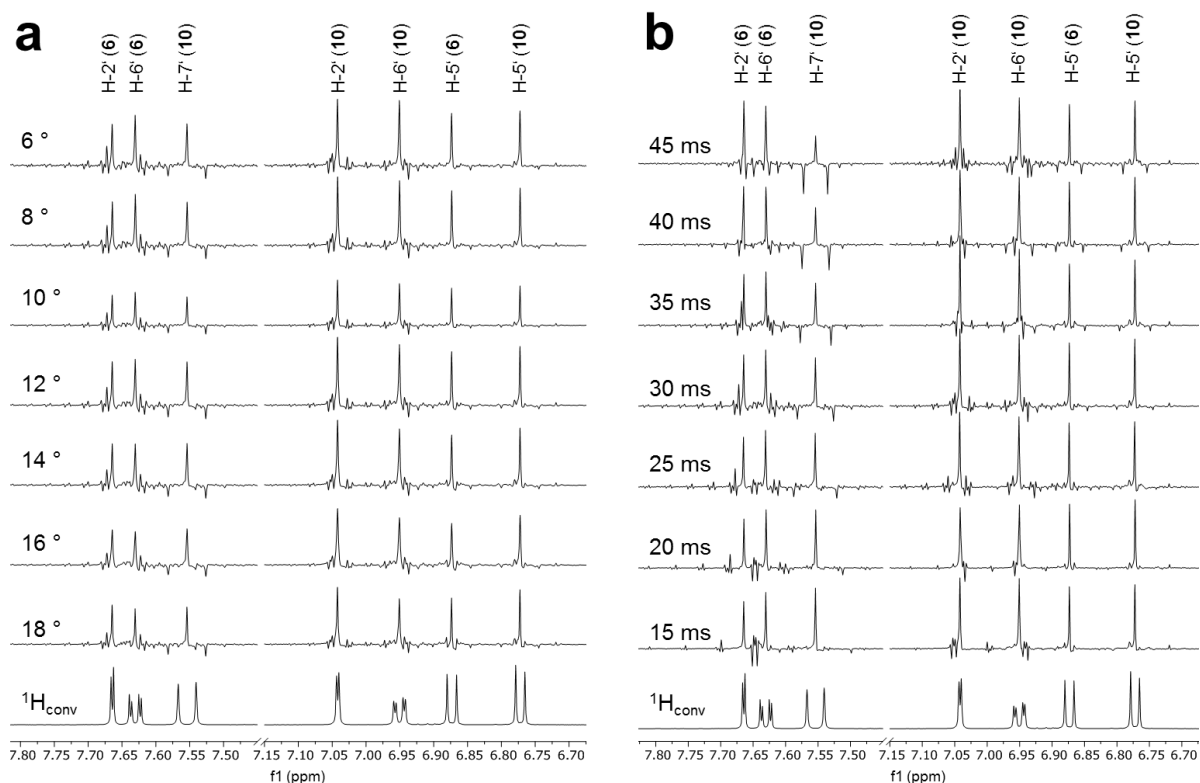


Figure 7.2 PSYCHE spectra section of compound mixture of chlorogenic acid (**10**) and rutin (**6**) with: (a) constant τ PS (30 ms) and varied pulse angle, and (b) constant pulse angle (10°) and varied τ PS.

In our study, we decided to use a swept pulse flip angle = 10° , τ PS = 15 ms, and a pulse width = 30 ms, in order to enable the decoupling of larger J from aromatic and olefinic protons of secondary metabolites. The decision was based on the taxonomic marker compounds described for *Hypericum* species, such as flavonoids, xanthenes, phloroglucinols, and organic acids.^{9,39,52,101}

Furthermore, the influence of homodecoupling on binning was investigated, where particularly large coupling constants are problematic due to a higher signal width.

In general, it can be concluded that the spectroscopist decides already which signals are detected better and which worse by setting the parameters. This is fine for structure elucidation and targeted approaches, but for untargeted metabolomics, it is definitely a drawback.

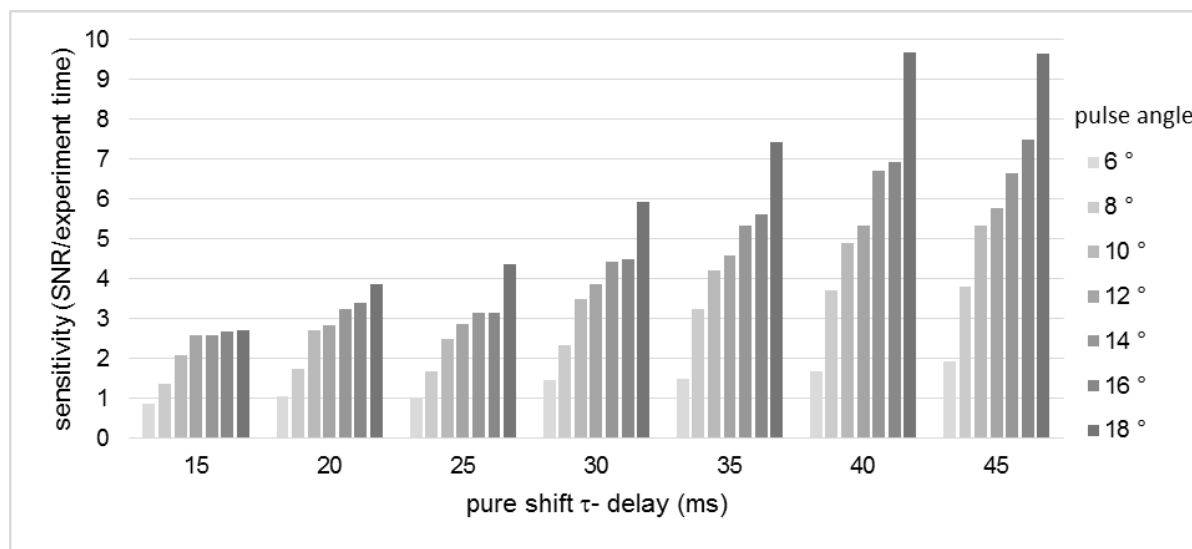


Figure 7.3 Effect of the variation of pulse angle and τ PS on the sensitivity (given as signal to noise ratio (SNR) per experiment time) of the PSYCHE spectra.

7.3.3 Suitability of PSYCHE and $^1\text{H-NMR}$ for metabolomics studies

It was investigated whether PSYCHE has advantages over a conventional $^1\text{H-NMR}$ during classical metabolomics processing with subsequent multivariate data analysis. 21 *Hypericum* species (Table 6.2) represented by 29 genotypes with up to three biological replicates were analyzed with quantitative $^1\text{H-NMR}$, conventional $^1\text{H-NMR}$, and PSYCHE experiment. In contrast to other Pure Shift methods, a PSYCHE spectrum contains quantitative information and thus can be used for metabolomics experiments.²⁴¹ In the metabolomics workflow used, spectra were baseline corrected and referenced. Then a uniform binning was applied, which divides the spectrum into bins of the same size over the whole spectral width. Finally, the binned spectra were normalized to the internal standard HMDS and evaluated by principal component analysis (PCA).

7.3.4 Optimization of bin size

Binning is a form of data reduction. Spectra are cut in sections (bin), and the total integral of each bin is used as evaluation value. Ideally, a bin completely includes one signal and represents the total integral; and thus, the amount of substance in the sample. The integrals measured in the quantitative $^1\text{H-NMR}$ and the corresponding bin value of the PSYCHE spectrum were compared to estimate the signal representation. The integration of the signals presupposes a baseline separation from other peaks. For comparison of both methods, eleven samples of five genotypes of *H. perforatum* (including biological replicates) were used. This ensures that no overlap of

signals chosen for quantification occurs, since intraspecifically the quantities of the constituents change but regularly, not the chemical composition. Nine baseline separated signals derived from seven constituents (Table 7.1), were integrated and compared to the bins of PSYCHE and conventional $^1\text{H-NMR}$ (examples shown in Figure 7.4). Figure 7.5 visualizes the procedure on the example of the doublet of the methyl group $\text{H}_3\text{-12}$ of hyperforin (**1**) at 1.08 ppm. Integrals and bin values were combined in an XY-diagram, and the coefficient of determination of the resulting regression line was used as a quality parameter. This procedure is inspired by Ludwig et al. (2010), who evaluated the quantitation of 2D J-resolved NMR experiments.²⁴⁶

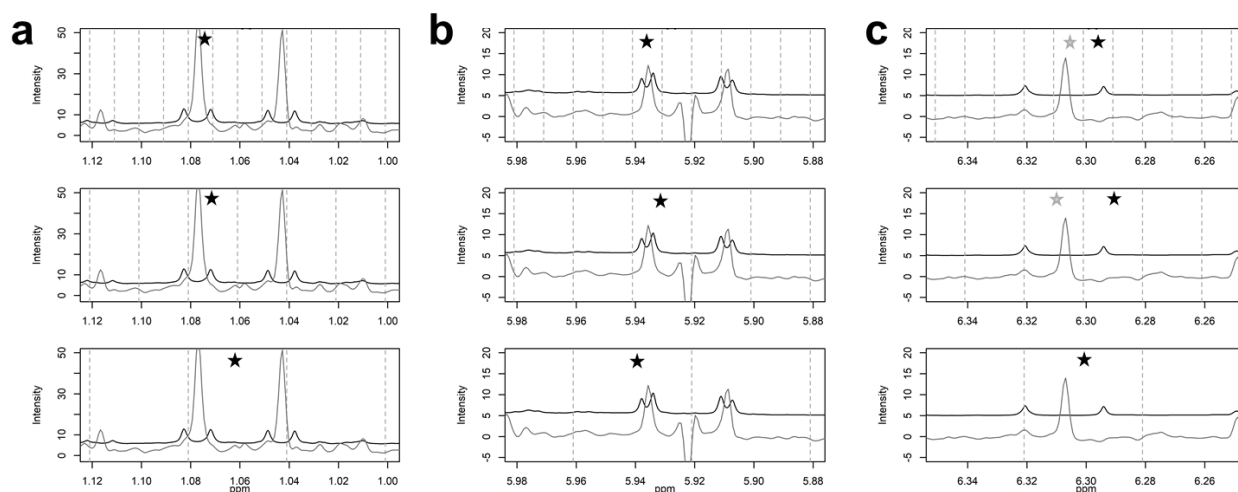


Figure 7.4 Selected sections of conventional $^1\text{H-NMR}$ (black) and PSYCHE (grey) spectra of a representative *Hypericum perforatum* sample evaluated with different bin sizes to display signal at: (a) 1.08 ppm (d, 6.5 Hz) $\text{H}_3\text{-12}$ of hyperforin (**5**); (b) 5.94 ppm (d, 2.4 Hz) H-6 of epicatechin/catechin (**4**); (c) 6.31 ppm (d, 15.8 Hz) H-8' of chlorogenic acid (**1**). For each signal, three bin sizes are shown (from up to down: 0.01, 0.02, and 0.04 ppm). The dashed vertical grey line marks the borders of each bin. The chosen bin for evaluation is highlighted with an asterisk.

For uniform binning, different bin sizes were used by different authors. We tested the regularly used bin sizes 0.01 ppm²⁵¹, 0.02 ppm²⁵², and 0.04 ppm²²⁹. In Figure 7.4 (extended in Appendix 7.1), it can be seen that the application of the bin size 0.01 ppm often results in split signals. In contrast, bins with boundaries of 0.04 ppm include multiple signals (Figure 7.4a, 0.04 ppm). Anderson et al. (2010) calculated that bins of 0.01 and 0.02 ppm enclose in average one peak per bin, whereas bins of 0.04 ppm contain four.²⁴⁸ To check if different bin sizes lead to a changing quantitative correlation to the concentration of the selected ingredients, the coefficient of determination (R^2) were compared (Table 7.1). This comparison was performed for conventional $^1\text{H-NMR}$ and PSYCHE experiments. No significant difference (Wilcoxon-Mann-Whitney, $p > 0.05$) between the bin sizes could be determined (Appendix 7.2). However, we decided to use 0.02 ppm for further processing because the R^2 average was highest, and the same size was successfully used in other studies.^{252,253}

One goal of this analysis was to check if PSYCHE may have advantages in comparison to conventional $^1\text{H-NMR}$, generated by uniform binning. Figure 7.6 shows the boxplots of the R^2 values, generated from the PSYCHE and the conventional $^1\text{H-NMR}$ experiments. The R^2 values

are not significantly different from each other (Wilcoxon-Mann-Whitney, $p > 0.05$). However, the data from conventional $^1\text{H-NMR}$ tends to reach slightly higher R^2 values. This indicates that the PSYCHE experiment has no advantage in terms of quantification during the binning process.

Table 7.1 Summary of the coefficient of determination (R^2) after linear regression between ^1H qNMR integrals and bin values of different experiments with bin size 0.02 ppm (Appendix 7.1).

Compound	Assignment	δ [ppm] multiplicity (J)	R^2 of experiment	
			$^1\text{Hconv}$	PSYCHE
Chlorogenic acid (10)	H-8'	6.31 d (15.8 Hz)	0.6553	0.8485
Chlorogenic acid (10)	H-2'	7.05 d (2.1 Hz)	0.7865	0.5748
Rutin (6)	H-6'''	1.12 d (6.2 Hz)	0.8176	0.7331
Hyperoside (7)	H-2'	7.83 (d 2.2 Hz)	0.9288	0.8810
Epicatechin/Catechin (14)	H-6	5.94 d (2.4 Hz)	0.9740	0.7974
Epicatechin/Catechin (14)	H-2'	6.97 d (1.9 Hz)	0.8104	0.7525
Hyperforin (1)	H ₃ -12	1.08 d (6.5 Hz)	0.8842	0.8380
Sucrose (12)	H-3'	4.09 d (8.2 Hz)	0.9945	0.9080
Shikimic acid (11)	H-4	4.36 m ($v_{1/2}$ 4.7 Hz)	0.8412	0.8706

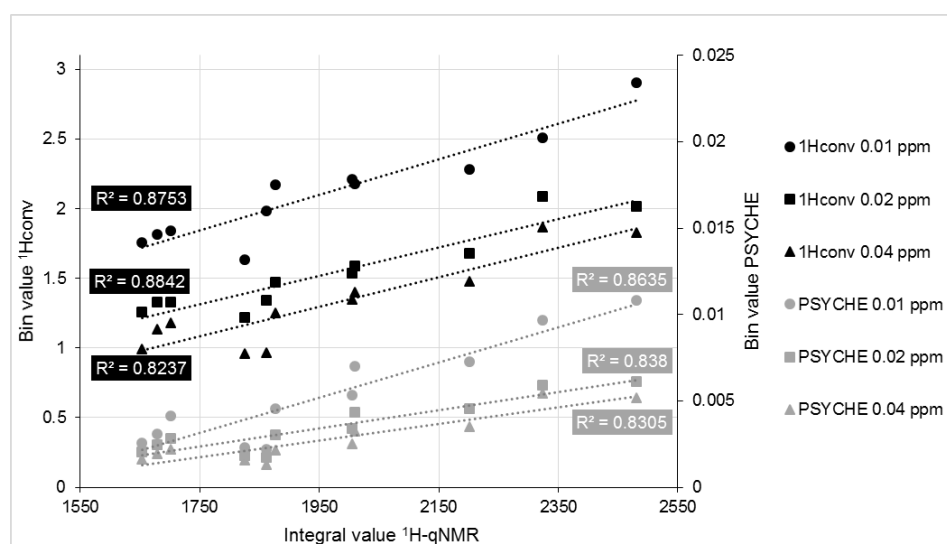


Figure 7.5 Linear regression between the integral of the signal at 1.08 ppm from the $^1\text{H-qNMR}$ measurement and the corresponding bin value of the conventional $^1\text{H-NMR}$ (1Hconv, black) and the PSYCHE (grey) experiment with different bin sizes. The coefficient of determination R^2 was calculated for each graph.

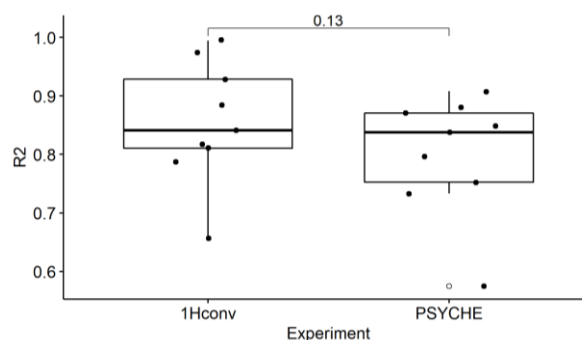


Figure 7.6 Boxplot of the coefficient of determination (R^2 , Table 1) for the conventional $^1\text{H-NMR}$ (1Hconv) and the PSYCHE experiment with a bin size of 0.02 ppm. Significance Test: Wilcoxon-Mann-Whitney paired (shown is p -value).

7.3.5 Multivariate data analysis of different *Hypericum* species

For 21 *Hypericum* species with up to three biological replicates, both a conventional $^1\text{H-NMR}$ and a PSYCHE spectrum were recorded. All data sets were identically processed, as described in the material and methods section. As already known from earlier studies,^{39,52,148,205,254} the interspecific variance within the genus *Hypericum* is huge. The obtained spectra are highly diverse (Appendix 6.1, Appendix 7.3), so no alignment or peak picking was carried out. The spectra are exceptionally varying in the region of olefinic and aromatic protons (5.5–8.0 ppm) and in signals belonging to phloroglucinol related resonances (1.0–2.0 ppm). Consequently, spectra were binned without alignment and peak picking to avoid mismatched signals. The scores plot of the PCA performed with conventional $^1\text{H-NMR}$ data is displayed in Figure 7.7a. The scores plot based on PSYCHE data demonstrate similar clusters and differ only marginally (Figure 7.7b).

Bin tables were freed from bins that did not exceed the limit of detection (LOD, calculated as three times the standard deviation of the noise regions (10.0–11.0 ppm)) to exclude bins with no informational content. In the case of PSYCHE data, due to signal narrowing by homodecoupling, the signal width is reduced, so that the signal will be expected to be distributed over a lower number of bins, ideally it is found in one bin only. However, the number of bins beyond the LOD was comparable for both methods. Within the bin table of the conventional $^1\text{H-NMR}$ data 9%, and in the case of PSYCHE data 10% of the bins could be removed. Utilizing the reduced data for PCA, no changes are observed in the first four principal components (PCs), indicating that the separation in PCA is independent of noise bins. It can be concluded that for the *Hypericum* data set, uniform binning with excluding noise data has general no advantage in comparison to conventional $^1\text{H-NMR}$. Santacruz et al. (2020) used adaptive intelligent binning and were able to reduce their initial data matrix of 168 bins from the conventional $^1\text{H-NMR}$ to 113 bins from the PSYCHE spectra.²³² Although they did not comment on that, this bin reduction could be a reason for better PLS results based on PSYCHE than conventional $^1\text{H-NMR}$ data.

Even if the noise bin reduction has no impact on the PCA result in the *Hypericum* data set, the PSYCHE spectra simplify the assignment of signals of interest. In Figure 7.8 three examples of interesting regions are highlighted. Regarding region A (Figure 7.8) two peaks at 1.55 ppm could correspond to one doublet with J 5.6 Hz, but in the PSYCHE spectrum, it is unambiguous that these are two singlets with similar chemical shifts. In contrast, the signal in region B (Figure 7.8) is a true doublet (J 6.5 Hz), which becomes a singlet by homodecoupling. Intriguing is also the region C (Figure 7.8), where the spectroscopist would initially assume a quartet, but the PSYCHE spectrum reveals that these are three overlapping signals with similar coupling constants. These few examples already show that the PSYCHE experiment can provide important information for the interpretation of overlapping regions.

The application of PSYCHE to complex samples like plant extracts is helpful for structural elucidation and reveals also combined with multivariate data analysis good results. However, the results are comparable to the conventional $^1\text{H-NMR}$, and therefore, the PSYCHE will never

replace the $^1\text{H-NMR}$ in this kind of analysis. Especially choosing the right parameters is difficult, so that still different types of artifacts are included. Additional optimization of the gradient amplitude and the duration of the pulse (for a given bandwidth) might also improve the results. Furthermore, new modified pulse sequences can even improve the spectral purity of the basic PSYCHE, like the triple spin-echo PSYCHE²⁴³ and the SAPPHIRE-PSYCHE²³⁴, which shows the great potential of these methods.

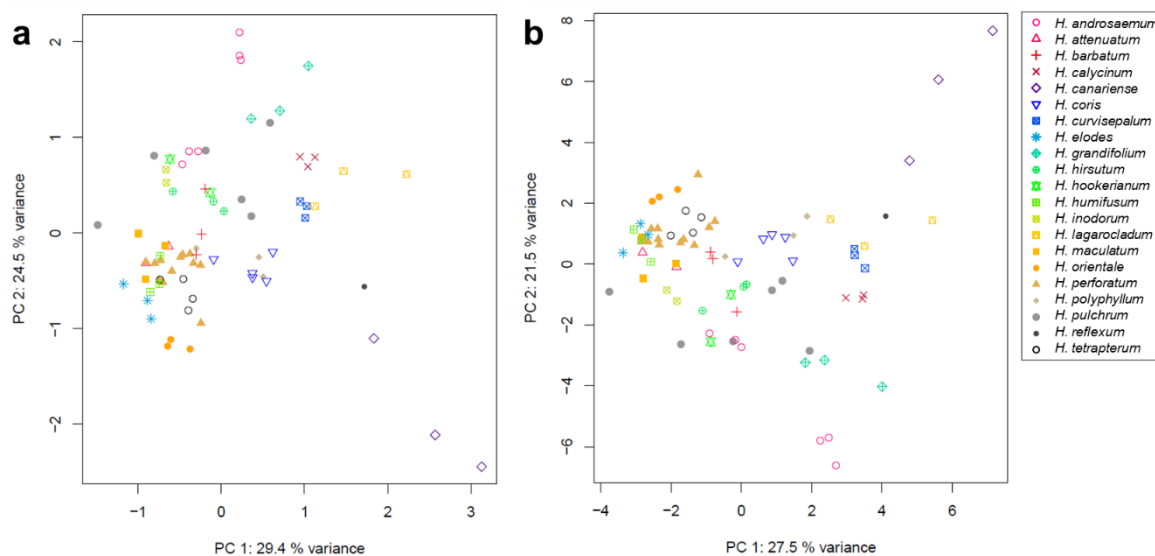


Figure 7.7 Scores plot of principal component analysis (PCA) of *Hypericum* species based on (a) conventional $^1\text{H-NMR}$ and (b) PSYCHE spectra.

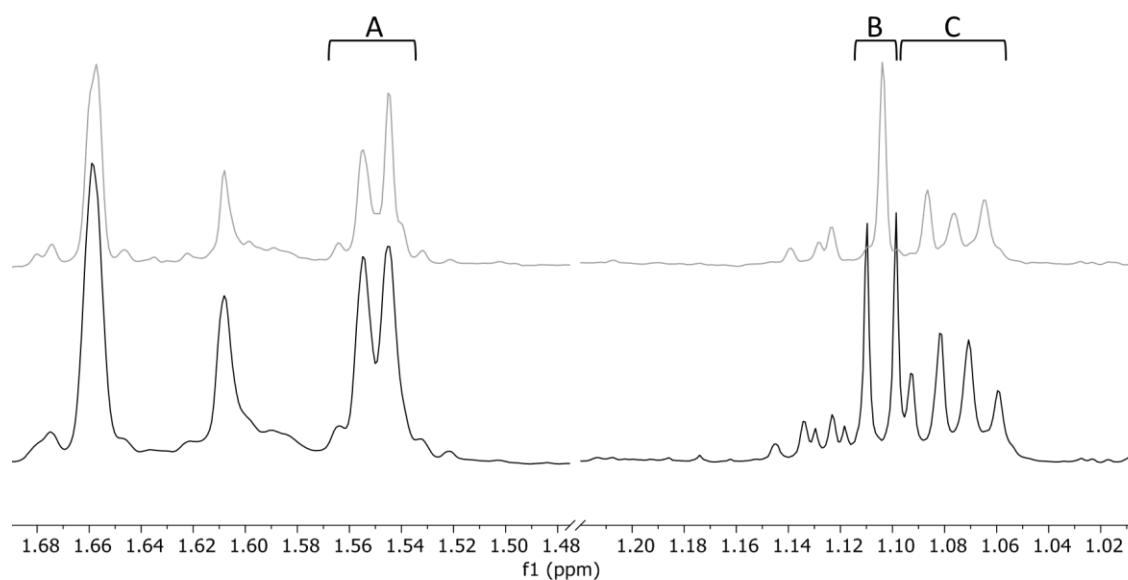


Figure 7.8 Regions of interest (A, B, C) in the conventional $^1\text{H-NMR}$ spectrum (black) of *Hypericum canariense* where the PSYCHE spectrum (grey) helps to identify signals and their multiplicity.

7.4 Conclusion

In this study, the PSYCHE experiment was optimized for *Hypericum* plant extracts, representing complex mixtures of various compound classes. The swept pulse angle and the pure shift τ -delay were adapted to get satisfactory results for coupling constants larger than 6 Hz. This requires a measuring time of 32 min, 16 times longer than the conventional $^1\text{H-NMR}$. However, the method presented here was much shorter than the ones reported before. The extraction of quantitative information was possible from the binned PSYCHE spectra and not significantly different to the results from the binned conventional $^1\text{H-NMR}$. With the uniform binning method, which is used for the spectra of compositionally different *Hypericum* extracts, the gain in resolution through homodecoupling did not affect multivariate data analyses such as PCA. So, the PCA of both methods leads to similar results. However, PSYCHE spectra are able to support the data interpretation and compound identification of NMR spectra by simplifying crowded spectral parts. This is the real power of Pure Shift methods. Therefore, we recommend to implement Pure Shift methods in the follow-up experiments of the NMR metabolomics workflow and use them to interpret the data.

8 General discussion and conclusion

Plants are in contrast to most other organisms bound to their location. Therefore, they stay in chemical contact with their environment, react, and adapt to changing conditions. This leads to an enormous wealth of plant secondary metabolites, which possess a variety of effects. Well-known from ethnomedicine, plant-based pharmaceuticals and dietary supplements are gaining more importance (2016 US sales increased 7.7%).¹ Often full extracts, i.e. complex mixtures of primary and secondary metabolites, are administered. Those compound combinations sometimes show a greater effect than the isolated active ingredient alone,²⁰¹ due to synergistic effects or better availability of the active ingredient by improving its solubility.

Hypericum perforatum (St. John's Wort) represents the most important antidepressant herb grown commercially. Furthermore, also divers other effects were reported. The valuable compounds mainly responsible for its activity are phloroglucinols, naphthodianthrones, and flavonoids. To guarantee consistent biological effects, the production of complex mixtures with a constant composition is necessary. The success depends highly on the used plant material and the extraction method. The starting plant material underlies various fluctuations caused by abiotic or biotic influences. In addition, different reproduction strategies are known for *Hypericum*, and intercrossing offers further genetic variability. The main goal of this thesis was to characterize the metabolite variance caused by the genetic settings only, i.e. as independent as possible from environmental factors. Therefore, the plants used were cultivated and sampled under identical conditions, which is not the case for comparative studies where plants are collected from the wild.⁶⁰⁻⁶² Further, those studies are typically restricted to the main compounds, whereas as strength of this work, an untargeted analysis was performed to include also minor abundant compounds present in *Hypericum*.

Metabolite variance of *Hypericum perforatum*

The production of reproducible pharmaceutical *H. perforatum* extracts is challenging because of the intraspecific variation. The metabolite composition varies greatly depending on the genotype (GT) used, as the study in chapter 2 has shown. From the 20 pharmaceutical extracts, generated from the aerial parts of different GTs, some lacked characteristic main constituents (European Pharmacopoeia¹⁵), such as the flavonoid glycoside rutin (**6**), while one GT exhibited a completely different phloroglucinol profile. This compositional change could have severe consequences as most neuroactive effects are attributed to phloroglucinols, such as hyperforin (**1**) and adhyperforin (**2**).^{82,168,255,256} Furthermore, rutin (**6**) is reported as a crucial ingredient for the antidepressant activity of *Hypericum* extracts, causing bioavailability.¹⁷⁴ The lack of these ingredients in a pharmaceutical drug could lead to a drastic loss of efficacy.

This first study (chapter 2) could show how variable *H. perforatum* ingredients are within pharmaceutical extracts of different GTs. However, the small sample set, the lack of biological

replicates, and the long preparation process were not suitable to answer the following raised scientific questions sufficiently:

1. What is the difference in the metabolite profiles of GTs from North America and Europe?
2. Are the variations of the constituents in leaves also reflected in the flowers?
3. Is the genetic variance reflected in the metabolite profiles of the GTs?
4. Can GTs with very different chemotypes in combination with transcriptomic analyses be used to elucidate biosynthetic processes?

In order to answer these questions, a meaningful data set containing leaves and flowers of 93 *H. perforatum* GTs (originating from North America and Europe and cultivated under the same conditions) was sampled in parallel for metabolomics and transcriptomics analysis. Furthermore, biological replicates were produced to distinguish the variance between individuals from those of GTs (chapter 4).

The variances in the metabolite profiles of different GTs described in chapter 2 were confirmed in the more extensive data set, showing that more than 20% of the investigated GTs contain no rutin (**6**) or a strongly reduced rutin content. Three of the GTs exhibited a significantly different phloroglucinol profile characterized by the absence of highly prenylated phloroglucinols such as hyperforin (**1**) and adhyperforin (**2**). Instead, the occurrence of less prenylated precursors (e.g. **15**, **16**) was observed. Furthermore, astilbin (**19**) and acetylated flavonoid glycosides (chapter 4: P13, P14, P23, P24) influenced the discrimination of the GTs.

The metabolic differences between the GTs could be observed equally in both organs, leaves and flowers. However, it should be emphasized that the flavonoid and the phloroglucinol composition differed between the organs. In contrast, the naphthodianthrone composition was organ-independent.

The study indicated that genetic characteristics such as ploidy, genetic background⁶⁴, and reproduction strategy were not related to the variance of the metabolite profiles. Further, the compositional differences cannot be explained by the origin of the plants from the different continents. Interestingly, samples from close sample sites possess similar metabolite profiles. An exception is the Great Lake region in North America, where a big metabolite variance can be observed, which is in line with the assumption of Molins et al. (2014)⁶⁴, who identified the area as genetically highly variable.

Chapter 3 proved that comparative studies of GTs are ideal for transcriptomic analysis. The investigation of pistil tissue from different GTs with and without dark glands allowed the identification of compounds involved in hypericin biosynthesis, based on PLS analysis. The transcriptome analysis confirmed two potential enzymes involved in hypericin formation (POCP, OKS) and suggested BBE to catalyze the first dehydrogenation step that forms the second bond connecting the anthrone subunits to yield protohypericins (**18**, **27**) (Figure 3.4).⁶⁷ The correlation

of skyrin-glycosides (**29**, **30**) to dark gland occurrence could be a possible indication to a biosynthetic relation of skyrin and hypericin. This would suggest a first linkage of the two anthrone halves via the 5 and 5' bond leading to penicilliopsin (**23**) and hydroxyenicilliopsin (**25**). However, intensive MS/MS studies (chapter 4) gave hints that the fragmentation behavior fits better to compounds with 10-10' linkage, such as emodin dianthrone (**24**) and hydroxy emodin dianthrone (**26**). The question remains open if the hypericin and skyrin biosyntheses are connected. However, recent MALDI investigations¹⁸³ show that skyrin derivatives (e.g. **29**) are found in the dark glands or in the immediate vicinity. In addition, two precursors of skyrin have been proposed,¹⁰¹ which would imply the independence of the biosynthesis of the two compounds (skyrin and hypericin) and just a co-location in the dark glands. A general problem of mass spectrometric investigations of naphthodianthrone represents the identical molecular composition of some flavonoids and anthrones, e.g. emodin corresponds to apigenin and skyrin to biapigenin or amentoflavone. Also, the fragmentation is often identical on the MS² level, which makes chromatographic separation indispensable for clear identification. To confirm the proposed precursors unambiguously, isolation and unequivocal structure determination is necessary.

The high sample number and the constant naphthodianthrone composition of the leaf and flower data from the multi-genotype study also enabled correlation analysis among the metabolites (chapter 4). Besides the already described compounds, two minor compounds correlated to hypericin (**3**) and pseudohypericin (**4**). Based on MS/MS investigations, compounds **35** and **36** were tentatively assigned. Further, emodin-8-*O*-glucoside (**34**) was correlated, which could be the water-soluble transportable form of the precursor emodin (**20**). Beyond the focus on the hypericin biosynthesis, the untargeted study design allows the usage of the acquired data to answer additional questions.

The correlation network analysis showed that correlation occur in particular, within the compound classes of phloroglucinols and naphthodianthrone. Those metabolites are stored in the translucent and dark glandular structures of *Hypericum* plants, respectively. In order to predict the content of these valuable ingredients, an automated gland counting tool was developed. Using two phenotypes, it was shown that the hypericin content was higher with a rising number of dark glands and hyperforin was increased in the phenotype with more translucent glands (chapter 5). The developed gland counting tool could serve as a rapid screening method to predict valuable genotypes.

Metabolite variance of different *Hypericum* species

H. perforatum is almost the only commercially used *Hypericum* species, although for an increasing number of species active compounds were isolated. The metabolite profiling of 21 species exhibited in particular polyprenylated acylphloroglucinols as species-specific constituents, well suited to distinguish between species (chapter 6). Interestingly, this cannot be applied to *H. perforatum*, because the most prominent phloroglucinol hyperforin (**1**) is also present in other

members of the genus (e.g. *H. curvisepalum*, *H. lagarocladum*). The only unique compound for *H. perforatum* was rutin (**6**), making it – pending further species studies – suitable as marker compound, as suggested in the European Pharmacopeia. However, it must be pointed out again that the absence of rutin does not necessarily mean that *H. perforatum* is not present since some GTs did not contain this compound either (chapter 2 and 4).

Within the exploration of the bioactivity some of the investigated *Hypericum* species possess antibacterial effects against gram-positive and gram-negative bacteria (chapter 6). Since the metabolite profiles of these species differ significantly, it is assumed that different active compounds are responsible for their bioactivity. To determine possible structure-activity relationships, data from the 2D NMR method HMBC was correlated to the growth inhibition of *A. fischeri* utilizing an Activity Correlation Analysis (ACorA). In general, the correlated structural elements point out to prenylated phloroglucinols. In *H. canariense* and *H. reflexum* 2-*O*-geranyl-methylpropanoyl-phloroglucinol (**41**) was identified as a candidate. This compound, firstly isolated from *H. punctatum*, has been described as active against bacterial growth before.⁶⁵ To validate the results and determine details of its particular bioactivity, compound **41** should be isolated or synthesized. Detailed phytochemical studies should be performed to identify the active phloroglucinols from *H. orientale* and *H. elodes*. In general, it can be concluded that many *Hypericum* species are still not or not sufficiently investigated, although they are likely to contain numerous promising bioactive constituents.

Analytical aspects of the metabolite investigations

In this thesis, TLC, NMR, and LC-MS methods were utilized to analyze the metabolite profiles of *Hypericum*. TLC analysis as a rapid screening method is implemented in the European Pharmacopeia. It provides information about the presence of naphthodianthrones such as hypericin (**3**), flavonoid glycosides (**6-9**), and organic acids (**10**). However, the simultaneous detection of phloroglucinols is not possible. NMR represents a useful tool for detecting all main constituents, comprising primary metabolites and secondary metabolites such as flavonoids, organic acids, xanthenes, and phloroglucinols within crude plant extracts. The low sensitivity, nevertheless, limits the detection of minor compounds. In addition to the low concentration, the keto-enol-tautomerism prevents hypericins (**3**, **4**, **18**, **27**) from being detected. In contrast to that, LC-MS allows the detection of most secondary metabolites due to its high sensitivity. However, coelution can lead to ion suppression and, due to the different levels of ionization, over- and under-representation of compounds can occur. The combination of different methods, as shown in chapters 2 and 6, allows a comprehensive assessment of the chemical fingerprints of *Hypericum* extracts.

The generation of statistically meaningful data sets requires different processing steps depending on the analytical method. The processing of 1D NMR spectra is comparatively simple, because it is limited to two-dimensional information (intensity and proton chemical shift δ_H) and a

manageable number of variables is generated. In contrast, 2D NMR methods, such as HSQC and HMBC contain three-dimensional information (δ_H , δ_C , intensity). Therefore, these spectra are divided into a high number of pixels (chapter 6). To reduce the number of variables, only pixels exceeding the limit of detection were used for evaluation (chapter 6). $^1\text{H-NMR}$ has the great advantage of being a quantitative method. This is also true for the 2D NMR experiment HSQC. However, in HMBC, the signal intensity depends on the strength of the carbon-proton coupling. Thus, the use in metabolomics experiments is not very common. The loss of quantifiability is definitely a disadvantage, but the comparability of the signals between different samples is still given. This is in analogy to the different ionization efficiencies of the compounds in mass spectrometry. In contrast to the HSQC (J^1), the HMBC shows proton-carbon couplings via two, three, or more bonds, so essentially more information per proton signal. Furthermore, quaternary signals are covered, which are naturally not detectable with HSQC experiments. Therefore, mainly HMBC was used for the metabolomics experiments to gain additional structural information (chapter 6). The acquisition of the 2D NMR spectra with non-uniform sampling (NUS, 50%) significantly shortened the experiment time, making the setup more attractive for large data sets.²⁵⁷

In 2D NMR experiments, the resolution of the proton signals is improved by spreading the signals in the additional δ_C dimension. Other methods to reduce the overlap of NMR signals in complex $^1\text{H-NMR}$ spectra are Pure Shift methods. Chapter 7 shows that PSYCHE is a suitable method to simplify most multiplets to singlets. However, the increase in resolution led in multivariate statistics to results comparable to the conventional $^1\text{H-NMR}$ experiment. Nevertheless, adding PSYCHE to the metabolomics workflow can be an enrichment reducing the complexity of overlapping signals. Therefore, the structure identification and interpretation of the results is accelerated. Although the method possesses great potential for metabolomics studies,^{232,234,245} it might be more powerful for less diverse samples, where peak picking or intelligent binning can be applied. The recent and future developments of the PSYCHE pulse sequence will lead to better results concerning signal-to-noise ratio and artifact suppression, also leading to shorter measurement time.^{234,235}

Due to the high sensitivity of LC-MS measurements, various metabolites can be detected simultaneously. It has to be mentioned that the method development is much more time-consuming. Due to the low robustness, the analysis of huge numbers of samples in a comparable way is critical and often leads to batch effects. The number of samples in one batch was limited in our case by deterioration in mass accuracy of the Orbitrap MS system during acquisition. This issue could be addressed by using a different MS-system with an included automatic recalibration. However, also in this kind of MS-systems, the number of samples is limited because cleaning steps, caused by the complex matrix of the crude plant extracts, are necessary. These batch effects can be corrected afterwards with batch correction algorithms (chapters 4 and 6), but it is an additional processing step with the possibility to introduce mistakes.

Another issue of analyzing *Hypericum* extracts by liquid chromatography is the abundance of naphthodianthrones. On the one hand, those compounds are poorly soluble once they are separated from the matrix. On the other hand, the naphthodianthrones interact strongly with the stationary phase, wherefore a long flushing time with 100% organic solvent of the column is required to circumvent carry over. In addition, those compounds are highly sensitive to pH changes, showing retention time shifts during long batches due to changes in the mobile phase over time. Therefore buffered solvent systems were used to counteract possible misalignment.

Finally, in the LC-MS data set, each variable is characterized by two values, the m/z value and the retention time. This makes peak picking possible and simplifies the evaluation method because each feature stands theoretically for one compound, neglecting adducts and isotopes. The bottleneck of this method is the identification of the thousands of detected signals. Apart from the comparison with reference compounds, only putative assignment is possible.

In summary, this thesis contributes to the major understanding of the complexity of natural products within *Hypericum*. The interaction of advanced analytical techniques, such as LC-MS and NMR, combined with multivariate data analysis, gives in-depth insights into the different levels of variance. This information can be used to determine valuable properties of extracts, identify marker compounds, get insights into biosynthetic pathways, select breeding lines for specific applications, or identify targets for further phytochemical investigations.

References

1. Smith T, Kawa K, Eckl V, Morton C and Stredneyd R. Herbal Supplement Sales in US Increase 7.7% in 2016. *Market report* **2017**. 115:56-65.
2. Newman DJ and Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **2020**. 83:770-803.
3. Paracelsus. *Einliche Tractat: 1. von natürlichen Dingen, 2. Beschreibung etlicher Kreuter*. Müller, **1582**.
4. Robson NKB. **2003**. *Hypericum botany*. In *Hypericum - The genus Hypericum*. Ernst E (Ed.) Taylor and Francis: London and New York.
5. Wurdack KJ and Davis CC. Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *Am J Bot* **2009**. 96:1551-1570.
6. Ruhfel BR, Bittrich V, Bove CP, Gustafsson MH, Philbrick CT, Rutishauser R, Xi Z and Davis CC. Phylogeny of the clusoid clade (Malpighiales): evidence from the plastid and mitochondrial genomes. *Am J Bot* **2011**. 98:306-325.
7. Stevens PF. **2007**. Hypericaceae. In *The Families and Genera of Vascular Plants*, Vol 11. Kubitzki K (Ed.) Springer Verlag: Berlin, Heidelberg, 194-201.
8. Nürk NM and Crockett SL. Morphological and Phytochemical Diversity among *Hypericum* Species of the Mediterranean Basin. *Med Aromat Plant Sci Biotechnol* **2011**. 5:14-28.
9. Crockett SL and Robson NKB. Taxonomy and Chemotaxonomy of the Genus *Hypericum*. *Med Aromat Plant Sci Biotechnol* **2011**. 5:1-13.
10. Barnes J, Anderson LA and Phillipson JD. St John's wort (*Hypericum perforatum* L.): a review of its chemistry, pharmacology and clinical properties. *J Pharm Pharmacol* **2001**. 53:583-600.
11. Hofrichter J, Krohn M, Schumacher T, Lange C, Feistel B, Walbroel B, Heinze H-J, Crockett S, Sharbel TF and Pahnke J. Reduced Alzheimer's disease pathology by St. John's wort treatment is independent of hyperforin and acilitated by ABCC1 and microglia activation in mice. *Curr Alzheimer Res* **2013**. 10:1057-1069.
12. Sutar I, Oyardi O, Akkol EK and Ozcelik B. Antimicrobial effect of the extracts from *Hypericum perforatum* against oral bacteria and biofilm formation. *Pharm Biol* **2016**. 54:1065-1070.
13. Mayer JG. **2014**. Johanniskraut ist Arzneipflanze des Jahres 2015. Würzburg; URL [<https://www.uni-wuerzburg.de/aktuelles/einblick/single/news/johanniskr/>]; accessed 30.10.2020
14. Vogel M. **2018**. Heilpflanze des Jahres 2019 ist das Johanniskraut. NHV Theophrastus; URL [https://nhv-theophrastus.de/site/index.php?option=com_content&view=article&id=268:heilpflanze-des-jahres-2019-ist-das-johanniskraut-&catid=36:presstexte&Itemid=56];
15. European Pharmacopeia 10.0 In *St John's Wort*, **2017**. Vol 01/2017:1438. Directorate for the Quality of Medicines & HealthCare of the Council of Europe: 1631-1633.
16. Zhang R, Ji Y, Zhang X, Kennelly EJ and Long C. Ethnopharmacology of *Hypericum* species in China: A comprehensive review on ethnobotany, phytochemistry and pharmacology. *J Ethnopharmacol* **2020**. 254:112686.
17. Tanemossu SA, Franke K, Arnold N, Schmidt J, Wabo HK, Tane P and Wessjohann LA. Rare biscoumarin derivatives and flavonoids from *Hypericum riparium*. *Phytochemistry* **2014**. 105:171-177.
18. Tatsis EC, Boeren S, Exarchou V, Troganis AN, Vervoort J and Gerothanassis IP. Identification of the major constituents of *Hypericum perforatum* by LC/SPE/NMR and/or LC/MS. *Phytochemistry* **2007**. 68:383-393.
19. Sirvent T, Krasnoff SB and Gibson DM. Induction of hypericins and hyperforins in *Hypericum perforatum* in response to damage by herbivores. *J Chem Ecol* **2003**. 29:2667-2681.
20. Guillet G, Podeszinski C, Regnault-Roger C, Arnason JT and Philogène BJR. Behavioral and Biochemical Adaptations of Generalist and Specialist Herbivorous Insects Feeding on *Hypericum perforatum* (Guttiferae). *Environ Entomol* **2000**. 29:135-139.
21. Dudek-Peric AM, Golab J, Garg AD and Agostinis P. Melanoma targeting with the loco-regional chemotherapeutic, Melphalan: From cell death to immunotherapeutic efficacy. *Oncoimmunology* **2015**. 4:e1054600.
22. Garg AD, Krysko DV, Vandenabeele P and Agostinis P. Hypericin-based photodynamic therapy induces surface exposure of damage-associated molecular patterns like HSP70 and calreticulin. *Cancer Immunol Immunother* **2012**. 61:215-221.
23. Garg AD, Nowis D, Golab J and Agostinis P. Photodynamic therapy: illuminating the road from cell death towards anti-tumour immunity. *Apoptosis* **2010**. 15:1050-1071.

24. Krysko O, Love Aaes T, Bachert C, Vandenabeele P and Krysko DV. Many faces of DAMPs in cancer therapy. *Cell Death Dis* **2013**. 4:e631.
25. Košuth J, Smelcerovic A, Borsch T, Zuehlke S, Karppinen K, Michael S, Hohtola A and Cellárová E. The hyp-1 gene is not a limiting factor for hypericin biosynthesis in the genus *Hypericum*. *Funct Plant Biol* **2011**. 38:35-43.
26. Gill M and Giménez A. Austrovenetin, the principal pigment of the toadstool *Dermocybe austroveneta*. *Phytochemistry* **1991**. 30:951-955.
27. Soták M, Czerankova O, Klein D, Jurcackova Z, Li L and Cellarova E. Comparative Transcriptome Reconstruction of Four *Hypericum* Species Focused on Hypericin Biosynthesis. *Front Plant Sci* **2016**. 7:1039.
28. Bridi H, Meirelles GC and von Poser GL. Structural diversity and biological activities of phloroglucinol derivatives from *Hypericum* species. *Phytochemistry* **2018**. 155:203-232.
29. Yang XW, Grossman RB and Xu G. Research Progress of Polycyclic Polyprenylated Acylphloroglucinols. *Chem Rev* **2018**. 118:3508-3558.
30. Soelberg J, Jorgensen LB and Jager AK. Hyperforin accumulates in the translucent glands of *Hypericum perforatum*. *Annals of botany* **2007**. 99:1097-1100.
31. Beerhues L. Hyperforin. *Phytochemistry* **2006**. 67:2201-2207.
32. Paniego NB, Zuubier KWM, Fung S-Y, van der Heijden R, Scheffer JJC and Verpoorte R. Phlorisoverlerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *Eur J Biochem* **1999**. 262:612-616.
33. Adam P, Arigoni D, Bacher A and Eisenreich W. Biosynthesis of Hyperforin in *Hypericum perforatum*. *J Med Chem* **2002**. 45:4786-4793.
34. Klundt T, Bocola M, Lutge M, Beuerle T, Liu B and Beerhues L. A single amino acid substitution converts benzophenone synthase into phenylpyrone synthase. *J Biol Chem* **2009**. 284:30957-30964.
35. Karppinen K, Hokkanen J, Tolonen A, Mattila S and Hohtola A. Biosynthesis of hyperforin and adhyperforin from amino acid precursors in shoot cultures of *Hypericum perforatum*. *Phytochemistry* **2007**. 68:1038-1045.
36. Klingauf P, Beuerle T, Mellenthin A, El-Moghazy SA, Boubakir Z and Beerhues L. Biosynthesis of the hyperforin skeleton in *Hypericum calycinum* cell cultures. *Phytochemistry* **2005**. 66:139-145.
37. Boubakir Z, Beuerle T, Liu B and Beerhues L. The first prenylation step in hyperforin biosynthesis. *Phytochemistry* **2005**. 66:51-57.
38. Dinamarca MC, Cerpa W, Garrido J, Hancke JL and Inestrosa NC. Hyperforin prevents beta-amyloid neurotoxicity and spatial memory impairments by disaggregation of Alzheimer's amyloid-beta-deposits. *Mol Psychiatry* **2006**. 11:1032-1048.
39. Kucharíková A, Kusari S, Sezgin S, Spitteller M and Cellarova E. Occurrence and Distribution of Phytochemicals in the Leaves of 17 In vitro Cultured *Hypericum* spp. Adapted to Outdoor Conditions. *Front Plant Sci* **2016**. 7:1-15.
40. Chrubasik-Hausmann S, Vlachoianis J and McLachlan AJ. Understanding drug interactions with St John's wort (*Hypericum perforatum* L.): impact of hyperforin content. *J Pharm Pharmacol* **2019**. 71:129-138.
41. Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genom* **2001**. 2:155-168.
42. Wolfender J-L, Marti G, Thomas A and Bertrand S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J Chromatogr A* **2015**. 1382:136-164.
43. Kim HK, Choi YH and Verpoorte R. NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol* **2011**. 29:267-275.
44. Pinto CR. **2017**. Chemometrics Methods and Strategies in Metabolomics. In *Metabolomics: From Fundamentals to Clinical Application*. Sussulini A (Ed.) Springer
45. Wold S, Sjöström M and Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* **2001**. 58:109-130.
46. Worley B and Powers R. Multivariate Analysis in Metabolomics. *Curr Metabolomics* **2013**. 1:92-107.
47. Ioku K, Tsushida T, Takei Y, Nakatani N and Terao J. Antioxidative activity of quercetin and quercetin monoglucosides in solution and phospholipid bilayers. *Biochim Biophys Acta* **1995**. 1234:99-104.
48. Sosa S, Pace R, Bornancin A, Morazzoni P, Riva A, Tubaro A and Della Loggia R. Topical anti-inflammatory activity of extracts and compounds from *Hypericum perforatum* L. *J Pharm Pharmacol* **2007**. 59:703-709.
49. Tian J, Zhang F, Cheng J, Guo S, Liu P and Wang H. Antidepressant-like activity of adhyperforin, a novel constituent of *Hypericum perforatum* L. *Sci Rep* **2014**. 4:5632.

-
50. Velingkar VS, Gupta GL and Hegde NB. A current update on phytochemistry, pharmacology and herb–drug interactions of *Hypericum perforatum*. *Phytochem Rev* **2017**. 16:725-744.
 51. Butterweck V, Christoffel V, Nahrstedt A, Petereit F, Spengler B and Winterhoff H. Step by step removal of hyperforin and hypericin: activity profile of different *Hypericum* preparations in behavioral models. *Life Sci* **2003**. 73:627-639.
 52. Porzel A, Farag MA, Mülbradt J and Wessjohann LA. Metabolite profiling and fingerprinting of *Hypericum* species: a comparison of MS and NMR metabolomics. *Metabolomics* **2014**. 10:574-588.
 53. Nahrstedt A and Butterweck V. Biologically active and other chemical constituents of the herb of *Hypericum perforatum* L. *Pharmacopsychiatry* **1997**. 30 Suppl 2:129-134.
 54. Buytaert E, Callewaert G, Hendrickx N, Scorrano L, Hartmann D, Missiaen L, Vandenneede JR, Heirman I, Grooten J and Agostinis P. Role of endoplasmic reticulum depletion and multidomain proapoptotic BAX and BAK proteins in shaping cell death after hypericin-mediated photodynamic therapy. *FASEB J* **2006**. 20:756-758.
 55. Zobayed SM, Afreen F, Goto E and Kozai T. Plant-environment interactions: Accumulation of hypericin in dark glands of *Hypericum perforatum*. *Annals of botany* **2006**. 98:793-804.
 56. Raclariu AC, Paltinean R, Vlase L, Labarre A, Manzanilla V, Ichim MC, Crisan G, Brysting AK and de Boer H. Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding, TLC and HPLC-MS. *Sci Rep* **2017**. 7:1-12.
 57. Farag MA and Wessjohann LA. Metabolome classification of commercial *Hypericum perforatum* (St. John's Wort) preparations via UPLC-qTOF-MS and chemometrics. *Planta Med* **2012**. 78:488-496.
 58. Booker A, Agapouda A, Frommenwiler DA, Scotti F, Reich E and Heinrich M. St John's wort (*Hypericum perforatum*) products - an assessment of their authenticity and quality. *Phytomedicine : international journal of phytotherapy and phytopharmacology* **2018**. 40:158-164.
 59. Scotti F, Lobel K, Booker A and Heinrich M. St. John's Wort (*Hypericum perforatum*) Products - How Variable Is the Primary Material? *Front Plant Sci* **2018**. 9:1973.
 60. Bozin B, Kladar N, Grujic N, Anackov G, Samojlik I, Gavaric N and Conic BS. Impact of origin and biological source on chemical composition, anticholinesterase and antioxidant properties of some St. John's wort species (*Hypericum* spp., Hypericaceae) from the Central Balkans. *Molecules (Basel, Switzerland)* **2013**. 18:11733-11750.
 61. Rusalepp L, Raal A, Pussa T and Maeorg U. Comparison of chemical composition of *Hypericum perforatum* and *H. maculatum* in Estonia. *Biochem Syst Ecol* **2017**. 73:41-46.
 62. Zeljkovic SC, Karalija E, Paric A, Muratovic E and Tarkowski P. Environmental Factors do not Affect the Phenolic Profile of *Hypericum perforatum* Growing Wild in Bosnia and Herzegovina. *Natural product communications* **2017**. 12:1465-1468.
 63. Bagdonaitė E, Mártonfi P, Repčák M and Labokas J. Variation in concentrations of major bioactive compounds in *Hypericum perforatum* L. from Lithuania. *Ind Crop Prod* **2012**. 35:302-308.
 64. Molins MP, Corral JM, Aliyu OM, Koch MA, Betzin A, Maron JL and Sharbel TF. Biogeographic variation in genetic variability, apomixis expression and ploidy of St. John's wort (*Hypericum perforatum*) across its native and introduced range. *Annals of botany* **2014**. 113:417-427.
 65. Sarkisian SA, Janssen MJ, Matta H, Henry GE, Laplante KL and Rowley DC. Inhibition of bacterial growth and biofilm production by constituents from *Hypericum* spp. *Phytother Res* **2012**. 26:1012-1016.
 66. Ernst E. *Medicinal and Aromatic Plants - Industrial Profiles*. Taylor and Francis, **2003**: 115.
 67. Rizzo P, Altschmied L, Stark P, Rutten T, Gundel A, Scharfenberg S, Franke K, Baumlein H, Wessjohann L, Koch M, Borisjuk L and Sharbel TF. Discovery of key regulators of dark gland development and hypericin biosynthesis in St. John's Wort (*Hypericum perforatum*). *Plant Biotechnol J* **2019**. 17:2299-2312.
 68. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R and Human Serum Metabolome C. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2011**. 6:1060-1083.
 69. Tautenhahn R, Patti GJ, Rinehart D and Siuzdak G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* **2012**. 84:5035-5039.
 70. Stacklies W, Redestig H, Scholz M, Walther D and Selbig J. pcaMethods--a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **2007**. 23:1164-1167.
 71. Passos CS, Simoes-Pires CA, Nurisso A, Soldi TC, Kato L, de Oliveira CM, de Faria EO, Marcourt L, Gottfried C, Carrupt PA and Henriques AT. Indole alkaloids of Psychotria as multifunctional cholinesterases and monoamine oxidases inhibitors. *Phytochemistry* **2013**. 86:8-20.

72. Wangenstein H, Samuelsen AB and Malterud KE. Antioxidant activity in extracts from coriander. *Food Chem* **2004**. 88:293-297.
73. Dos Santos CHC, de Carvalho MG, Franke K and Wessjohann L. Dammarane-type triterpenoids from the stem of *Ziziphus glaziovii* Warm. (Rhamnaceae). *Phytochemistry* **2019**. 162:250-259.
74. Avato P and Guglielmi G. Determination of Major Constituents in St. John's Wort Under Different Extraction Conditions. *Pharm Biol* **2008**. 42:83-89.
75. Mártonfi P, Repčák M and Mártonfiová L. Secondary metabolites during ontogenetic phase of reproductive structures in *Hypericum maculatum*. *Biologia* **2006**. 61:473-478.
76. Umek A, Kreft S, Kartnig T and Heydel B. Quantitative phytochemical analyses of six hypericum species growing in slovenia. *Planta Med* **1999**. 65:388-390.
77. Kusari S, Zuhlke S, Borsch T and Spiteller M. Positive correlations between hypericin and putative precursors detected in the quantitative secondary metabolite spectrum of *Hypericum*. *Phytochemistry* **2009**. 70:1222-1232.
78. Taguchi G. **2016**. In *Molecular Breeding and Nutritional Aspects of Buckwheat*. Meiliang Zhou, Ivan Kreft, Sun-Hee Woo, Nikhil Chrungoo and Wieslander G (Eds.) Academic Press, 377-386.
79. Gutmann H, Poller B, Buter KB, Pfrunder A, Schaffner W and Drewe J. *Hypericum perforatum*: which constituents may induce intestinal MDR1 and CYP3A4 mRNA expression? *Planta Med* **2006**. 72:685-690.
80. Madabushi R, Frank B, Drewelow B, Derendorf H and Butterweck V. Hyperforin in St. John's wort drug interactions. *Eur J Clin Pharmacol* **2006**. 62:225-233.
81. Adiwidjaja J, Boddy AV and McLachlan AJ. Physiologically Based Pharmacokinetic Modelling of Hyperforin to Predict Drug Interactions with St John's Wort. *Clin Pharmacokinet* **2019**. 58:911-926.
82. Guo Y, Zhang N, Sun W, Duan X, Zhang Q, Zhou Q, Chen C, Zhu H, Luo Z, Liu J, Li XN, Xue Y and Zhang Y. Bioactive polycyclic polyprenylated acylphloroglucinols from *Hypericum perforatum*. *Org Biomol Chem* **2018**. 16:8130-8143.
83. Zhang JJ, Yang XW, Liu X, Ma JZ, Liao Y and Xu G. 1,9-seco-Bicyclic Polyprenylated Acylphloroglucinols from *Hypericum uralum*. *J Nat Prod* **2015**. 78:3075-3079.
84. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reilly MD, Thaden JJ and Viant MR. Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG). Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**. 3:211-221.
85. Hernandez MF, Falé PLV, Araújo MEM and Serralheiro MLM. Acetylcholinesterase inhibition and antioxidant activity of the water extracts of several *Hypericum* species. *Food Chem* **2010**. 120:1076-1082.
86. Kraus B, Wolff H, Heilmann J and Elstner EF. Influence of *Hypericum perforatum* extract and its single compounds on amyloid-beta mediated toxicity in microglial cells. *Life Sci* **2007**. 81:884-894.
87. Manczak M, Anekonda TS, Henson E, Park BS, Quinn J and Reddy PH. Mitochondria are a direct site of A-beta accumulation in Alzheimer's disease neurons: implications for free radical generation and oxidative damage in disease progression. *Hum Mol Genet* **2006**. 15:1437-1449.
88. Zou Y, Lu Y and Wei D. Antioxidant Activity of a Flavonoid-Rich Extract of *Hypericum perforatum* L. in Vitro. *J Agr Food Chem* **2004**. 52:5032-5039.
89. Jadhav HR and Bhutani KK. Antioxidant properties of Indian medicinal plants. *Phytother Res* **2002**. 16:771-773.
90. Sagratini G, Ricciutelli M, Vittori S, Ozturk N, Ozturk Y and Maggi F. Phytochemical and antioxidant analysis of eight *Hypericum* taxa from Central Italy. *Fitoterapia* **2008**. 79:210-213.
91. Silva BA, Ferreres F, Malva JO and Dias ACP. Phytochemical and antioxidant characterization of *Hypericum perforatum* alcoholic extracts. *Food Chem* **2005**. 90:157-167.
92. Zheleva-Dimitrova D, Nedialkov P and Kitanov G. Radical scavenging and antioxidant activities of methanolic extracts from *Hypericum* species growing in Bulgaria. *Pharmacognosy magazine* **2010**. 6:74-78.
93. Galla G, Barcaccia G, Schallau A, Puente Molins M, Baumlein H and Sharbel TF. The cytological basis of apospory in *Hypericum perforatum* L. *Sex Plant Reprod* **2011**. 24:47-61.
94. Rizzo P. **2016**. Novel insights on female gametophyte development in the apomictic model species *Boechera* spp. and *Hypericum* spp. Halle University.
95. Schallau A, Arzenton F, Johnston AJ, Hahnel U, Koszegi D, Blattner FR, Altschmied L, Haberer G, Barcaccia G and Baumlein H. Identification and genetic analysis of the APOSPORY locus in *Hypericum perforatum* L. *Plant J* **2010**. 62:773-784.
96. Kusari S, Sezgin S, Nigutova K, Cellarova E and Spiteller M. Spatial chemo-profiling of hypericin and related phytochemicals in *Hypericum* species using MALDI-HRMS. *Anal Bioanal Chem* **2015**. 407:4779-4791.

-
97. Ciccarelli D, Andreucci AC and Pagni AM. Translucent Glands and Secretory Canals in *Hypericum perforatum* L. (Hypericaceae): Morphological, Anatomical and Histochemical Studies During the Course of Ontogenesis. *Annals of botany* **2001**. 88:637-644.
 98. Bramanti E, Lenci F and Sgarbossa A. Effects of hypericin on the structure and aggregation properties of beta-amyloid peptides. *Eur Biophys J* **2010**. 39:1493-1501.
 99. Sgarbossa A, Buselli D and Lenci F. In vitro perturbation of aggregation processes in beta-amyloid peptides: a spectroscopic study. *FEBS Lett* **2008**. 582:3288-3292.
 100. Karppinen K, Derzso E, Jaakola L and Hohtola A. Molecular Cloning and Expression Analysis of hyp-1 Type PR-10 Family Genes in *Hypericum perforatum*. *Front Plant Sci* **2016**. 7:526.
 101. Kimáková K, Kimáková A, Idkowiak J, Stobiecki M, Rodziewicz P, Marczak L and Cellarova E. Phenotyping the genus *Hypericum* by secondary metabolite profiling: emodin vs. skyrin, two possible key intermediates in hypericin biosynthesis. *Anal Bioanal Chem* **2018**. 410:7689-7699.
 102. Karppinen K and Hohtola A. Molecular cloning and tissue-specific expression of two cDNAs encoding polyketide synthases from *Hypericum perforatum*. *J Plant Physiol* **2008**. 165:1079-1086.
 103. Bais HP, Vepachedu R, Lawrence CB, Stermitz FR and Vivanco JM. Molecular and biochemical characterization of an enzyme responsible for the formation of hypericin in St. John's wort (*Hypericum perforatum* L.). *J Biol Chem* **2003**. 278:32413-32422.
 104. Onelli E, Rivetta A, Giorgi A, Bignami M, Cocucci M and Patrignani G. Ultrastructural studies on the developing secretory nodules of *Hypericum perforatum*. *Flora* **2002**. 197:92-102.
 105. Soták M, Czeranková O, Klein D, Nigutová K, Altschmied L, Li L, Jose A, Wurtele ES and Čellárová E. Differentially Expressed Genes in Hypericin-Containing *Hypericum perforatum* Leaf Tissues as Revealed by De Novo Assembly of RNA-Seq. *Plant Mol Biol Rep* **2016**. 34:1027-1041.
 106. Hevia F, Berti M and Wilckens R. Quality and yield in St. John's wort (*Hypericum perforatum* L.) harvested in different phenological stages. *Acta Agronomica Hungarica* **2002**. 50:349-358.
 107. Koch MA, Scheriau C, Betzin A, Hohmann N and Sharbel TF. Evolution of cryptic gene pools in *Hypericum perforatum*: the influence of reproductive system and gene flow. *Annals of botany* **2013**. 111:1083-1094.
 108. Rutten T, Kruger C, Melzer M, Stephan UW and Hell R. Discovery of an extended bundle sheath in *Ricinus communis* L. and its role as a temporal storage compartment for the iron chelator nicotianamine. *Planta* **2003**. 217:400-406.
 109. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **2011**. 17:10-12.
 110. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **2011**. 29:644-652.
 111. Bray NL, Pimentel H, Melsted P and Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **2016**. 34:525-527.
 112. Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**. 15:550.
 113. Murthy HN, Kim YS, Park SY and Paek KY. Hypericins: biotechnological production from cell and organ cultures. *Appl Microbiol Biotechnol* **2014**. 98:9187-9198.
 114. Southwell IA and Campbell MH. Hypericin content variation in *Hypericum perforatum* in australia. *Phytochemistry* **1991**. 30:475-478.
 115. Fornasiero RB, Bianchi A and Pinetti A. Anatomical and Ultrastructural Observations in *Hypericum perforatum* L. Leaves. *J Herbs Spices Med Plants* **1998**. 5:21-33.
 116. Holscher D, Shroff R, Knop K, Gottschaldt M, Crecelius A, Schneider B, Heckel DG, Schubert US and Svatos A. Matrix-free UV-laser desorption/ionization (LDI) mass spectrometric imaging at the single-cell level: distribution of secondary metabolites of *Arabidopsis thaliana* and *Hypericum* species. *Plant J* **2009**. 60:907-918.
 117. Kucharikova A, Kimakova K, Janfelt C and Cellarova E. Interspecific variation in localization of hypericins and phloroglucinols in the genus *Hypericum* as revealed by desorption electrospray ionization mass spectrometry imaging. *Physiol Plant* **2016**. 157:2-12.
 118. Wirz A, Simmen U, Heilmann J, Calis I, Meier B and Sticher O. Bisanthraquinone glycosides of *Hypericum perforatum* with binding inhibition to CRH-1 receptors. *Phytochemistry* **2000**. 55:941-947.
 119. Falk H and Meyer J. On the Homo- and Heteroassociation of Hypericin. *Monatshefte für Chemie* **1994**. 125:753-762.

120. Crispin MC. **2014**. Metabolism and localization of polyketides in the genus *Hypericum*. Vol Doctor of Philosophy. Iowa State University: Ames, Iowa.
121. Farrow SC and Facchini PJ. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front Plant Sci* **2014**. 5:524.
122. Schuler MA and Werck-Reichhart D. Functional genomics of P450s. *Annu Rev Plant Biol* **2003**. 54:629-667.
123. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T and Fernie AR. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol Biochem* **2013**. 72:21-34.
124. Dixon DP, Skipsey M and Edwards R. Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* **2010**. 71:338-350.
125. Labrou NE, Papageorgiou AC, Pavli O and Flemetakis E. Plant GSTome: structure and functional role in xenome network and plant stress response. *Curr Opin Biotechnol* **2015**. 32:186-194.
126. Wagner U, Edwards R, Dixon DP and Mauch F. Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. *Plant Mol Biol* **2002**. 49:515-532.
127. Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K and Theodoulou FL. Plant ABC proteins - a unified nomenclature and updated inventory. *Trends Plant Sci* **2008**. 13:151-159.
128. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci* **2015**. 20:576-585.
129. Xu Z, Escamilla-Trevino LL, Zeng L, Lalgondar M, Bevan DR, Winkel BSJ, Mohamed A, Cheng C-L, Shih M-C, Poulton JE and Esen A. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol Biol* **2004**. 55:343-367.
130. Reddy VS, Shlykov MA, Castillo R, Sun EI and Saier MH, Jr. The major facilitator superfamily (MFS) revisited. *FEBS J* **2012**. 279:2022-2035.
131. Lee EJ and Facchini P. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. *Plant Cell* **2010**. 22:3489-3503.
132. Daniel B, Konrad B, Toplak M, Lahham M, Messenlehner J, Winkler A and Macheroux P. The family of berberine bridge enzyme-like enzymes: A treasure-trove of oxidative reactions. *Arch Biochem Biophys* **2017**. 632:88-103.
133. Käll L, Krogh A and Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Res* **2007**. 35:W429-432.
134. Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y and Taura F. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *J Biol Chem* **2004**. 279:39767-39774.
135. Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y and Morimoto S. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS Lett* **2007**. 581:2929-2934.
136. Wehrens R, Hageman JA, van Eeuwijk F, Kooke R, Flood PJ, Wijnker E, Keurentjes JJ, Lommen A, van Eekelen HD, Hall RD, Mumm R and de Vos RC. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* **2016**. 12:88.
137. Galili T, O'Callaghan A, Sidi J and Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **2018**. 34:1600-1602.
138. Wei T and Simko V. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). **2017**.
139. Cirak C, Radusiene J, Janulis V and Ivanauskas L. Pseudohypericin and hyperforin in *Hypericum perforatum* from Northern Turkey: variation among populations, plant parts and phenological stages. *J Integr Plant Biol* **2008**. 50:575-580.
140. Ang CY, Hu L, Heinze TM, Cui Y, Freeman JP, Kozak K, Luo W, Liu FF, Mattia A and DiNovi M. Instability of St. John's wort (*Hypericum perforatum* L.) and degradation of hyperforin in aqueous solutions and functional beverage. *J Agric Food Chem* **2004**. 52:6156-6164.
141. Fuzzati N, Gabetta B, Strepponi I and Villa F. High-performance liquid chromatography-electrospray ionization mass spectrometry and multiple mass spectrometry studies of hyperforin degradation products. *J Chromatogr A* **2001**. 926:187-198.
142. Vajs V, Vugdeliija S, Trifunović S, Karadžić I, Juranić N, Macura S and Milosavljević S. Further degradation product of hyperforin from *Hypericum perforatum* (St. John's Wort). *Fitoterapia* **2003**. 74:439-444.
143. Anne P, Francois DG, Vicky Chan Fook T and Philippe G. Effect on Light on Hypericins Contents in Fresh Flowering Top Parts and in an Extract of St. John's Wort (*Hypericum perforatum*). *Planta Med* **2001**. 67:254-259.

-
144. Smelcerovic A and Spiteller M. Phytochemical analysis of nine *Hypericum* L. species from Serbia and the F.Y.R. Macedonia. *Die Pharmazie* **2006**. 61:251-252.
 145. Sanchez-Illana A, Pineiro-Ramos JD, Sanjuan-Herraez JD, Vento M, Quintas G and Kuligowski J. Evaluation of batch effect elimination using quality control replicates in LC-MS metabolite profiling. *Anal Chim Acta* **2018**. 1019:38-48.
 146. Brunius C, Shi L and Landberg R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. *Metabolomics* **2016**. 12:173.
 147. Bagdonaitė E, Janulis V, Ivanauskas L and Labokas J. Between species diversity of *Hypericum perforatum* and *H. maculatum* by the content of bioactive compounds. *Natural product communications* **2012**. 7:199-200.
 148. Crockett SL, Schaneberg B and Khan IA. Phytochemical profiling of new and old world *Hypericum* (St. John's Wort) species. *Phytochem Anal* **2005**. 16:479-485.
 149. Zdunic G, Godjevac D, Savikin K and Petrovic S. Comparative Analysis of Phenolic Compounds in Seven *Hypericum* Species and Their Antioxidant Properties. *Natural product communications* **2017**. 12:1805-1811.
 150. Oniga I, Toiu A, Benedec D, Tomuta I and Vlase L. Phytochemical analysis of *Hypericum maculatum* in order to obtain standardized extracts. *Farmacia* **2016**. 64:171-174.
 151. Kartnig T, Göbel I and Heydel B. Production of Hypericin, Pseudohypericin and Flavonoids in Cell Cultures of Various *Hypericum* Species and their Chemotypes. *Planta Med* **1996**. 62:51-53.
 152. Radusienė J, Bagdonaitė E and Kazlauskas S. Morphological and Chemical Evaluation on *Hypericum perforatum* and *H. maculatum* in Lithuania. *Acta Hort* **2004**. 55-62.
 153. Nigutová K, Kusari S, Sezgin S, Petijova L, Henzelyova J, Balintova M, Spiteller M and Cellarova E. Chemometric evaluation of hypericin and related phytochemicals in 17 in vitro cultured *Hypericum* species, hairy root cultures and hairy root-derived transgenic plants. *J Pharm Pharmacol* **2017**.
 154. Prodan I, Sevastre B, Toiu A-M, Benedec D, Oniga I, Deliu C and Marcus I. Antitumour Activity of *Hypericum perforatum* and *Hypericum maculatum* in Ehrlich Ascitic Carcinoma. *Vet Med* **2009**. 66:176-181.
 155. Mártonfi P, Repcák M and Mihoková L. *Hypericum maculatum* Crantz subsp. *maculatum* X *H. perforatum* L. (Hypericaceae): corroboration of natural hybridization by secondary metabolite analysis. *Folia Geobot Phytotax* **1996**. 31:245-250.
 156. Nedialkov PT, Momekov G, Kokanova-Nedialkova ZK and Heilmann J. Polyprenylated Phloroglucinols from *Hypericum maculatum*. *Natural product communications* **2015**. 10:1231-1235.
 157. Zheleva-Dimitrova D, Nedialkov P, Girreser U and Kitanov G. Benzophenones and flavonoids from *Hypericum maculatum* and their antioxidant activities. *Nat Prod Res* **2012**. 26:1576-1583.
 158. Sampson AW and Parker k. St. Johnswort on Range Lands of California. *University of California Experimental Station Bulletin* **1930**. 503:1-48.
 159. Bütter R, Orlaccio C, Soldati A and Berger K. Significance of Genetic and Environmental Aspects in the Field Cultivation of *Hypericum perforatum*. *Planta Med* **1998**. 64:431-437.
 160. Tekel'ová D, Repcák M, Zemková E and Tóth J. Quantitative Changes of Dianthrone, Hyperforin and Flavonoids Content in the Flower Ontogenesis of *Hypericum perforatum*. *Planta Med* **2000**. 66:778-780.
 161. Chauhan RS, Rana JC, Gupta N, Sharma TR, Sharma SK and Jana S. Genetic and Genome Resources in Buckwheat – Present Status and Future Perspectives. *Eur J Plant Sci Biotechnol* **2010**. 4:33-44.
 162. Filippini R, Piovan A, Borsarini A and Caniato R. Study of dynamic accumulation of secondary metabolites in three subspecies of *Hypericum perforatum*. *Fitoterapia* **2010**. 81:115-119.
 163. Franke R, Schenk R and Bauermann U. Variability in *Hypericum perforatum* L. Breeding Lines. *Acta Hort* **1999**. 502:167-173.
 164. Avato P. A Survey on the *Hypericum* Genus: Secondary Metabolites and Bioactivity. *Stud Nat Prod Chem* **2005**. 30:603-634.
 165. Barnes J, Arnason JT and Roufogalis BD. St John's wort (*Hypericum perforatum* L.): botanical, chemical, pharmacological and clinical advances. *J Pharm Pharmacol* **2019**. 71:1-3.
 166. Müller WE, Singer A and Wonnemann M. Hyperforin - Antidepressant Activity by a Novel Mechanism of Action. *Pharmacopsychiatry* **2001**. 2001:98-102.
 167. Roz N, Mazur Y, Hirshfeld A and Rehavi M. Inhibition of vesicular uptake of monoamines by hyperforin. *Life Sciences* **2002**. 71:2227-2237.
 168. Butterweck V, Nahrstedt A, Evans J, Hufeisen S, Rauser L, Savage J, Popadak B, Ernsberger P and Roth BL. *In vitro* receptor screening of pure constituents of St. John's wort reveals novel interactions with a number of GPCRs. *Psychopharmacology* **2002**. 162:193-202.

169. Suzuki O, Katsumata M, Oya M, Bladt S and Wagner H. Inhibition of Monoamine Oxidase by Hypericin. *Planta Med* **1984**. 50:272-274.
170. Müller WE, Rolli M, Schäfer C and Hafner U. Effects of *Hypericum* Extract (LI 160) in Biochemical Models of Antidepressant Activity. *Pharmacopsychiatry* **1997**. 30:102-107.
171. Jürgenliemk G and Nahrstedt A. Dissolution, solubility and cooperativity of phenolic compounds from *Hypericum perforatum* L. in aqueous systems. *Die Pharmazie* **2003**. 58:200-203.
172. Verjee S, Kelber O, Kolb C, Abdel-Aziz H and Butterweck V. Permeation characteristics of hypericin across Caco-2 monolayers in the presence of single flavonoids, defined flavonoid mixtures or *Hypericum* extract matrix. *J Pharm Pharmacol* **2019**. 71:58-69.
173. Butterweck V, Liefländer-Wulf U, Winterhoff H and Nahrstedt A. Plasma Levels of Hypericin in Presence of Procyanidin B2 and Hyperoside: A Pharmacokinetic Study in Rats. *Planta Med* **2003**. 69:189-192.
174. Nöldner M and Schötz K. Rutin is Essential for the Antidepressant Activity of *Hypericum perforatum* Extracts in the Forced Swimming Test. *Planta Med* **2002**. 68:577-580.
175. Calsolaro V and Edison P. Neuroinflammation in Alzheimer's disease: Current evidence and future directions. *Alzheimers Dement* **2016**. 12:719-732.
176. Brenn A, Grube M, Jedlitschky G, Fischer A, Strohmeier B, Eiden M, Keller M, Groschup MH and Vogelgesang S. St. John's Wort reduces beta-amyloid accumulation in a double transgenic Alzheimer's disease mouse model-role of P-glycoprotein. *Brain Pathol* **2014**. 24:18-24.
177. Cao Z, Wang F, Xiu C, Zhang J and Li Y. *Hypericum perforatum* extract attenuates behavioral, biochemical, and neurochemical abnormalities in Aluminum chloride-induced Alzheimer's disease rats. *Biomed Pharmacother* **2017**. 91:931-937.
178. Li H, Wan L, Hashi Y and Chen S. Fragmentation study of a 8-C-glycosyl isoflavone, puerarin, using electrospray ion trap time-of-flight mass spectrometry at high resolution. *Rapid Commun Mass Spectrom* **2007**. 21:2497-2504.
179. Francioso A, Franke K, Villani C, Mosca L, D'Erme M, Frischbutter S, Brandt W, Sanchez-Lamar A and Wessjohann L. Insights into the Phytochemistry of the Cuban Endemic Medicinal Plant *Phyllanthus orbicularis*: Fideloside, a Novel Bioactive 8-C-glycosyl 2,3-Dihydroflavonol. *Molecules (Basel, Switzerland)* **2019**. 24.
180. Ye M, Han J, Chen H, Zheng J and Guo D. Analysis of phenolic compounds in rhubarbs using liquid chromatography coupled with electrospray ionization mass spectrometry. *J Am Soc Mass Spectrom* **2007**. 18:82-91.
181. Hecka A, Maunit B, Aubriet F and Muller JF. Study of active naphthodianthrone St John's Wort compounds by electrospray ionization Fourier transform ion cyclotron resonance and multi-stage mass spectrometry in sustained off-resonance irradiation collision-induced dissociation and infrared multiphoton dissociation modes. *Rapid Commun Mass Spectrom* **2009**. 23:885-898.
182. Piperopoulos G, Lotz R, Wixforth A, Schmierer T and Zeller K-P. Determination of naphthodianthrone in plant extracts from *Hypericum perforatum* L. by liquid chromatography–electrospray mass spectrometry. *J Chromatogr B* **1997**. 695:309-316.
183. Revuru B, Balintova M, Henzelyova J, Cellarova E and Kusari S. MALDI-HRMS Imaging Maps the Localization of Skyrin, the Precursor of Hypericin, and Pathway Intermediates in Leaves of *Hypericum* Species. *Molecules (Basel, Switzerland)* **2020**. 25.
184. Pradeep M, Kachlicki P and Franklin G. Simultaneous determination of naphthodianthrone, emodin, skyrin and new bisanthrones in *Hypericum perforatum* L. *in vitro* shoot cultures. *Ind Crop Prod* **2020**. 144:112003.
185. Kubin A, Wierrani F, Burner U, Alth G and Grünberger W. Hypericin - The Facts About a Controversial Agent. *Curr Pharm Des* **2005**. 11:233-253.
186. Wolfender J-L, Verotta L, Belvisi L, Fuzzati N and Hostettmann K. Structural Investigations of Isomeric Oxidised Forms of Hyperforin by HPLC-NMR and HPLC-MSn. *Phytochem Anal* **2003**. 14:290-297.
187. Schmidt J. Negative ion electrospray high-resolution tandem mass spectrometry of polyphenols. *J Mass Spectrom* **2016**. 51:33-43.
188. Zhang L, Tai Y, Wang Y, Meng Q, Yang Y, Zhang S, Yang H, Zhang Z, Li D and Wan X. The proposed biosynthesis of procyanidins by the comparative chemical analysis of five *Camellia* species using LC-MS. *Sci Rep* **2017**. 7:46131.
189. Fornal CA, Metzler CW, Mirescu C, Stein SK and Jacobs BL. Effects of Standardized Extracts of St. John's Wort on the Single-Unit Activity of Serotonergic Dorsal Raphe Neurons in Awake Cats: Comparisons with Fluoxetine and Sertraline. *Neuropharmacology* **2001**. 25:858-870.

-
190. Murch SJ and Saxena PK. St. John's wort (*Hypericum perforatum* L.): Challenges and strategies for production of chemically consistent plants. *Can J Plant Sci* **2006**. 1:765-771.
191. Huang L-K and Wang M-JJ. Image Tresholding by minimizing the measures of fuzziness. *Pattern recognition* **1995**. 28:41-51.
192. Li MN, Dong X, Gao W, Liu XG, Wang R, Li P and Yang H. Global identification and quantitative analysis of chemical constituents in traditional Chinese medicinal formula Qi-Fu-Yin by ultra-high performance liquid chromatography coupled with mass spectrometry. *J Pharm Biomed Anal* **2015**. 114:376-389.
193. Cirak C, Radusiené J, Karabük B and Janulis V. Variation of bioactive substances and morphological traits in *Hypericum perforatum* populations from Northern Turkey. *Biochem Syst Ecol* **2007**. 35:403-409.
194. Kirakosyan A, Gibson DM and Sirven T. A Comparative Study of *Hypericum perforatum* Plants as Sources of Hypericins and Hyperforins. *Journal of Herbs, Spices & Medicinal Plants* **2003**. 10:73-88.
195. Hong D, Yin F, Hu LH and Lu P. Sulfonated xanthenes from *Hypericum sampsonii*. *Phytochemistry* **2004**. 65:2595-2598.
196. Duong TH, Beniddir MA, Nguyen VK, Aree T, Gallard JF, Mac DH, Nguyen HH, Bui XH, Boustie J, Nguyen KP, Chavasiri W and Le Pogam P. Sulfonic Acid-Containing Flavonoids from the Roots of *Phyllanthus acidus*. *J Nat Prod* **2018**. 81:2026-2031.
197. Yu-Lin Huang C-CC, Feng-Lin Hsu, Chieh-Fu Chen. Tannins, Flavonol Sulfonates, and a Norlignan from *Phyllanthus virgatus*. *Journal of Natural products* **1998**. 61:1194-1197.
198. Sayed AE-T, Salib JY, Shafik NH, Elkaram AS, Micky JA and Farag AA. New Flavonoids from the aerial parts of *Polygonum equisetiforme* SM (Polygonaceae). *Int J Pharm Pharm Sci* **2017**. 9:166-170.
199. Seabra R and Alves AC. Quercetin-3-glucuronide-3'-sulphate from hypericum elodes. *Phytochemistry* **1988**. 27:3019-3020.
200. Kaserer T, Schuster D and Rollinger JM. **2018**. Chemoinformatics in Natural Product Research. In *Applied Chemoinformatics*. 207-236.
201. Fobofou Tanemossu SA. **2016**. Metabolomic Analysis, Isolation, Characterization and Synthesis of Bioactive Compounds from *Hypericum* species (Hypericaceae). In *Naturwissenschaftlichen Fakultät II - Chemie, Physik und Mathematik*. Martin-Luther-Universität Halle-Wittenberg: Halle.
202. Cirak C, Radusiene J, Jakstas V, Ivanauskas L, Yayla F, Seyis F and Camas N. Secondary metabolites of *Hypericum* species from the Drosanthe and Olympia sections. *S Afr J Bot* **2016**. 104:82-90.
203. Lazzara S, Carrubba A and Napoli E. Variability of Hypericins and Hyperforin in *Hypericum* Species from the Sicilian Flora. *Chemistry & biodiversity* **2020**. 17:e1900596.
204. Smelcerovic A, Verma V, Spitteller M, Ahmad SM, Puri SC and Qazi GN. Phytochemical analysis and genetic characterization of six *Hypericum* species from Serbia. *Phytochemistry* **2006**. 67:171-177.
205. Umek A, Kreft S, Kartnig T and Heydel B. Quantitative Phytochemical Analyses of Six *Hypericum* Species Growing in Slovenia. *Planta Med* **1999**. 65:388-390.
206. Kitanov G. Hypericin and pseudohypericin in some *Hypericum* species. *Biochem Syst Ecol* **2001**. 29:171-178.
207. Saddiqe Z, Naem I and Maimoona A. A review of the antibacterial activity of *Hypericum perforatum* L. *J Ethnopharmacol* **2010**. 131:511-521.
208. Saroglou V, Marin PD, Rancic A, Veljic M and Skaltsa H. Composition and antimicrobial activity of the essential oil of six *Hypericum* species from Serbia. *Biochem Syst Ecol* **2007**. 35:146-152.
209. Kitanov GM and Nedialkov PT. Mangiferin and isomangiferin in some *Hypericum* species. *Biochem Syst Ecol* **1998**. 26:647-653.
210. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York:Springer-Verlag, **2016**.
211. Michels B, Franke K, Weiglein A, Sultani H, Gerber B and Wessjohann LA. Rewarding compounds identified from the medicinal plant *Rhodiola rosea*. *J Exp Biol* **2020**. 223.
212. Bejaoui A, Ben Salem I, Rokbeni N, M'Rabet Y, Boussaid M and Boulila A. Bioactive compounds from *Hypericum humifusum* and *Hypericum perforatum*: inhibition potential of polyphenols with acetylcholinesterase and key enzymes linked to type-2 diabetes. *Pharm Biol* **2017**. 55:906-911.
213. Hu L-H and Sim K-Y. Sampsoniones A–M, a Unique Family of Caged Polyprenylated Benzoylphloroglucinol Derivatives, from *Hypericum sampsonii*. *Tetrahedron* **2000**. 56:1379-1386.
214. Bonkanka CX, Smelcerovic A, Zuehlke S, Rabanal RM, Spitteller M and Sanchez-Mateo Cdel C. HPLC-MS analysis and anti-oedematogenic activity of *Hypericum grandifolium* Choisy (Hypericaceae). *Planta Med* **2008**. 74:719-725.
215. Bottega S, Garbari F and Pagni AM. *Hypericum elodes* L. (Clusiaceae): The secretory structures of the flower. *Isr J Plant Sci* **2004**. 52:51-57.

216. Piovan A, Filippini R, Caniato R, Borsarini A, Bini Maleci L and Cappelletti EM. Detection of hypericins in the "red glands" of *Hypericum elodes* by ESI-MS/MS. *Phytochemistry* **2004**. 65:411-414.
217. Aramaki Y, Chiba K and Tada M. Absolute stereochemistry of chinesin I and II. *J Chem Soc Perkin Trans* **1995**. 1:683-688.
218. Cirak C and Radusiene J. Factors affecting the variation of bioactive compounds in *Hypericum* species. *Biologia Futura* **2019**. 70:198-209.
219. Aziz N, Sauve RJ, Long D and Cherry M. Genetic and Phytochemical Diversity Assessment Among Eleven *Hypericum* Accessions via AFLP and HPLC Analyses. *J Herbs Spices Med Plants* **2008**. 12:97-105.
220. Nürk NM, Madriñán S, Carine MA, Chase MW and Blattner FR. Molecular phylogenetics and morphological evolution of St. John's wort (*Hypericum*; Hypericaceae). *Mol Phylogenetics Evol* **2013**. 66:1-16.
221. Hielscher-Michael S, Griehl C, Buchholz M, Demuth HU, Arnold N and Wessjohann LA. Natural Products from Microalgae with Potential against Alzheimer's Disease: Sulfolipids Are Potent Glutaminyl Cyclase Inhibitors. *Mar Drugs* **2016**. 14.
222. Michels K. **2011**. Entwicklung einer LC-MS basierten Methode zur Identifizierung von aktivitätsrelevanten Metaboliten in komplexen Mischungen. In *Naturwissenschaftlichen Fakultät II - Chemie, Physik und Mathematik*. Martin-Luther-Universität Halle-Wittenberg: Halle.
223. Henry GE, Campbell MS, Zelinsky AA, Liu Y, Bowen-Forbes CS, Li L, Nair MG, Rowley DC and Seeram NP. Bioactive acylphloroglucinols from *Hypericum densiflorum*. *Phytother Res* **2009**. 23:1759-1762.
224. Stark P, Zab C, Porzel A, Franke K, Rizzo P and Wessjohann LA. PSYCHE—A Valuable Experiment in Plant NMR-Metabolomics. *Molecules (Basel, Switzerland)* **2020**. 25:1-13.
225. Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M and Wishart DS. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**. 9:1-39.
226. Larive CK, Barding GA, Jr. and Dinges MM. NMR spectroscopy for metabolomics and metabolic profiling. *Anal Chem* **2015**. 87:133-146.
227. Wojtowicz W, Zabek A, Deja S, Dawiskiba T, Pawelka D, Glod M, Balcerzak W and Mlynarz P. Serum and urine ¹H NMR-based metabolomics in the diagnosis of selected thyroid diseases. *Sci Rep* **2017**. 7:1-13.
228. Consonni R and Cagliani LR. The potentiality of NMR-based metabolomics in food science and food authentication assessment. *Magn Reson Chem* **2019**. 57:558-578.
229. Farag MA, Porzel A, Schmidt J and Wessjohann LA. Metabolite profiling and fingerprinting of commercial cultivars of *Humulus lupulus* L. (hop): a comparison of MS and NMR methods in metabolomics. *Metabolomics* **2011**. 8:492-507.
230. Sevastos A, Kalampokis IF, Panagiotopoulou A, Pelecanou M and Aliferis KA. *Fusarium graminearum* ¹H NMR metabolomics. *Data Brief* **2018**. 19:1162-1165.
231. Farag MA, Porzel A, Al-Hammady MA, Hegazy ME, Meyer A, Mohamed TA, Westphal H and Wessjohann LA. Soft Corals Biodiversity in the Egyptian Red Sea: A Comparative MS and NMR Metabolomics Approach of Wild and Aquarium Grown Species. *J Proteome Res* **2016**. 15:1274-1287.
232. Santacruz L, Hurtado DX, Doohan R, Thomas OP, Puyana M and Tello E. Metabolomic study of soft corals from the Colombian Caribbean: PSYCHE and ¹H-NMR comparative analysis. *Sci Rep* **2020**. 10:1-11.
233. Oliveira AI, Pinho C, Sarmiento B and Dias AC. Neuroprotective Activity of *Hypericum perforatum* and Its Major Components. *Front Plant Sci* **2016**. 7:1004.
234. Lopez JM, Cabrera R and Maruenda H. Ultra-Clean Pure Shift ¹H-NMR applied to metabolomics profiling. *Sci Rep* **2019**. 9:1-8.
235. Moutzouri P, Chen Y, Foroozandeh M, Kiraly P, Phillips AR, Coombes SR, Nilsson M and Morris GA. Ultraclean pure shift NMR. *Chem Commun* **2017**. 53:10188-10191.
236. Zangger K and Sterk H. Homonuclear Broadband-Decoupled NMR Spectra. *J Magn Reson* **1997**. 124:486-489.
237. Farag MA, Mahrous EA, Lübken T, Porzel A and Wessjohann L. Classification of commercial cultivars of *Humulus lupulus* L. (hop) by chemometric pixel analysis of two dimensional nuclear magnetic resonance spectra. *Metabolomics* **2014**. 10:21-32.
238. Mahrous EA and Farag MA. Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review. *J Adv Res* **2015**. 6:3-15.
239. Aguilar JA, Faulkner S, Nilsson M and Morris GA. Pure shift ¹H NMR: a resolution of the resolution problem? *Angew Chem Int Ed Engl* **2010**. 49:3901-3903.
240. Castanar L. Pure shift ¹H NMR: what is next? *Magn Reson Chem* **2017**. 55:47-53.
241. Zangger K. Pure shift NMR. *Prog Nucl Magn Reson Spectrosc* **2015**. 86-87:1-20.

-
242. Aue WP, Karhan J and Ernst RR. Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy. *J Chem Phys* **1976**. 64:4226-4227.
243. Foroozandeh M, Adams RW, Kiraly P, Nilsson M and Morris GA. Measuring couplings in crowded NMR spectra: pure shift NMR with multiplet analysis. *Chem Comm* **2015**. 51:15410-15413.
244. Foroozandeh M, Morris GA and Nilsson M. PSYCHE Pure Shift NMR Spectroscopy. *Chemistry* **2018**. 24:13988-14000.
245. Bo Y, Feng J, Xu J, Huang Y, Cai H, Cui X, Dong J, Ding S and Chen Z. High-resolution pure shift NMR spectroscopy offers better metabolite discrimination in food quality analysis. *Food Res Int* **2019**. 125:1-10.
246. Ludwig C and Viant MR. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem Anal* **2010**. 21:22-32.
247. Rodriguez-Martinez A, Posma JM, Ayala R, Harvey N, Jimenez B, Neves AL, Lindon JC, Sonomura K, Sato TA, Matsuda F, Zalloua P, Gauguier D, Nicholson JK and Dumas ME. J-Resolved ¹H NMR 1D-Projections for Large-Scale Metabolic Phenotyping Studies: Application to Blood Plasma Analysis. *Anal Chem* **2017**. 89:11405-11412.
248. Anderson PE, Mahle DA, Doom TE, Reo NV, DelRaso NJ and Raymer ML. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. *Metabolomics* **2010**. 7:179-190.
249. Foroozandeh M, Adams RW, Meharry NJ, Jeannerat D, Nilsson M and Morris GA. Ultrahigh-resolution NMR spectroscopy. *Angew Chem Int Ed Engl* **2014**. 53:6990-6992.
250. Aguilar JA. **2014**. Introduction to pure shift NMR. mestrelab: Durham University.
251. Agnolet S, Jaroszewski JW, Verpoorte R and Staerk D. H NMR-based metabolomics combined with HPLC-PDA-MS-SPE-NMR for investigation of standardized *Ginkgo biloba* preparations. *Metabolomics* **2010**. 6:292-302.
252. Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, Fornier M, Hudis C, Norton L, Luchinat C and Di Leo A. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. *Mol Oncol* **2015**. 9:128-139.
253. Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AW and Verpoorte R. Metabolomic differentiation of *Cannabis sativa* cultivars using ¹H NMR spectroscopy and principal component analysis. *J Nat Prod* **2004**. 67:953-957.
254. Ayan AK and Çirak C. Hypericin and Pseudohypericin Contents in Some *Hypericum* Species Growing in Turkey. *Pharm Biol* **2008**. 46:288-291.
255. Zirak N, Shafiee M, Soltani G, Mirzaei M and Sahebkar A. *Hypericum perforatum* in the treatment of psychiatric and neurodegenerative disorders: Current evidence and potential mechanisms of action. *J Cell Physiol* **2019**. 234:8496-8508.
256. Mennini T and Gobbi M. The antidepressant mechanism of *Hypericum perforatum*. *Life Sci* **2004**. 75:1021-1027.
257. Schlippenbach TV, Oefner PJ and Gronwald W. Systematic Evaluation of Non-Uniform Sampling Parameters in the Targeted Analysis of Urine Metabolites by ¹H,¹H 2D NMR Spectroscopy. *Sci Rep* **2018**. 8:4249.
258. Kartnig T and Brantner A. Secondary Constituents in Cell Cultures of *Hypericum perforatum* and *Hypericum maculatum*. *Planta Med* **1990**. 56.
259. Zapesochnay OO, Kuptsova LP, Kyshtymova TV, Ban'kovskii AI and Mel'nikova TM. Flavonoids of *Hypericum maculatum* and *H. inodorum*. *Khimiya Prirodnikh Soedinenii*, **1966**. 3:279-280.
260. Coste A, Vlase L, Halmagyi A, Deliu C and Coldea G. Effects of plant growth regulators and elicitors on production of secondary metabolites in shoot cultures of *Hypericum hirsutum* and *Hypericum maculatum*. *Plant Cell Tiss Organ Cult* **2011**. 106:279-288.
261. Kladar N, Srdenovic B, Grujic N, Bokic B, Rat M, Anackov G and Bozin B. Ecologically and ontogenetically induced variations in phenolic compounds and biological activities of *Hypericum maculatum* subsp *maculatum*, Hypericaceae. *Rev Bras Bot* **2015**. 38:703-715.

Appendix

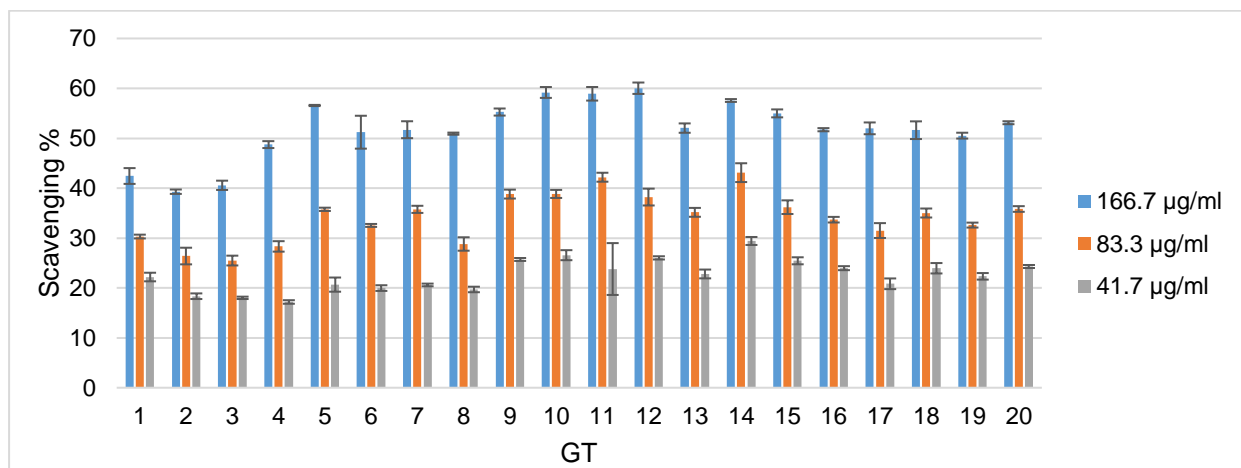
Appendix 2.1: Information of *H. perforatum* plant material

Genotype	Accession (IPK)	Extract amount [g]	Locality	Latitude	Longitude	Ploidy	Apomeiosis %	Parthenogenesis %	Fertilization %	Polyembryony %	Genetic background according to Mollins et al. 2013		
											red %	green %	blue %
1	H06-1383	3.895	Green Lake WI US	43.85	89.30	4C	63.3	60.0	40.0	23.3	0	100	0
2	H06-3246	3.863	Praha Czech	50.08	14.46	6C	13.6	40.9	59.1	0	0	0	100
3	H06-3324	3.694	Bonn Germany	50.73	7.10	4C	79.2	41.7	58.3	4.2	100	0	0
4	H06-3087	5.264	Adliswil Switzerland	47.30	8.56	4C	64.7	52.9	47.1	11.8	0	95	5
5	H013-0508a	4.571											
6	H06-2844	3.746	Badia Polesine Italy	45.08	11.48	4C	91.7	87.5	12.5	4.2	20	0	80
7	H06-1644	4.452	Rideau River ON Canada	45.00	75.62	6C	8.3	20.8	79.2	0	0	100	0
8	H06-1498	5.080	Kewaunee MI US	44.61	88.11	6C	13.0	34.8	65.2	0	0	0	100
9	H06-3251	4.305	Praha Czech	50.08	14.46	4C	95.2	90.5	9.5	0	100	0	0
10	H013-0508b	4.609											
11	H06-2745	4.588	Bolzano Italy	46.51	12.15	4C	75.0	66.7	33.4	4.2	0	0	100
12	H06-1640	4.058	Afton MN US	44.90	93.06	4C	66.7	66.7	33.3	16.7	60	40	0
13	H06-1449	4.422	Point Beach WI US	44.26	87.56	6C	0	0	100	0	0	0	100
14	H06-2886	3.950	Badia Polesine Italy	45.08	11.48	6C	9.1	27.3	72.7	0	0	100	0
15	H06-38	5.102	St/4(F ₁ 2xAn)1/10/1*										
16	H06-1359	3.802	Tuscola IL US	39.80	88.28	4C	90.5	80.9	14.3	4.8	0	100	0
17	H06-343	5.235	St/4(F ₁ 1xNo)1/8/4*										
18	H06-1654	4.809	Rideau River ON Canada	45.00	75.62	4C	71.4	66.7	33.3	0	0	0	100
19	H06-1378	4.422	ESGR MI US	42.50	83.70	4C	91.7	87.5	12.5	4.2	0	0	100
20	H06-54	5.056	St/4(F ₁ 2xNo)1a/8/1*										

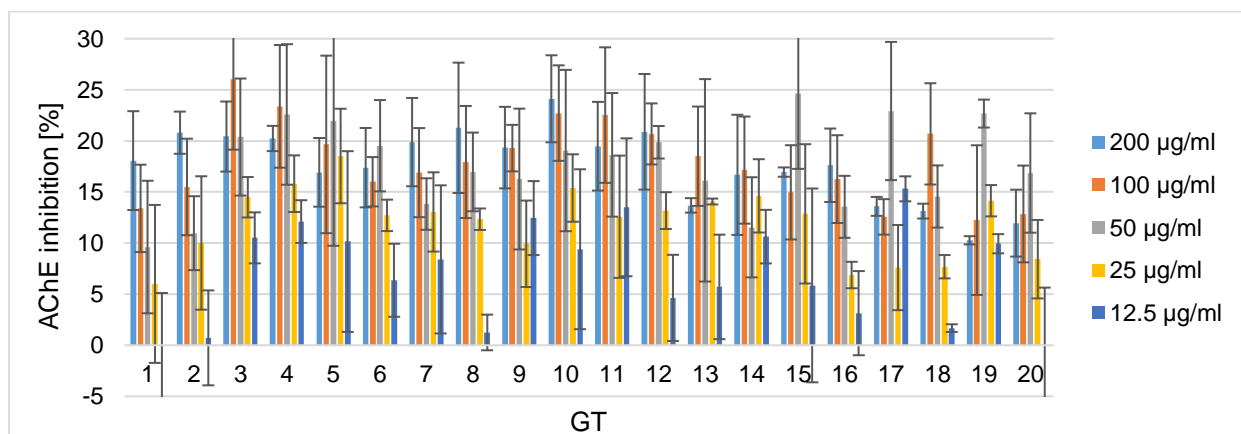
* crossing was performed in Gatersleben (IPK), using accessions from Norway (No) and Italy (Ac, Ancona)

Appendix 2.2 Biological activities of 20 *H. perforatum* genotypes (GTs)

a) DPPH-radical scavenger activity



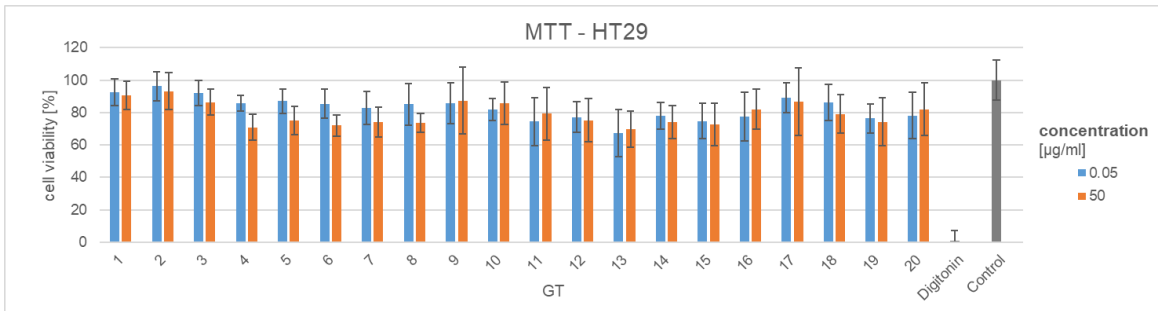
b) Acetylcholinesterase inhibition



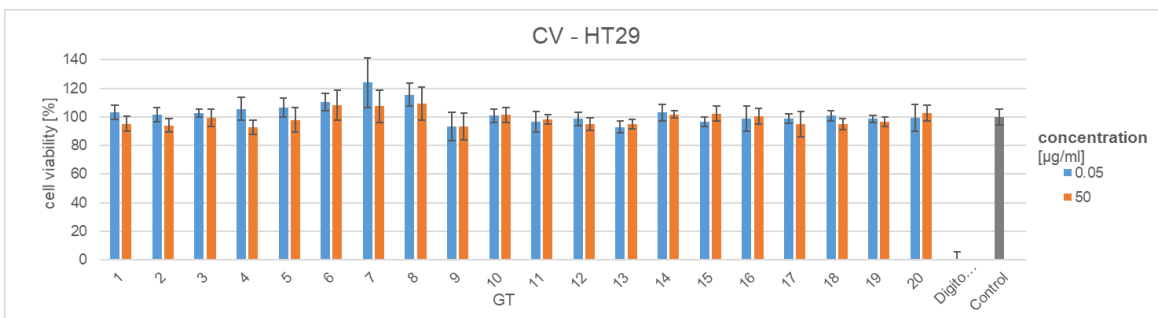
c) Cytotoxicity cell assay

Digitonin was used as positive control and DMSO as a negative control.

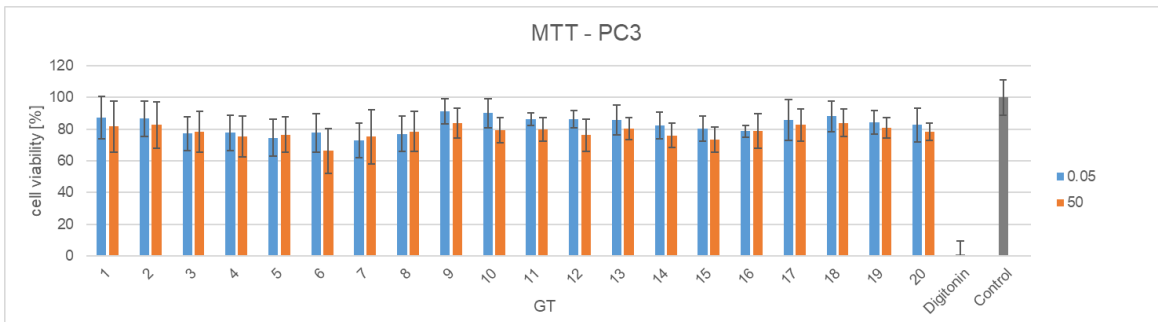
MTT assay of colon cancer cell line (HT29).



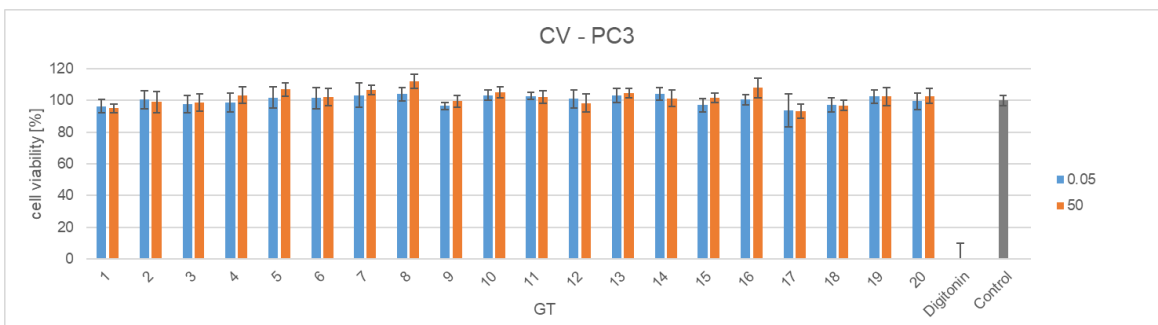
Cell viability assay with crystal violet (CV) of colon cancer cell line (HT29)



MTT assay of human prostate cancer cell line (PC3)



Cell viability assay with crystal violet (CV) of human prostate cancer cell line (PC3)



Appendix 3.1: Plant material information of 93 *H. perforatum* GTs

GT	Genotype identifier		Sampling characteristics				Genotype origin			Reproduction characteristics					Genetic background			DG
	LC-MS samples	Line ID	Organ	Quartil	Day	Flower	Locality	Lat	Lon	Ploidy	Apomeiosis	Parthenogenesis	Fertilization	Poly-embryony	re	gr	bl	
1	D6009_F, D6017_F, D6029_F, D5069_F, D4048_F, D3069_F, D2020_F, D1020_F, D1041_F, D3069_L, D3090_L, D4027_L, D5048_L, D5069_L, D5101_L, D5102_L, D4048_L	HYPR-01	F	4	1, 1, 2, 3, 4, 5, 6, 6, 6	f	NA	NA	NA	2C	NA	NA	100	NA	NA	NA	NA	115.7
2	D1021_F, D1042_F, D2021_F, D3070_F, D4049_F, D3070_L, D4049_L, D5070_L	HYPR-03	F	4	1, 1, 2, 3, 4, 4, 5	f	NA	NA	2C	NA	NA	100	NA	NA	NA	NA	95.2	
3	D3068_F, D4047_F, D5068_F, D3068_L, D4047_L, D5068_L	HEID-919132	F	4	3, 4, 5, 4, 5	m	Heidelberg University	NA	NA	2C	NA	NA	100	NA	NA	NA	0	
4	D6001_F, D5081_F, D4011_F, D4011_L, D3081_L, D5081_L	H06-1460	F	2	4, 5, 6, 3, 4, 5	f	Point Beach WI US	44.26	87.56	4C	0	0	100	0	0	0	100	29.8
5	D1033_F, D1012_F, D2012_F, D3086_L, D4095_L, D5046_L	H06-3194	F	1	1, 1, 2, 3, 4, 5	f	Veip Netherlands	52	5.59	4C	5.3	0	100	0	70	30	0	1
6	D5056_F, D5100_F, D4035_F, D3058_L, D4035_L, D5056_L	H06-3281	F	4	5, 5, 4, 3, 4, 5	f	Prana Czech	50.08	14.46	6C	5.6	11.1	88.9	0	0	0	100	0
7	D1023_F, D1002_F, D2002_F, D3076_L, D4080_L, D5031_L	H06-3191	F	1	1, 1, 2, 3, 4, 5	f	Veip Netherlands	52	5.59	4C	30.0	15.0	85.0	15	70	30	0	0
8	D3015_F, D4008_F, D5078_F, D3015_L, D4008_L, D5078_L	H06-3220	F	2	3, 4, 5, 3, 4, 5	f	Silwood England	51.5	0.46	4C	29.4	23.5	76.5	0	100	0	0	0
9	D3027_F, D4050_F, D5001_F, D3027_L, D4050_L, D5001_L	H06-1887	F	3	3, 4, 5, 3, 4, 5	f	Iron Mountain MI US	45.83	88.08	6C	19.0	23.8	76.2	0	0	0	0	3.8
10	D3004_F, D4078_F, D5029_F, D3004_L, D4078_L, D5029_L	H06-3190	F	1	3, 4, 5, 3, 4, 5	f	Veip Netherlands	52	5.59	4C	34.6	26.9	73.1	15.4	70	30	0	11.4
11	D1007_F, D1028_F, D2007_F, D3082_L, D4087_L, D5038_L	H06-2886	F	1	1, 1, 2, 3, 4, 5	f	Badia Polesine Italy	45.08	11.48	6C	9.1	27.3	72.7	0	0	100	0	95.9
12	D1031_F, D1010_F, D2010_F, D3085_L, D4092_L, D5043_L	H06-3184	F	1	1, 1, 2, 3, 4, 5	f	Veip Netherlands	52	5.59	4C	27.3	27.3	72.7	9.1	50	50	0	27.5
13	D3025_F, D4024_F, D5094_F, D3025_L, D4024_L, D5094_L	H06-3341	F	2	3, 4, 5, 3, 4, 5	f	Bonn Germany	50.73	7.1	6C	8.3	29.2	70.8	0	100	0	0	0.8
14	D4051_F, D5002_F, D6004_F, D3028_L, D4051_L, D5002_L	H06-1453	F	3	4, 5, 6, 3, 4, 5	l	Point Beach WI US	44.26	87.56	4C	72.0	72	68	4	0	0	100	0
15	D3012_F, D4004_F, D5074_F, D3012_L, D4004_L, D5074_L	H06-3200	F	2	3, 4, 5, 3, 4, 5	f	Silwood England	51.5	0.46	4C	38.4	34.6	65.4	19.2	100	0	0	0
16	D5063_F, D6021_F, D6013_F, D3064_L, D4042_L, D5063_L	H06-1498	F	4	5, 6, 6, 3, 4, 5	m	Kewannee MI US	44.61	88.11	6C	13.0	34.8	65.2	0	0	0	100	0
17	D4056_F, D5007_F, D3033_L, D4056_L, D5007_L	H06-1698	F	3	4, 5, 3, 4, 5	l	Meminee MI US	45.21	87.75	6C	13.0	34.8	65.2	0	0	0	0	0
18	D1025_F, D1004_F, D2004_F, D3078_L, D4082_L, D5033_L	H06-3085	F	1	1, 1, 2, 3, 4, 5	f	Adliswil Switzerland	47.3	8.56	6C	13.0	34.8	65.2	0	0	100	0	43.2
19	D3026_F, D4025_F, D5095_F, D3026_L, D4025_L, D5095_L	H06-3265	F	2	3, 4, 5, 3, 4, 5	f	Prana Czech	50.08	14.46	6C	17.6	29.4	64.7	0	0	0	100	0

GT	LC-MS samples	Line ID	O	Q	Day	F	Locality	Lat	Lon	PI	Apo	Part	Fert	Poly	re	gr	bl	DG
20	D3046_F, D4069_F, D5020_F D3046_L, D4069_L, D5020_L	H06-3199	F L	3 3	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	45.8	37.5	62.5	12.5	100	0	0	0
21	D1038_F, D1017_F, D2017_F D4001_L, D4015_L, D5085_L	H06-3337	F L	2 2	1, 1, 2 3, 4, 5	f	Bonn Germany	50.73	7.1	4C	68.9	37.9	62.1	34.5	100	0	0	3.4
22	D3018_F, D4016_F, D5086_F D3018_L, D4016_L, D5086_L	H06-3233	F L	2 2	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	38.46	38.46	61.53	23.1	100	0	0	3.8
23	D3045_F, D4068_F, D5019_F D3045_L, D4068_L, D5019_L	H06-3196	F L	3 3	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	43.47	39.13	60.86	0	100	0	0	0
24	D3052_F, D5050_F, D6018_F D3052_L, D4028_L, D5050_L	H06-2913	F L	4 4	3, 5, 6 3, 4, 5	m	Badia Polesine Italy	45.08	11.48	6C	20	40	60	0	0	100	0	79.6
25	D3005_F, D4079_F, D5030_F D3005_L, D4079_L, D5030_L	H06-3189	F L	1 1	3, 4, 5 3, 4, 5	f	Veip Netherlands	52	5.59	4C	50	40	60	26.7	80	20	0	10
26	D3019_F, D4017_F, D5087_F D3019_L, D4017_L, D5087_L	H06-3287	F L	2 2	3, 4, 5 3, 4, 5	f	Praha Czech	50.08	14.46	4C	51.8	40.7	59.3	18.5	100	0	0	0
27	D3021_F, D4019_F, D5089_F D3021_L, D4019_L, D5089_L	H06-3246	F L	2 2	3, 4, 5 3, 4, 5	f	Praha Czech	50.08	14.46	6C	13.6	40.9	59.1	0	0	0	100	0
28	D2018_F, D3073_F, D4020_F, D5090_F, D6003_F, D6016_F, D6025_F	H06-1819	F	2	2, 3, 4, 5, 6, 6	f	Carney MI US	45.6	87.03	4C	58.8	47.1	58.8	0	100	0	0	0
29	D3073_L, D4020_L, D5090_L D4070_F, D5098_F, D6027_F, D6007_F	H06-3331	F	2	3, 4, 5	l	Bonn Germany	50.73	7.1	6C	23.5	41.2	58.8	0	100	0	0	0
30	D4063_F, D5097_F, D5014_F D3040_L, D4063_L, D5014_L	H06-3075	F	3	4, 5, 5	f	Adliswil Switzerland	47.3	8.56	4C	58.3	41.7	58.3	4.2	0	0	0	0
33	D3023_F, D4022_F, D5092_F D3023_L, D4022_L, D5092_L	H06-3197	F L	2 2	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	42.9	42.9	57.1	28.6	100	0	0	5.2
34	D3024_F, D4023_F, D5093_F D3024_L, D4023_L, D5093_L	H06-3214	F L	2 2	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	50	43.8	56.3	0	100	0	0	19.0
35	D3066_F, D4044_F, D5065_F D3066_L, D4044_L, D5065_L	H06-1823	F L	4 4	3, 4, 5 3, 4, 5	f	Carney MI US	45.6	87.03	4C	52.6	47.4	52.6	5.3	100	0	0	0
36	D3043_F, D4066_F, D5017_F D3043_L, D4066_L, D5017_L	H06-1964	F L	3 3	3, 4, 5 3, 4, 5	f	Tecumseh MI US	42	89.66	4C	52.6	47.4	52.6	0	100	0	0	8.2
38	D3039_F, D4062_F, D5013_F D3039_L, D4062_L, D5013_L	H06-3279	F L	3 3	3, 4, 5 3, 4, 5	m	Praha Czech	50.08	14.46	4C	52.4	47.6	52.4	4.8	100	0	0	0
39	D3010_F, D4093_F, D5044_F D3010_L, D4093_L, D5044_L	H06-3286	F L	1 1	3, 4, 5 3, 4, 5	f	Praha Czech	50.08	14.46	6C	57.1	47.6	52.4	0	0	0	100	0
40	D3057_F, D4034_F, D6031_F D3057_L, D4034_L, D5055_L	H06-1533	F L	4 4	3, 4, 6 3, 4, 5	m	Kewaunee MI US	44.61	88.11	6C	50	45	50	0	0	0	100	6.0
41	D3022_F, D4021_F, D5091_F D3022_L, D4021_L, D5091_L	H06-3232	F L	2 2	3, 4, 5 3, 4, 5	f	Silwood England	51.5	0.46	4C	55	50	50	10	100	0	0	0.8
42	D3016_F, D4010_F, D5080_F D3016_L, D4010_L, D5080_L	H06-1427	F L	2 2	3, 4, 5 3, 4, 5	f	Green Lake WI US	43.85	89.3	4C	57.1	50	50	17.9	0	100	0	0.4
43	D3003_F, D4075_F, D5026_F D3003_L, D4075_L, D5026_L	H06-1665	F L	1 1	3, 4, 5 3, 4, 5	f	Rideau River ON Canada	45	75.62	4C	50	50	50	15.4	0	60	40	1.4
44	D1001_F, D1022_F, D2001_F D3074_L, D4076_L, D5027_L	H06-1979	F L	1 1	1, 1, 2 3, 4, 5	f	Tecumseh MI US	42	89.66	4C	61.1	50	50	0	0	100	0	83.2
45	D1030_F, D1009_F, D2009_F D3084_L, D4091_L, D5042_L	H06-2945	F L	1 1	1, 1, 2 3, 4, 5	f	Clapier France	44.1	7.41	4C	63.6	50	50	4.5	0	100	0	38.4
46	D6023_F D3075_L, D4077_L, D5028_L	H06-3155	F	1	6	l	DeSteege Netherlands	52	6.05	4C	77.8	51.9	48.1	14.8	100	0	0	108.8
47	D3017_F, D4013_F, D5083_F D3017_L, D4013_L, D5083_L	H06-1513	F L	2 2	3, 4, 5 3, 4, 5	f	Kewaunee MI US	44.61	88.11	4C	68	52	48	16	0	0	100	0

GT	LC-MS samples	Line ID	O	Q	Day	F	Locality	Lat	Lon	PI	Apo	Part	Fert	Poly	re	gr	bl	DG
50	D3011_F, D4003_F, D5073_F D3011_L, D4003_L, D5073_L	H06-3205	F	2	3, 4, 5	f	Silwood England	51.5	0.46	4C	58.3	52.8	47.2	36.1	100	0	0	2.6
51	D6014_F, D6022_F, D6033_F D3065_L, D4043_L, D5064_L	H06-2464	F	4	6, 6, 6	m	Granera Spain	41.73	2.06	4C	58.8	47.1	47.1	0	0	100	0	80.4
52	D3051_F, D4028_F, D6030_F D3051_L, D4028_L, D5049_L	H06-3087	F	4	3, 4, 5	m	Adliswil Switzerland	47.3	8.56	4C	64.7	52.9	47.1	11.8	0	95	5	47.2
53	D1006_F, D1027_F, D2006_F D3080_L, D4086_L, D5037_L	H06-2967	F	1	1, 1, 2	f	Clapier France	44.1	7.41	4C	100	96.5	45.5	0	0	100	0	37.0
54	D3060_F, D4037_F, D5058_F D3060_L, D4037_L, D5058_L	H06-3367	F	4	3, 4, 5	f	Hamburg Germany	52.86	10.58	4C	55.2	55.2	44.8	17.2	100	0	0	0
55	D3008_F, D4088_F, D5039_F D3008_L, D4088_L, D5039_L	H06-1681	F	1	3, 4, 5	f	Menominee MI US	45.21	87.75	4C	63.0	56.6	44.4	11.1	80	0	20	0
56	D1005_F, D1026_F, D2005_F D3079_L, D4083_L, D5034_L	H06-2887	F	1	1, 1, 2	f	Badia Polesine Italy	45.08	11.48	4C	72	56	44	8	0	0	100	102
58	D1014_F, D1035_F, D2014_F D3088_L, D4009_L, D5079_L	H06-3340	F	2	1, 1, 2	f	Bonn Germany	50.73	7.1	4C	62.2	56.8	43.2	35.1	30	70	0	2.4
59	D3001_F, D4073_F, D5024_F D3001_L, D4073_L, D5024_L	H06-2925	F	1	3, 4, 5	f	Badia Polesine Italy	45.08	11.48	4C	73.7	57.9	42.1	0	0	0	100	6.8
60	D1003_F, D1024_F, D2003_F D3077_L, D4081_L, D5032_L	H06-2874	F	1	1, 1, 2	f	Badia Polesine Italy	45.08	11.48	4C	77.8	59.3	40.7	14.8	0	0	100	0
61	D6011_F, D6020_F, D6032_F D3059_L, D4036_L, D5057_L	H06-2031	F	4	6, 6, 6	f	Bass Lake CA US	38.66	119.5	4C	84	60	40	12	20	80	0	0
63	D3014_F, D4006_F, D5076_F D3014_L, D4006_L, D5076_L	H06-3360	F	2	3, 4, 5	f	Hamburg Germany	52.86	10.58	4C	69.2	53.8	38.5	7.7	80	0	20	0
65	D4071_F, D5099_F, D5022_F, D6008_F, D6028_F D3048_L, D4071_L, D5022_L	H06-1435	F	3	4, 5, 5	f	Green Lake WI US	43.85	89.3	4C	95.7	87	13.0	0	0	100	0	0.8
66	D3062_F, D4039_F, D5060_F D3062_L, D4039_L, D5060_L	H06-2458	F	4	3, 4, 5	m	Corvallis OR US	44.61	123.2	4C	95.7	87	13.0	0	0	0	0	78
67	D3002_F, D4074_F, D5025_F D3002_L, D4074_L, D5025_L	H06-3154	F	1	3, 4, 5	f	DeSieg Netherlands	52	6.05	4C	45.7	87	13.0	50	100	0	0	102.2
68	D3032_F, D4055_F, D5006_F D3032_L, D4055_L, D5006_L	H06-1376	F	3	3, 4, 5	f	Tuscola IL US	39.8	88.28	4C	95.7	87	13.0	0	0	100	0	0
69	D3030_F, D4053_F, D5004_F D3030_L, D4053_L, D5004_L	H06-1378	F	3	3, 4, 5	f	ESGR MI US	42.5	83.7	4C	91.7	87.5	12.5	4.2	0	0	100	0.6
70	D3035_F, D4058_F, D5009_F D3035_L, D4058_L, D5009_L	H06-1765	F	3	3, 4, 5	f	Wausaukee WI US	45.45	87.63	4C	91.7	87.5	12.5	0	0	100	0	0
72	D3036_F, D4059_F, D6026_F D3036_L, D4059_L, D5010_L	H06-3395	F	3	3, 4, 6	m	Keweenaw MI US	44.61	88.11	4C	91.7	87.5	12.5	8.3	0	0	100	0
73	D1008_F, D1029_F, D2008_F D3083_L, D4090_L, D5041_L	H06-1877	F	1	1, 1, 2	f	Carney MI US	45.6	87.03	4C	95.8	87.5	12.5	0	100	0	0	0
74	D3031_F, D4054_F, D5005_F D3031_L, D4054_L, D5005_L	H06-1643	F	3	3, 4, 5	f	Rideau River ON Canada	45	75.62	4C	92	88	12	4	0	60	40	0
75	D1013_F, D1034_F, D2013_F D3087_L, D4007_L, D5077_L	H06-1591	F	2	1, 1, 2	f	Gillett WI US	44.84	88.61	4C	92	88	12	4	100	0	0	1.8
76	D3050_F, D4026_F, D5047_F D3050_L, D4026_L, D5047_L	H06-1994	F	4	3, 4, 5	f	Cazadero CA US	38.55	123.1	4C	82.4	88.2	11.8	0	0	0	0	75.6
77	D4018_F, D5088_F, D5096_F D3020_L, D4018_L, D5088_L	H06-3148	F	2	4, 5, 5	m	Altesheim Switzerland	47.48	7.61	4C	50	88.6	11.4	47.7	30	60	10	52.0
78	D3029_F, D4052_F, D5003_F D3029_L, D4052_L, D5003_L	H06-2377	F	3	3, 4, 5	f	Winchester OR US	43.26	123.4	4C	100	90	10	0	0	0	0	0.2
79	D3037_F, D4060_F, D5011_F D3037_L, D4060_L, D5011_L	H06-1960	F	3	3, 4, 5	f	Iron Mountain MI US	45.83	88.08	4C	95.2	90.5	9.5	0	0	0	100	0

GT	LC-MS samples	Line ID	O	Q	Day	F	Locality	Lat	Lon	PI	Apo	Part	Fert	Poly	re	gr	bl	DG
80	D3049_L, D4072_L, D5023_L	H06-2720	L	3	3, 4, 5		La Selva de Mar Spain	41.21	1.13	4C	95.2	90.5	9.5	0	0	100	0	0
81	D3034_F, D4057_F, D5008_F D3034_L, D4057_L, D5008_L	H06-2850	F	3	3, 4, 5	f	Badia Polesine Italy	45.08	11.48	4C	95.2	90.5	9.5	0	0	10	90	0
82	D3013_F, D4005_F, D5075_F D3013_L, D4005_L, D5075_L	H06-3251	F	2	3, 4, 5	f	Praha Czech	50.08	14.46	4C	95.2	90.5	9.5	0	100	0	0	0
83	D4061_F, D5012_F, D6005_F D3038_L, D4061_L, D5012_L	H06-1883	F	3	4, 5, 6	m	Iron Mountain MI US	45.83	88.08	4C	95.5	90.9	9.1	4.54	0	0	100	0.6
84	D3042_F D3042_L, D4065_L, D5016_L	H06-2701	F	3	3, 4, 5	f	La Selva de Mar Spain	41.21	1.13	4C	90.9	90.9	9.09	0	0	0	0	2.4
85	D1019_F, D1040_F, D2019_F D3091_L, D4040_L, D5061_L	H06-3033	F	4	1, 1, 2	f	Cerbere France	42.45	3.16	4C	95.5	90.9	9.09	0	0	100	0	107.2
86	D3055_F, D4032_F, D5053_F D3055_L, D4032_L, D5053_L	H06-2386	F	4	3, 4, 5	m	Winchester OR US	43.26	123.4	4C	100	91.3	8.7	0	0	100	0	0
87	D3044_F, D4067_F, D5018_F D3044_L, D4067_L, D5018_L	H06-2755	F	3	3, 4, 5	m	Bolzano Italy	46.51	12.15	4C	95.7	91.3	8.7	0	0	0	100	2.2
89	D1015_F, D1036_F, D2015_F D3089_L, D4012_L, D5082_L	H06-2857	F	2	1, 1, 2	f	Badia Polesine Italy	45.08	11.48	4C	95.7	91.3	8.7	0	0	5	95	58.3
90	D3041_F, D4064_F, D5015_F D3041_L, D4064_L, D5015_L	H06-1367	F	3	3, 4, 5	f	Tuscola IL US	39.8	88.28	4C	91.7	91.7	8.3	0	0	100	0	0
91	D1011_F, D1032_F, D2011_F, D3071_F, D4094_F, D5045_F D3071_L, D4094_L, D5045_L	H06-1450	F	1	1, 1, 2,	f	Point Beach WI US	44.26	87.56	4C	95.8	91.7	8.3	0	0	0	100	0
92	D3054_F, D4031_F, D5052_F D3054_L, D4031_L, D5052_L	H06-1406	F	4	3, 4, 5	m	Green Lake WI US	43.85	89.3	4C	96.2	92.3	7.7	7.7	0	100	0	1
93	D3061_F, D4038_F, D5059_F D3061_L, D4038_L, D5059_L	H06-2767	F	4	3, 4, 5	f	Bolzano Italy	46.51	12.15	4C	100	94.4	5.6	0	0	0	100	4.6
94	D4041_F, D5062_F, D6012_F D3063_L, D4041_L, D5062_L	H06-1998	F	4	4, 5, 6	m	Cazadero CA US	38.55	123.13	4C	84.2	94.7	5.3	0	0	0	0	64
95	D3066_L, D4033_L, D5054_L	H06-2706	L	4	3, 4, 5		La Selva de Mar Spain	41.21	1.13	4C	100	95	5	0	0	100	0	103.2
96	D3067_L, D4045_L, D5066_L	H06-2719	L	4	3, 4, 5		La Selva de Mar Spain	41.21	1.13	4C	100	95	5	0	0	100	0	45.2
97	D5051_F, D6010_F, D6019_F D3053_L, D4030_L, D5051_L	H06-2941	F	4	5, 6, 6	m	Clapier France	44.1	7.41	4C	90	95	5	0	0	100	0	52.2
98	D3006_F, D4084_F, D5035_F D3006_L, D4084_L, D5035_L	H06-1489	F	1	3, 4, 5	f	Kewaunee MI US	44.61	88.11	4C	81.8	95.5	4.5	4.5	0	0	100	0
99	D3007_F, D4085_F, D5036_F D3007_L, D4085_L, D5036_L	H06-1650	F	1	3, 4, 5	f	Rideau River ON Canada	45	75.62	4C	100	95.5	4.5	0	0	0	100	16
101	D1016_F, D1037_F, D2016_F, D3072_F, D4014_F, D5084_F, D6002_F, D6024_F, D6015_F, D3072_L, D4014_L, D5084_L	H06-1369	F	2	1, 1, 2,	f	Tuscola IL US	39.8	88.28	4C	91.3	95.7	4.3	0	0	100	0	0
102	D3009_F, D4089_F, D5040_F D3009_L, D4089_L, D5040_L	H06-1441	F	1	3, 4, 5	f	Green Lake WI US	43.85	89.3	4C	92	96	4	4	0	100	0	0
103	D4002_F, D4046_F, D5067_F D4002_L, D4046_L, D5067_L	HyPR-02	F	4	4, 4, 5	f	Iron Mountain MI US	45.83	88.08	4C	NA	NA	9.09	NA	NA	NA	NA	97.6

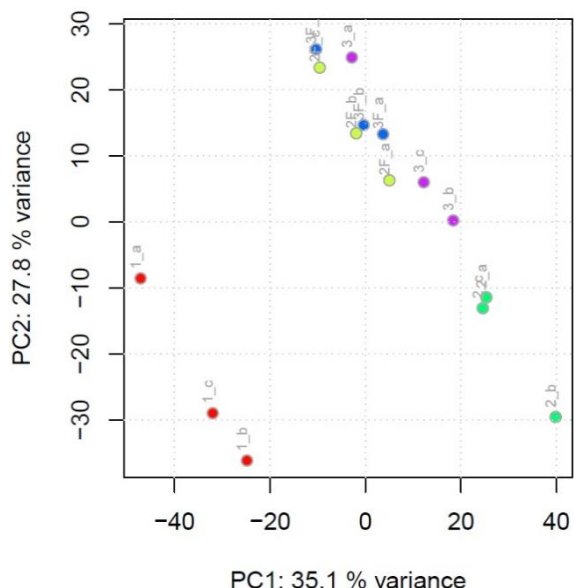
Appendix 3.2: Statistical analysis of selected correlating features of the PLS model

Results of the statistical tests of the selected correlating features of the PLS model (Table 3.1). Table sorted by correlation order with decreasing correlation factor. Analysis of variance conducted using the F-Test (p-value > 0.05 hypothesis rejected) with homogeneous variances in black and nonhomogeneous in green. In case of homoscedasticity, the t-Test was performed, in case of heteroscedasticity, the variance independent Welch two sample t-Test. Effect size defined by Cohen and power calculation was done feature-wise with the pooled variance.

Correlation order	Feature	p-value	p-value	p-value	EffectSize	Power
		F-Test	t-Test	WelchTest		
2	525.1/599	1.14E-05	3.01E-07	2.44E-05	3.951	1
3	519.1/698	8.35E-04	1.95E-07	1.14E-05	4.082	1
8	449.1/348	5.70E-03	3.81E-08	2.34E-06	4.604	1
10	313/501	1.71E-05	6.44E-08	9.90E-06	4.432	1
13	505.1/776	1.24E-04	7.71E-08	8.96E-06	4.374	1
15	699.1/664	7.68E-06	1.79E-06	6.98E-05	3.439	1
18	669.1/669	1.75E-03	1.12E-07	6.81E-06	4.255	1
22	503.1/820	1.77E-03	1.15E-08	1.67E-06	5.015	1
26	499.1/455	5.32E-05	8.66E-08	1.06E-05	4.336	1
29	532.1/426	2.44E-01	3.42E-05	6.14E-05	2.677	1
33	337.1/191	4.45E-01	4.30E-05	5.59E-05	2.622	0.999
35	509.1/684	1.62E-04	2.25E-09	1.13E-06	5.623	1
42	337.1/172	5.76E-01	9.47E-06	1.14E-05	2.998	1
44	463.1/445	1.29E-01	6.26E-08	4.42E-07	4.441	1
51	447.1/425	6.99E-01	3.48E-06	3.87E-06	3.26	1
54	301/461	6.99E-02	8.74E-09	1.69E-07	5.114	1
-46	609.1/399	2.48E-03	3.01E-04	1.17E-03	2.164	0.99

Appendix 4.1 *H. perforatum* extract stability

Methodology: The stability of the extracts during the UHPLC-MS analysis process was examined on the leaves QC (2 mg/ml, n=3). The freshly prepared samples (chapter 4.2.2) were aliquoted and part 1 was measured directly and the vial remained in the autosampler (-6°C), while part 2 was stored at -20 °C. Both samples were then measured after 42 h and both remained further in the autosampler for another injection after 24 h.



Results:

Sample treatments and colored legend

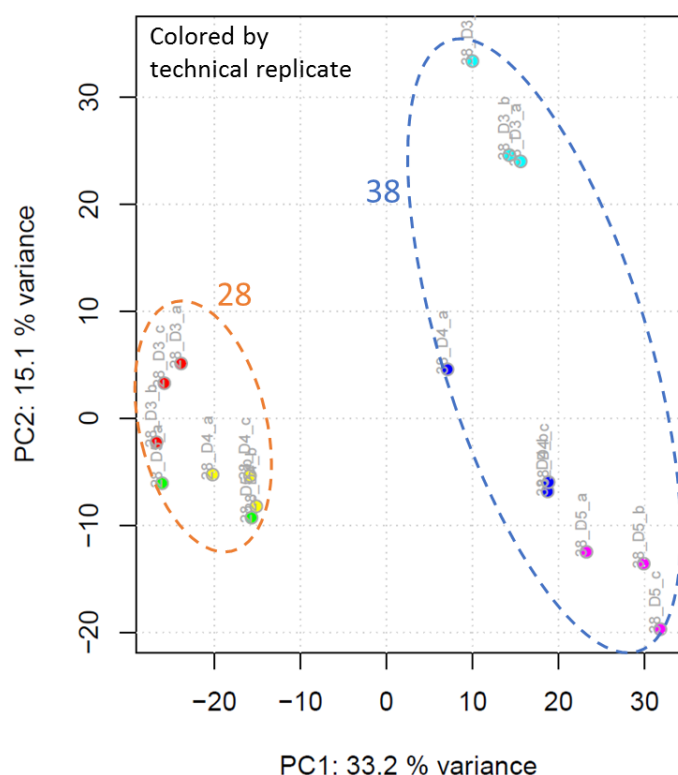
Sample abbreviation	Storage
1	fresh
2	after 42 h autosampler
3	after 66 h autosampler
2F	42 h frozen at -20 °C
3F	42 h frozen at -20 °C and 24 h in autosampler

PCA of the leaves QC samples after different storage conditions

Conclusion: The PCA shows that the fresh samples (1) differ from the autosampler stored samples (2-3). The difference between the samples, that were frozen after preparation 2F and 3F are really small, that is why all samples were frozen at least 12h to ensure a comparability.

Appendix 4.2 Evaluation of different variances within the 93 GT *H. perforatum* data set

Methodology: All replicates of two random lines (GT 28 and 38) were three times extracted and measured with UHPLC-ESI-HRMS to validate the technical and biological variance.



Results:

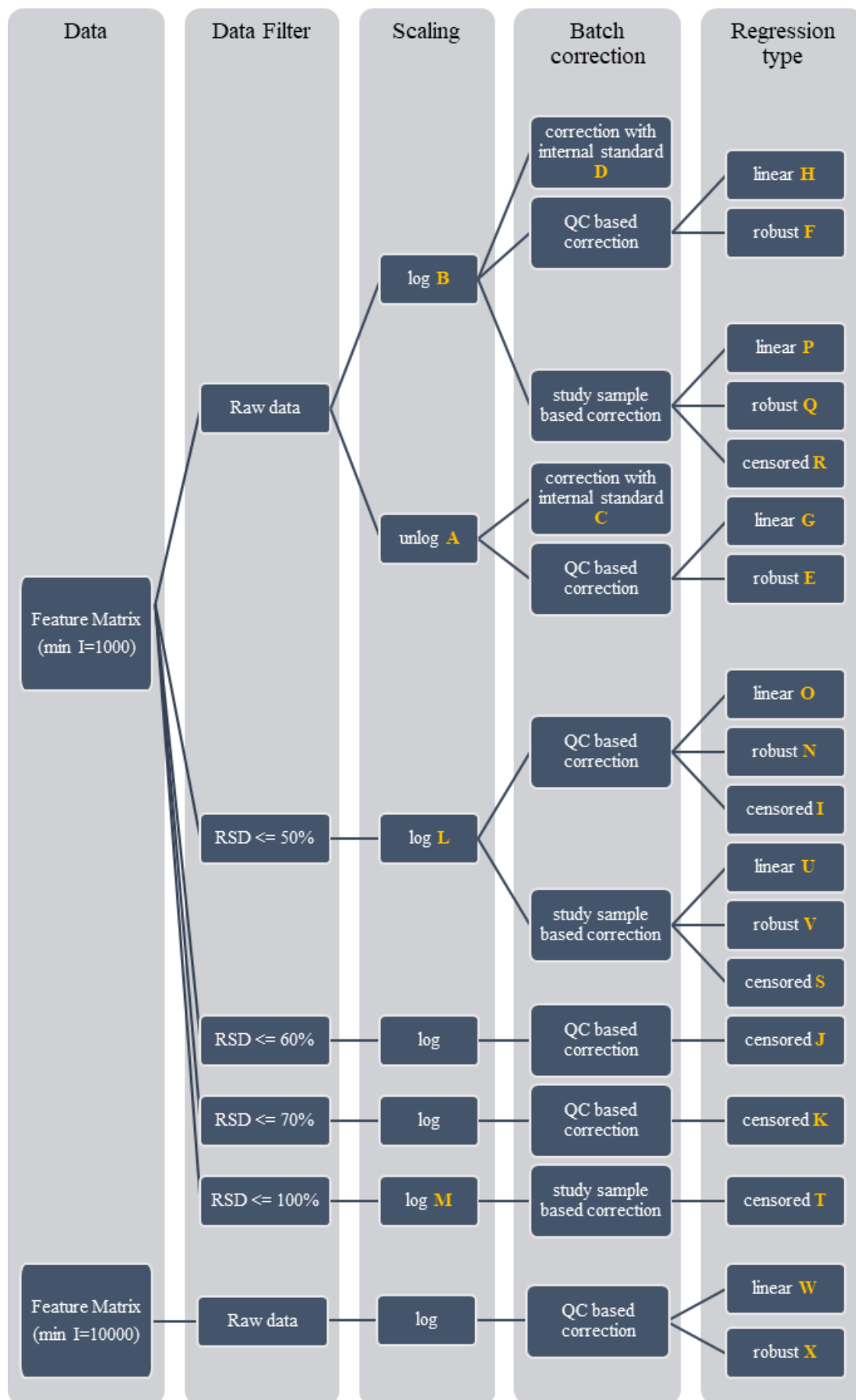
The main variance (PC1) is based on the different GTs 28 and 38. The technical replicates of the same extracted plant material cluster closer together (same color), than the biological replicates.

Conclusion: The PCA shows that the smallest variance in the data is coming from the technical aspects of the extraction and measuring process. The biological variation based on the replicates of three different fields that were harvested on three different days is smaller than the variation which is caused by the GT. This experiment shows that with our measuring method the technical variation can be neglected because the biological replicates will cause a higher variation.

Appendix 4.3 Batch correction methods with evaluation criteria

Applied Batch correction methods are shown in the scheme below (correction method abbreviation marked in yellow) and the results for the corresponding evaluation criteria are summarized in the Table.

Correction method	duplo	Bhattacharyya distance	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
A	0.376	49.388	0.509	0.484	0.429	0.528	0.526	0.422	0.423	0.417	0.522	0.467	0.526	0.528	0.424	0.422
B	0.387	27.819	3.970	11.338	0.050	0.050	0.052	0.053	0.053	0.050	0.054	0.050	0.050	0.050	0.079	0.087
C	0.403	8.651														
D	0.396	10.617														
E	0.420	0.206														
F	0.430	0.085														
G	0.424	0.240														
H	0.432	0.075														
I	0.537	0.070														
J	0.521	0.070														

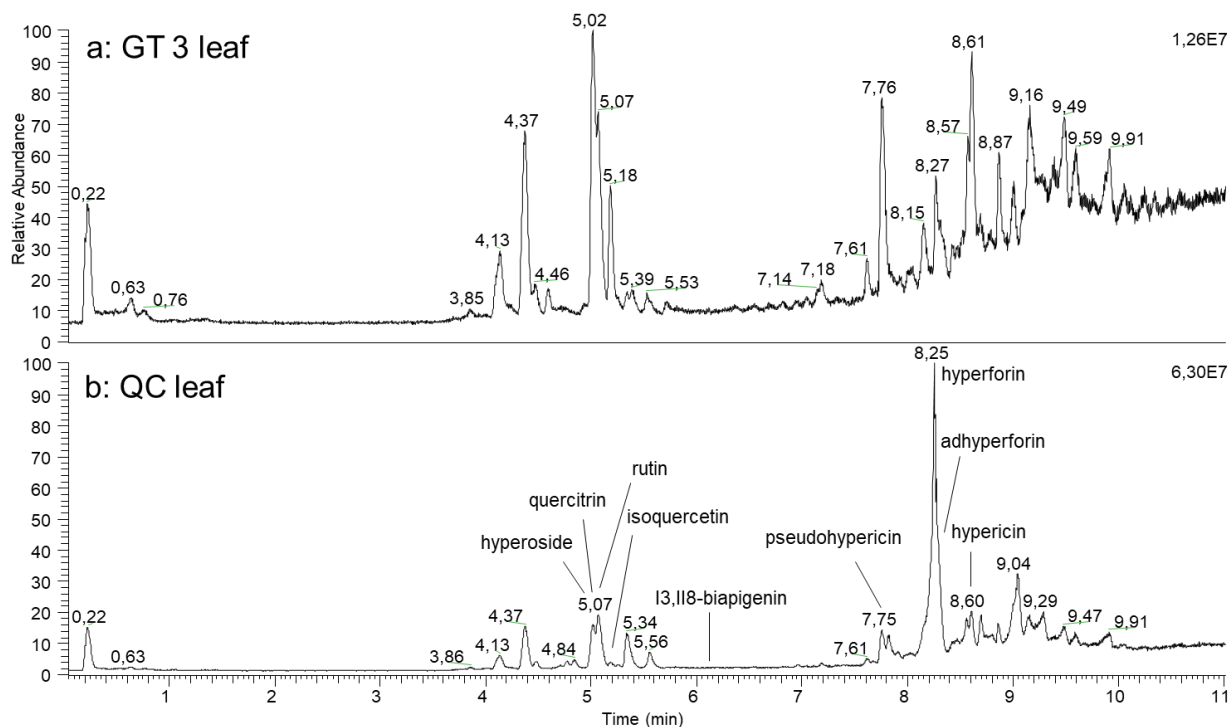


Appendix 4.4: Comparison of literature sources that characterized *H. maculatum*

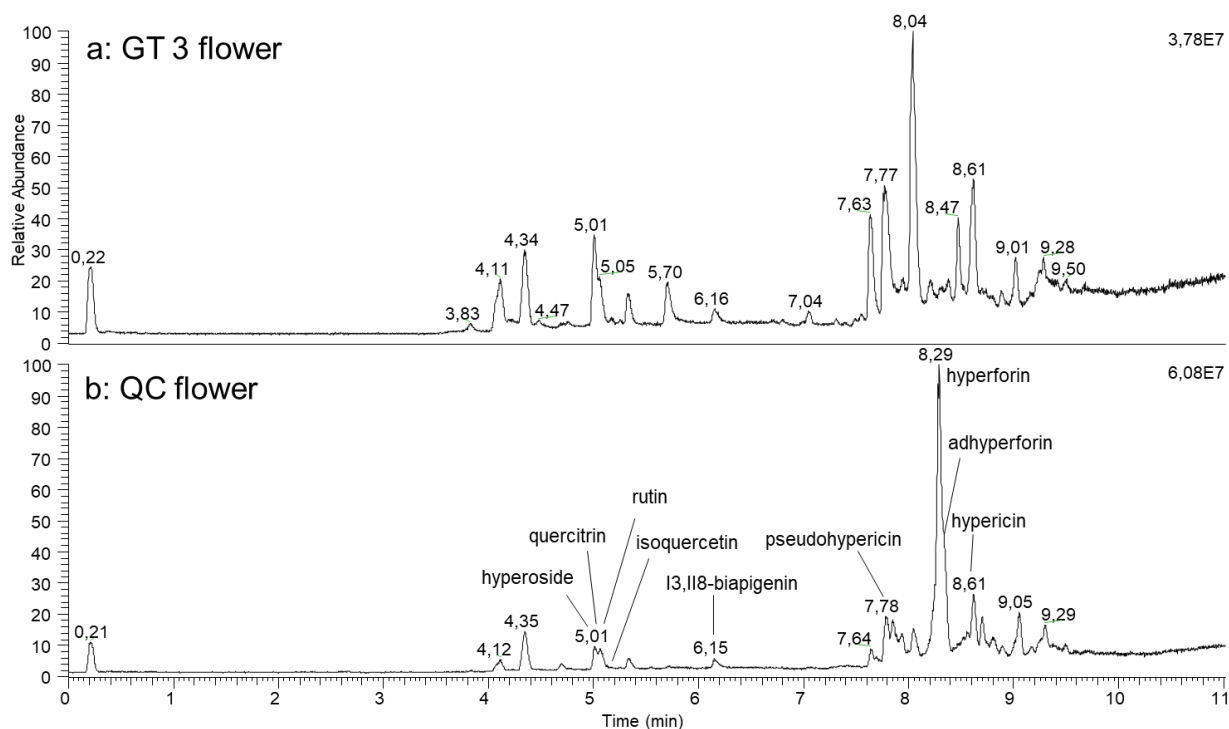
For studies that included *H. maculatum* (white) and *H. perforatum* (grey) the results of both species are shown. + compound detected; ++ in the same analysis the compound was significant higher than in the other species; - compound was not detected; (+) occurrence depends on harvest time; L in leaves only; F in flowers only.

Author	Plant origin	Total flavonoid content	Hypericin	Pseudohypericin	Emodin	Hyperforin	Chlorogenic acid	Caffeic acid	5-O-Caffeoylquinic acid	Catechin	Epicatechin	Biapigenin	Rutin	Quercetin	Quercitrin	Isoquercitrin	Amentoflavone	Hyperoside
Radusienė et al. (2004) ¹⁵²	Field	+	++ L										++		++ L	++ F		++ F
		+	+ L											+		+ L	+ F	
Zheleva- Dimitrova et al. (2012) ¹⁵⁷	wild															+		+
Kartnig et al. (1996) ²⁵⁸	<i>in vitro</i>		++	++								+	+	+	+	+	+	+
			+	+								+	+	+	+	+	+	+
Martonfi et al. (1996) ¹⁵⁵	wild		+	+		+						+	+	+	+	+		+
			+	+		-						+	-	+	+	+		+
Zapesochnaya et al. (1966) ²⁵⁹	wild		+	+			+	+						+	+			+
Coste et al. (2011) ²⁶⁰	<i>in vitro</i>		+	+		+												
Martonfi et al. (2006) ⁷⁵			+F	+F								+F		+F	+F	+F		+F
Raclariu et al. (2017) ⁵⁶	field		+			++												
			+			+												
Nigutova et al. (2017) ¹⁵³	<i>in vitro</i>	+	+	+	+	++							+		+			+
		+	+	+	+	+							+		+			+
Bagdonaite et al. (2012) ¹⁴⁷	field		+			++	+						+	+	++			+
			+			+	+						-	+	+			++
Crockett et al. (2005) ¹⁴⁸	wild		++	++		+							+	++	++	+	-	+
			+	+		+							-	+	+	+	+	-
Bozin et al. (2013) ⁶⁰	wild	+	+			+	+	+					+					
		++	-			-	++	++					++					
Kladar et al. (2015) ²⁶¹	wild		+			(+)	+						+		+		+	
Kusari et al. (2009) ⁷⁷	wild					++							+	++	+			+
						+							++	+	++			++
Kucharikova et al. (2016) ¹¹⁷	<i>in vitro</i>		+	+		+												
			+	+		-												
Umek et al. (1999) ⁷⁶	wild		++	+		+						+	++	+	++	+	+	+
			+	+		-						+	+	+	+	++	+	++
Oniga et al. (2016) ¹⁵⁰	wild						+						+	+	+	+		+
Rusalepp et al. (2017) ⁶¹	wild	+	+	+		++				+	+		++		++			
		++	+	+		+				+	++		+		+			
Smelcerovic and Spiteller (2006) ¹⁴⁴	wild		+			++									++			+
			+			+									+			++
Smelcerovic et al. (2006) ²⁰⁴	wild		+	+		+								+				+
Zdunic et al. (2017) ¹⁴⁹	wild	++							+				+					
		+							-				-					
Zheleva-Dimitrova et al. (2010) ⁹²	wild	++																
		+																

Appendix 4.5: Metabolite profiles of GT3 in comparison to the QC samples



Total ion chromatogram of *H. perforatum* a) GT3 leaf sample and b) QC leaf sample



Total ion chromatogram of *H. perforatum* a) GT3 flower sample and b) QC flower sample

Appendix 4.6 Naphthodianthrone composition of *H. perforatum* GTs

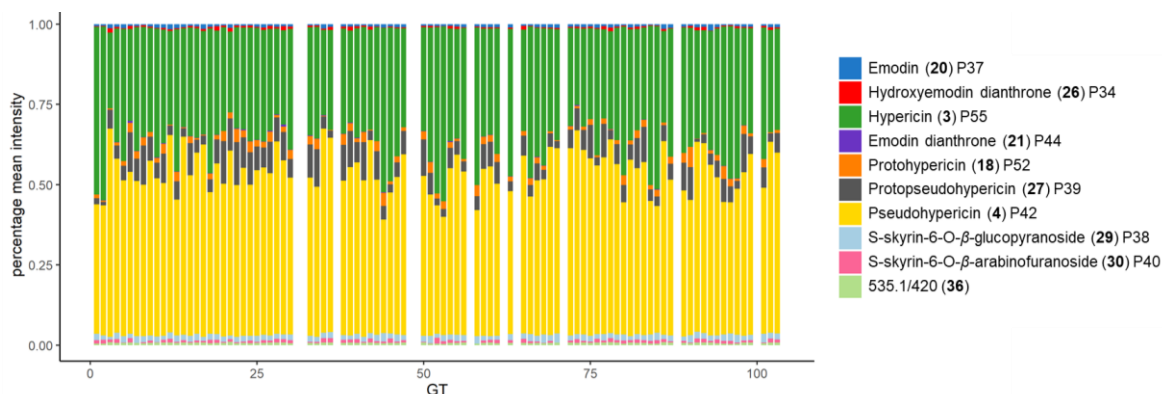


Figure 1 Percentage composition of major naphthodianthrone per GT in leaves

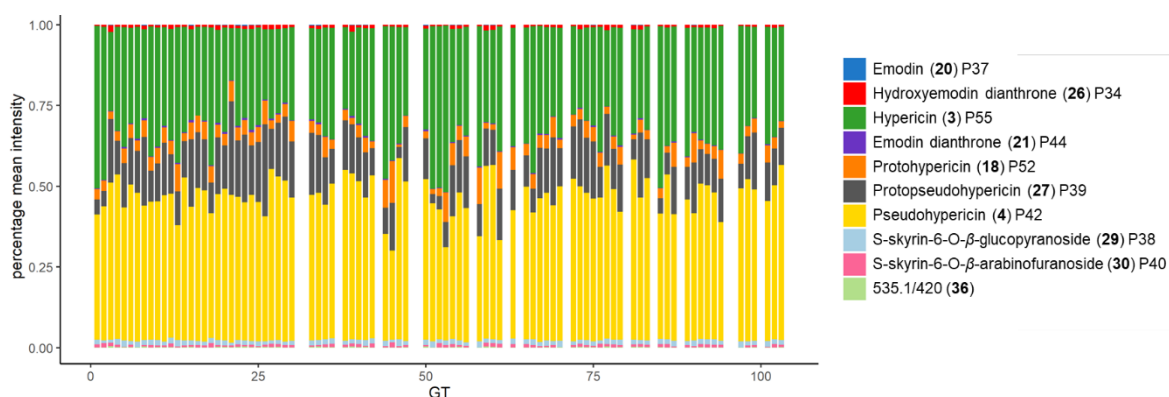


Figure 2 Percentage composition of major naphthodianthrone per GT in flowers

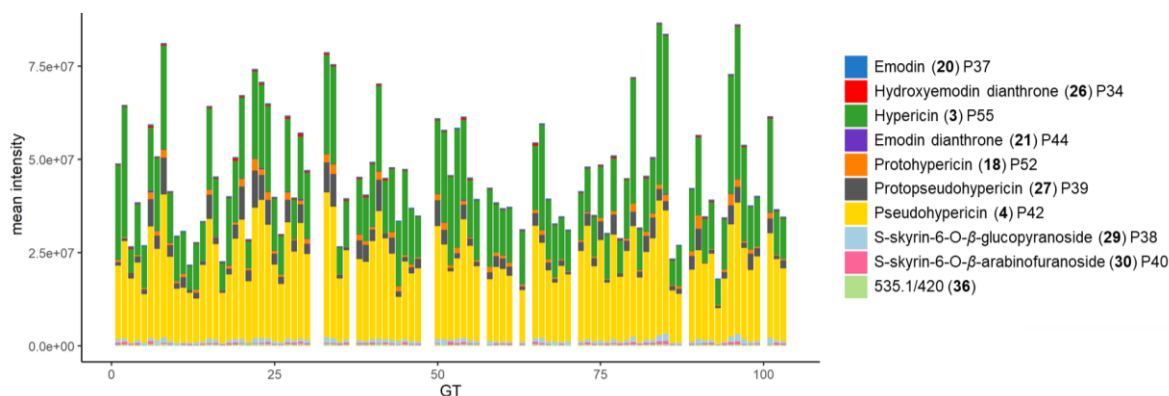


Figure 3 Composition of major naphthodianthrone per GT in leaves

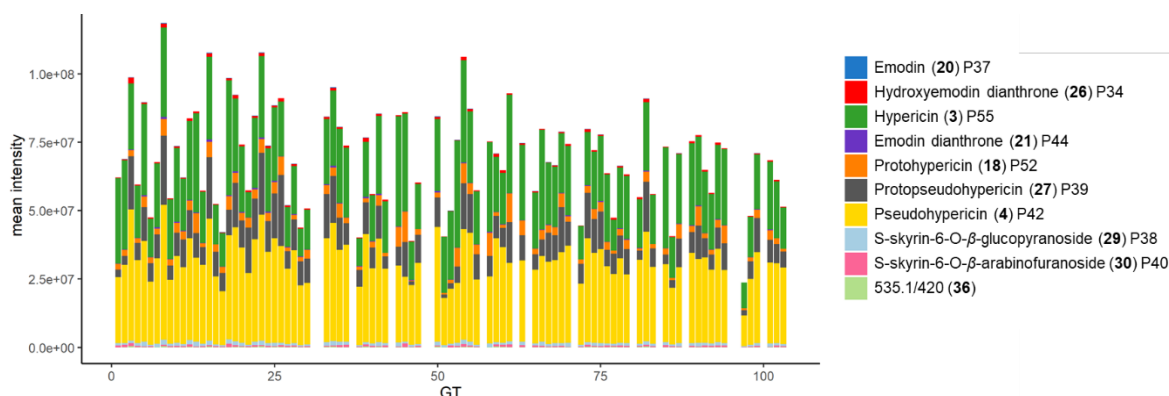


Figure 4 Composition of major naphthodianthrone per GT in flowers

Appendix 4.7 Phloroglucinol composition of *H. perforatum* GTs

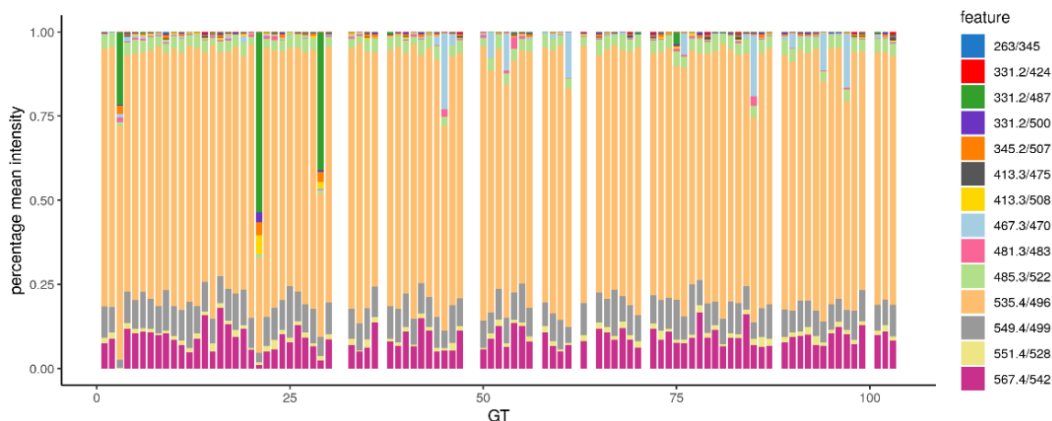


Figure 1 Percentage composition of major phloroglucinols per GT in leaves

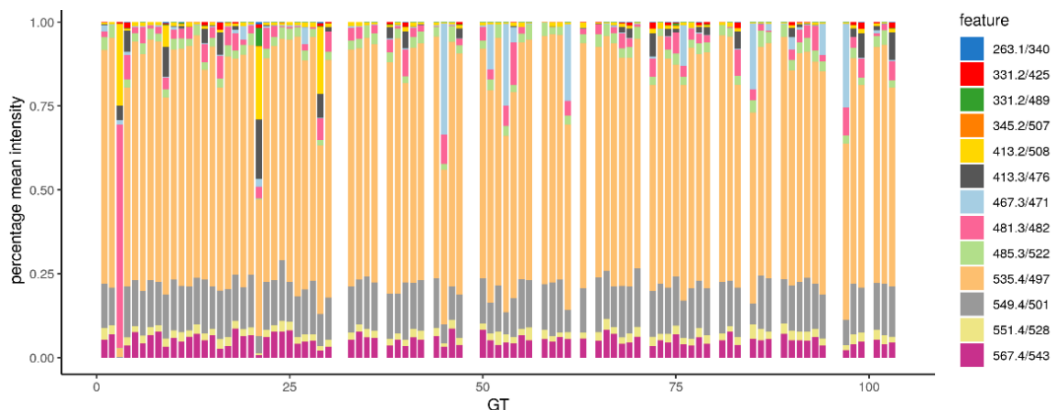


Figure 2 Percentage composition of major phloroglucinols per GT in flowers

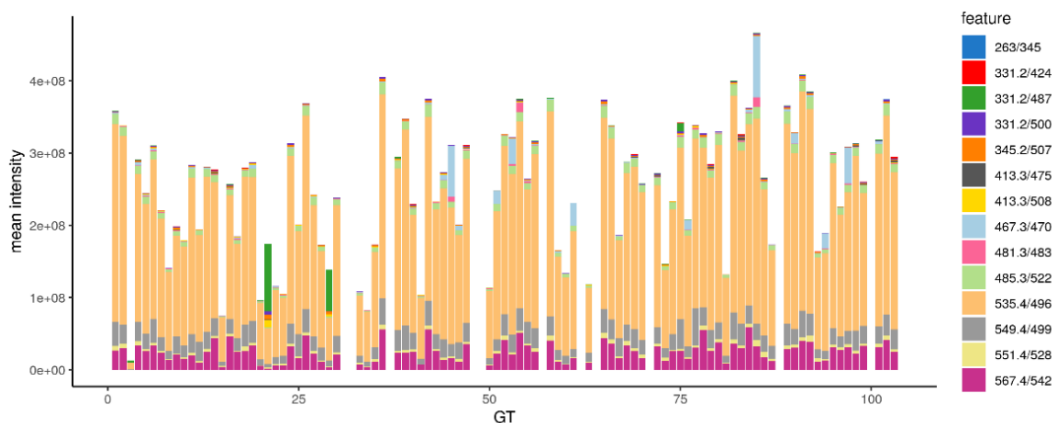


Figure 3 Composition of major phloroglucinols per GT in leaves

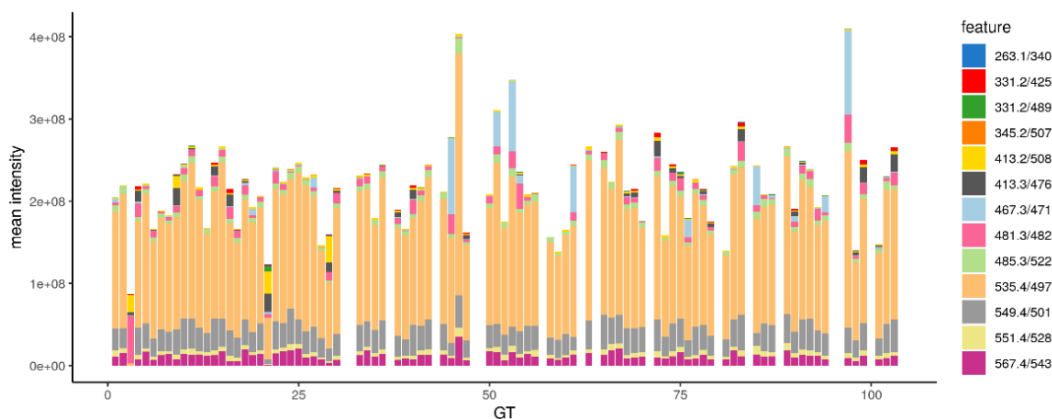


Figure 4 Composition of major phloroglucinols per GT in flowers

Appendix 4.8 Hyperforin correlated feature

Correlating features to hyperforin in *H. perforatum* flower data (correlation coefficient > 0.5). Mass fragmentation was obtained with ncd= 35.

Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	ppm	Fragments (intensity)	Identification
399.3/488	399.3265	8.1	C ₂₇ H ₄₃ O ₂	-0,561	399.3265 (50), 354.2516 (28), 329.2848 (100), 315.2327 (20), 273.2223 (30)	n.a.
425.3/495	425.2578	8.3	C ₂₃ H ₃₇ O ₇	7,156	-	n.a.
455.3/557	455.3168	9.3	C ₂₉ H ₄₃ O ₄	0,169	455.3168 (100), 437.3059 (25), 369.2433 (12), 317.1756 (35), 285.1493 (12), 195.0663 (20)	n.a.
457.3/522	457.3322	8.7	C ₂₉ H ₄₅ O ₄	0,277	-	n.a.
467.3/527	467.3168	8.8	C ₃₀ H ₄₃ O ₄	0,058	-	n.a.
467.3/549	467.3165	9.2	C ₃₀ H ₄₃ O ₄	0,935	467.3165 (5), 383.2594 (5), 275.1289 (100)	n.a.
471.3/543	471.3110	9.1	C ₂₉ H ₄₃ O ₅	-1,183	467.3165 (5), 383.2594 (5), 275.1289 (100)	n.a.
477.1/282	477.0668	4.7	C ₂₁ H ₁₇ O ₁₃	-1,412	301.0353 (100)	quercitrin (8)
481.3/554	481.3324	9.22	C ₃₁ H ₄₅ O ₄	0,077	437.3425 (8), 412.2619 (28), 385.2747 (6), 289.1444 (100), 275.1289 (17)	n.a.
485.3/522	485.3274	8.7	C ₃₀ H ₄₅ O ₅	0,396	457.3325 (100)	unidentified phloroglucinol II P56
495.3/557	495.3480	9.3	C ₃₂ H ₄₇ O ₄	-0,115	419.3322 (100)	n.a.
501.3/464	501.3014	7.7	C ₃₃ H ₄₁ O ₄	-0,732	501.0628 (90), 486.0382 (16), 457.0731 (10), 432.2307 (100), 383.2227 (14)	unidentified phloroglucinol I
501.3/505	501.3231	8.4	C ₃₀ H ₄₅ O ₆	1,830	501.0629 (100), 486.0381 (19), 483.0508 (10), 473.0668 (15), 457.0734 (11), 432.2307 (83), 383.2226 (11)	n.a.
507.3/464	507.3111	7.7	C ₃₂ H ₄₃ O ₅	-0,882	-	n.a.
507.3/491	507.3480	8.2	C ₃₃ H ₄₇ O ₄	0,073	-	n.a.
507.4/557	507.3837	9.3	C ₃₄ H ₅₁ O ₃	-1,298	-	n.a.
509.3/461	509.3	7.7	C ₃₂ H ₄₅ O ₅	-0,349	440.2566 (100), 439.2440 (14), 412.2618 (30), 383.2226 (22), 371.1855 (13), 341.2121 (10)	n.a.
509.3/514	509.3266	8.6	C ₃₂ H ₄₅ O ₅	-1,311	-	n.a.
511.3/543	511.3418	9.1	C ₃₂ H ₄₇ O ₅	-2,166	435.3272 (100), 417.3161 (3), 377.2851 (38)	n.a.
521.4/494	521.3630	8.2	C ₃₄ H ₄₉ O ₄	3,408	-	n.a.
525.3/482	525.3211	8.0	C ₃₂ H ₄₅ O ₆	-2,041	-	n.a.
535.4/497	535.3786	8.25	C ₃₅ H ₅₁ O ₄	-1,370	466.3095 (100), 397.2382 (34), 383.2234 (64), 315.1606 (42), 313.1814 (29)	hyperforin (1)
535.4/512	535.3784	8.5	C ₃₅ H ₅₁ O ₄	-1,631	491.3899 (8), 466.3093 (27), 397.2386 (100), 383.2230 (18), 315.1605 (12), 287.1291 (10), 275.1292 (97)	n.a.
535.4/530	535.3791	8.8	C ₃₅ H ₅₁ O ₄	-0,324	492.3250 (100), 465.3380 (75), 423.2545 (17), 343.1917 (15), 300.1369 (60), 257.0823 (10)	n.a.
535.4/553	535.3788	9.2	C ₃₅ H ₅₁ O ₄	-0,921	-	n.a.
539.3/480	539.3370	8.0	C ₃₃ H ₄₇ O ₆	-1,580	-	n.a.
539.4/543	539.3731	9.1	C ₃₄ H ₅₁ O ₅	-2,091	-	n.a.
541.4/460	541.3540	7.7	C ₃₃ H ₄₉ O ₆	0,993	-	n.a.
549.4/476	549.3569	7.9	C ₃₅ H ₄₉ O ₅	-3,054	480.2882 (100), 411.2178 (25), 383.2229 (38), 315.1605 (12)	n.a.
549.4/501	549.3945	8.31	C ₃₆ H ₅₃ O ₄	-0,734	480.3243 (100), 411.2539 (21), 397.2382 (38), 329.1756 (32), 313.1807 (26)	adhyperforin (2)
549.4/516	549.3944	8.6	C ₃₆ H ₅₃ O ₄	-0,953	505.4059 (7), 480.3267 (5), 411.2542 (95), 301.1446 (10), 289.1446 (100)	n.a.
549.4/562	549.4070	9.4	C ₃₆ H ₅₃ O ₄	0,558	480.3248 (62), 465.2996 (22), 411.2541 (100), 397.2385 (97), 329.1759 (55), 289.1447 (16)	n.a.

Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	ppm	Fragments (intensity)	Identification
551.4/557	551.3726	9.3	C ₃₅ H ₅₁ O ₅	0,747	523.3792 (15), 507.3841 (31), 456.3165 (100), 437.3057 (11), 399.3267 (7), 329.2850 (20)	furohyperforin isomer II (32b)
551.4/528	551.3741	8.71	C ₃₅ H ₅₁ O ₅	-0.25	483.3087 (25), 482.3041 (100), 467.2800 (11), 413.2335 (41), 411.2545 (89), 399.2181 (87), 331.1553 (65), 329.1766 (53)	furohyperforin isomer I (32)
565.4/563	565.3894	9.4	C ₃₆ H ₅₃ O ₅	-0,881	537.3947 (18), 521.3999 (25), 469.3320 (100), 451.3215 (14), 413.3423 (5), 329.2849 (14)	furoadhyperforin
567.4/470	567.3675	7.8	C ₃₅ H ₅₁ O ₆	-2.877	549.3586 (100), 535.3432 (37), 523.3796 (67), 498.2986 (45), 427.2489 (53)	n.a.
567.4/525	567.3683	8.8	C ₃₅ H ₅₁ O ₆	-1.520	539.3738 (18), 523.3791 (55), 498.2967 (20), 471.3114 (100), 453.3002 (18), 415.3216 (32), 330.2156 (25)	hydroperoxy furohyperforin
567.4/543	567.3696	9.05	C ₃₅ H ₅₁ O ₆	0.824	471.3116 (100), 497.3272 (14), 453.3007(15), 415.3218 (12)	unidentified phloroglucinol IV P59
569.4/464	569.3829	7.7	C ₃₅ H ₅₃ O ₆	-3.043	536.3467 (29), 500.3146 (100), 431.2439 (20), 383.2228 (90), 347.1863 (15), 275.1291 (8)	n.a.
581.4/474	581.3842	7.9	C ₃₆ H ₅₃ O ₆	-1.036	-	n.a.
581.4/549	581.3824	9.2	C ₃₆ H ₅₃ O ₆	1,114	553.3888 (8), 537.3939 (8), 497.3262 (22), 485.3263 (100), 467.3157 (27), 429.3366 (18), 411.3263 (5), 383.2585 (10)	hydroperoxy furoadhyperforin
583.4/505	583.3632	8.4	C ₃₅ H ₅₁ O ₇	0,913	-	n.a.
585.4/440	585.3784	7.3	C ₃₅ H ₅₃ O ₇	-2.199	-	n.a.
603.4/515	603.4409	8.6	C ₄₀ H ₅₉ O ₄	-1,630	-	n.a.
611.4/526	611.3939	8.8	C ₃₇ H ₅₅ O ₇	-2.220	-	n.a.
625.4/530	625.4103	8.8	C ₃₈ H ₅₇ O ₇	-1,067	-	n.a.
627.4/505	627.3896	8.4	C ₃₇ H ₅₅ O ₈	-1.007	-	n.a.
627.4/526	627.3892	8.8	C ₃₇ H ₅₅ O ₈	-1.645	-	n.a.
635.4/502	635.4360	8.4	C ₃₃ H ₆₃ O ₁₁	-2.543	-	n.a.
637.4/506	637.4475	8.4	C ₄₀ H ₆₁ O ₆	0.262	-	n.a.

Appendix 4.9 Correlating features to less prenylated phloroglucinols

Correlating features to less prenylated phloroglucinols in *H. perforatum* flower data (correlation coefficient > 0.5). Mass fragmentation was obtained with ncd= 35.

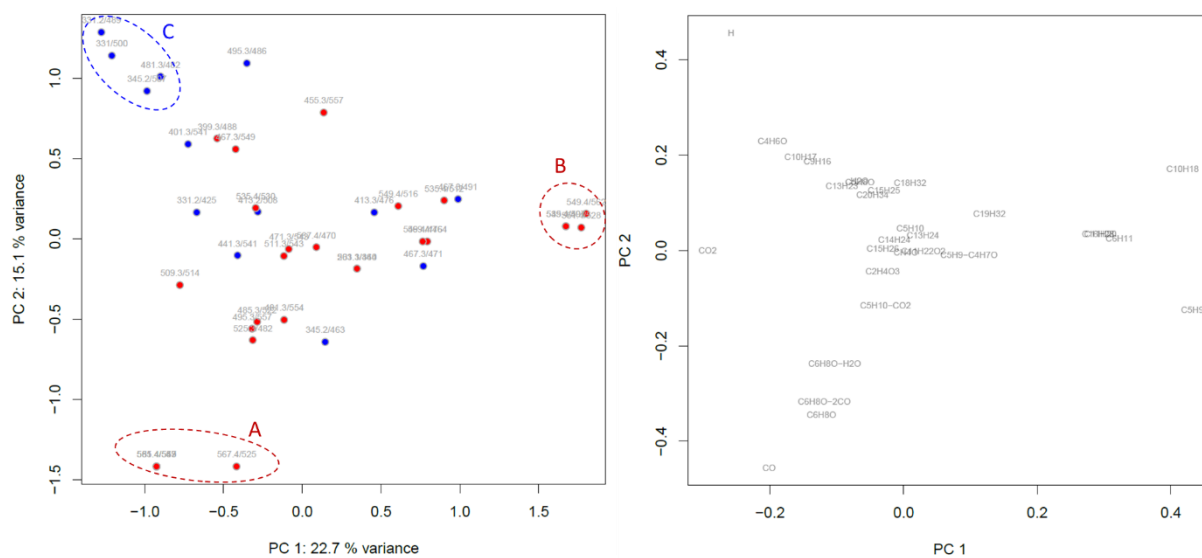
Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	ppm	Fragments (intensity)	Identification
263.1/340	263.1289	5.63	C ₁₅ H ₁₉ O ₄	-0.769	194.0583 (100), 166.0636 (12), 151.0039 (3)	Dimethylallyl-phlorisobutyrophenone (15)
331.2/500	331.1914	8.33	C ₂₀ H ₂₇ O ₄	-0.038	287.2018 (78), 261.1099 (5), 207.0664 (35) 194.0586 (100), 152.0118 (14)	2-O-geranyl-methylpropanoyl-phloroglucinol (41)
331.2/489	331.1914	8.14	C ₂₀ H ₂₇ O ₄	-0.189	331.1919 (24), 313.1812 (11), 287.2020 (100), 262.1214 (10), 261.1501 (17), 207.0666 (24), 194.0587 (6)	3-geranyl-methylpropanoyl-phloroglucinol (17)
331.2/425	331.1912	7.08	C ₂₀ H ₂₇ O ₄	-0.944	331.1912 (6), 262.1209 (5), 247.0974 (6), 194.0583 (100), 166.0636 (43), 151.0037 (13)	O-geranyl-methylpropanoyl-phloroglucinol (43)
345.2/463	345.1707	7.70	C ₂₀ H ₂₅ O ₅	-0.137	345.1707 (20), 317.1757 (100), 276.1003 (19), 263.0690 (10)	n.a.
345.2/507	345.2070	8.44	C ₂₁ H ₂₉ O ₄	-0.326	316.1996 (14), 301.2169 (64), 209.0818 (17), 208.0740 (100), 152.0116 (18)	2-O-geranyl-methylbutanoyl-phloroglucinol (42)
389.2/462	389.2330	7.7	C ₂₃ H ₃₃ O ₅	-0.790	-	n.a.
399.3/514	399.2547	8.6	C ₂₅ H ₃₅ O ₄	1.446	399.2538 (100), 370.0625 (6), 355.2639 (15), 329.2848 (16)	n.a.

Feature	[M-H]-	Rt (min)	Molecular Formula [M-H]-	ppm	Fragments (intensity)	Identification
401.3/541	401.2704	9.0	C ₂₅ H ₃₇ O ₄	0,068	401.2704 (100), 383.2593 (25), 357.2797 (7), 331.1913 (5), 313.1811 (7), 263.1289 (8), 231.1027 (4), 195.0660 (9)	unidentified phloroglucinol III P58
403.2/527	403.2494	8.8	C ₃₀ H ₄₃ O ₄	0,998	-	n.a.
413.3/476	413.2702	7.91	C ₂₆ H ₃₇ O ₄	1.275	369.2794 (21), 276.1365 (58), 275.1289 (35), 233.0819 (100), 221.0819 (21), 208.0741 (40)	deprenylhyper-polyphyllirin (38)
413.2/508	413.2333	8.46	C ₂₅ H ₃₃ O ₅	-0.236	413.2333 (20), 343.1913 (100)	n.a.
441.3/541	441.3016	9.0	C ₂₈ H ₄₁ O ₄	1,285	397.3107 (3), 365.2846 (100)	n.a.
449.1/300	449.1088	4.99	C ₂₁ H ₂₁ O ₁₁	-0.411	303.0506 (100), 285.0401 (81.75%), 151.0038 (23), 323.0768 (12)	astilbin (19)
457.3/448	457.2602	7.5	C ₂₇ H ₃₇ O ₆	1.417	-	n.a.
467.3/471	467.3163	7.82	C ₃₀ H ₄₃ O ₄	-0.863	398.2462 (100), 383.2229 (7), 329.1760 (5), 327.1968 (6), 271.1341 (6), 234.1264 (3)	hyperfirin (40)
467.3/491	467.3169	8.2	C ₃₀ H ₄₃ O ₄	0.657	451.2858 (21), 423.2539 (), 397.2384 (63), 383.2229 (74), 329.1759 (100), 315.1604 (74), , 287.1289 (61), 275.1290 (17), 219.0665 (14)	n.a.
481.3/482	481.3326	8.04	C ₃₁ H ₄₅ O ₄	0.638	481.3333 (10), 437.3430 (100), 411.2911 (80), 343.1918 (25), 301.2173 (10), 276.1368 (29), 233.0821 (57), 207.0666 (28)	hyperpolyphyllirin (39)
481.3/554	481.3328	9.2	C ₃₁ H ₄₅ O ₄	0,201	437.3425 (7), 412.2619 (25), 385.2747 (6), 289.1444 (100), 275.1289 (16)	n.a.
495.3/486	495.3489	8.1	C ₃₂ H ₄₇ O ₄	1,366	495.3489 (5), 451.3586 (35), 411.2904 (100), 357.2062 (24), 290.1523 (12), 247.0976 (26), 221.0821 (12)	adhyperpolyphyllirin
497.3/441	497.3276	7.4	C ₃₀ H ₄₃ O ₅	0.708	-	n.a.
497.3/477	497.3282	8.0	C ₃₁ H ₄₅ O ₅	1.854	-	n.a.
497.3/542	497.3278	9.0	C ₃₁ H ₄₅ O ₅	1.151	-	n.a.
509.3/514	509.3257	8.6	C ₃₂ H ₄₅ O ₅	1,654	481.3321 (100), 465.3372 (52), 439.2852 (22), 398.2460 (15), 371.2588 (26), 327.2692 (24)	n.a.
525.3/482	525.3222	8.0	C ₃₂ H ₄₅ O ₆	0.205	497.3267 (42), 481.3326 (100), 455.2772 (9), 414.2407 (7), 387.2537 (10), 371.2591 (39), 345.2066 (4)	n.a.
529.3/470	529.3185	7.8	C ₃₁ H ₄₅ O ₇	2.688	-	n.a.
651.2/291	651.1547	4.83	C ₂₉ H ₃₁ O ₁₇	-1.963	609.1457 (40), 591.1348 (100), 301.0353 (52), 300.0276 (90), 299.0196 (20), 271.0248 (18)	acetyl rutin I P13

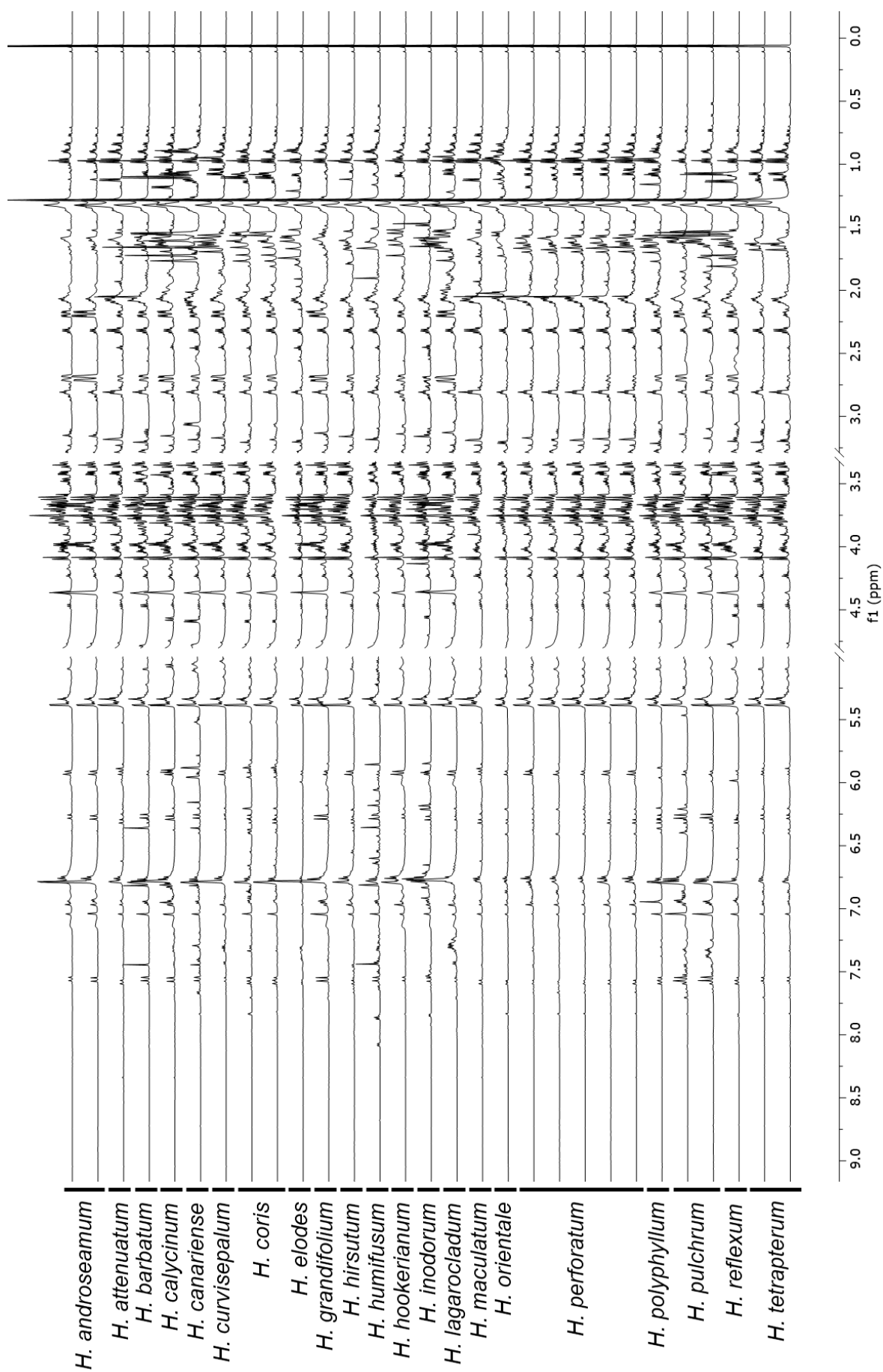
Appendix 4.10: MS² fragments of correlating polyprenylated acylphloroglucinols

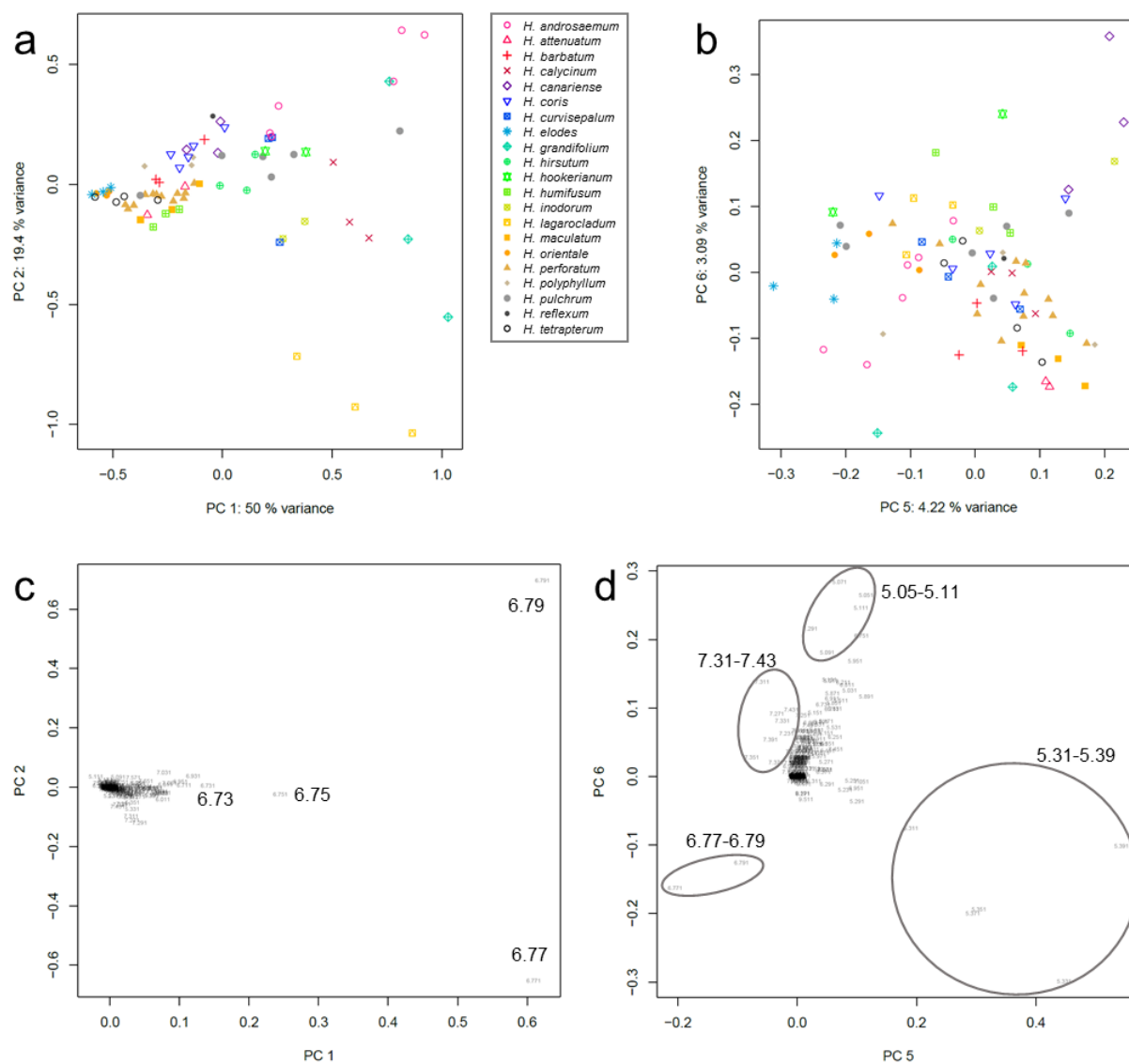
Losses that occur multiple times were listed for each feature with the *m/z* of the fragment and color-coded by the fragment intensity. This intensity matrix was used for PCA after log transformation. The PCA scores plot (see below) does not separate the compounds correlating with hyperforin (red), and the compounds correlating with phloroglucinols with less prenylated side chains (blue). So we concluded that those compounds are not distinguishable by one typical structural element.

Identification	Subgroup PCA	Feature	[M-H] ⁻	[M-H-H ₂ O] ⁻	[M-H-CO] ⁻	[M-H-CO ₂] ⁻	[M-H-CH ₃ O] ⁻	[M-H-C ₉ H ₁₆] ⁻	[M-H-C ₈ H ₆ O-H ₂ O] ⁻	[M-H-C ₈ H ₆ O-2CO] ⁻	[M-H-C ₈ H ₆ O] ⁻	[M-H-C ₈ H ₁₁] ⁻	[M-H-C ₅ H ₁₀] ⁻	[M-H-C ₅ H ₁₀ -CO ₂] ⁻	[M-H-C ₅ H ₉ -C ₄ H ₇ O] ⁻	[M-H-C ₅ H ₉] ⁻	[M-H-C ₅ H ₆ O] ⁻	[M-H-C ₄ H ₆ O] ⁻	[M-H-C ₂ H ₄ O ₃] ⁻	[M-H-C ₂₀ H ₃₄] ⁻	[M-H-C ₁₉ H ₃₂] ⁻	[M-H-C ₁₈ H ₃₂] ⁻	[M-H-C ₁₆ H ₂₆] ⁻	[M-H-C ₁₅ H ₂₆] ⁻	[M-H-C ₁₅ H ₂₅] ⁻	[M-H-C ₁₄ H ₂₄] ⁻	[M-H-C ₁₃ H ₂₄] ⁻	[M-H-C ₁₃ H ₂₃] ⁻	[M-H-C ₁₁ H ₂₀] ⁻	[M-H-C ₁₀ H ₁₈] ⁻	[M-H-C ₁₀ H ₁₇] ⁻						
		399.3/488	399																																		
		455.3/557	455																																		
		467.3/549		437																																	
		471.3/543																																			
P56		485.3/522			457				437																												
		495.3/557			465																																
P41		501.3/464																																			
		509.3/514																																			
		511.3/543																																			
		525.3/482																																			
1	B	535.4/497																																			
		535.4/512																																			
		535.4/530																																			
		549.4/476																																			
2	B	549.4/501																																			
		549.4/516																																			
	B	549.4/562																																			
32	B	551.4/528																																			
	A	551.4/557																																			
32b	A	555.4/563																																			
	A	567.4/470		549																																	
	A	567.4/525																																			
	A	569.4/464																																			
	A	581.4/549																																			
15		263.1/340																																			
P35		331.2/425																																			
17	C	331.2/489																																			
P51	C	331.5/500																																			
P53	C	345.2/463																																			
P58		401.3/541																																			
37		413.2/508																																			
38		441.3/476																																			
40		467.3/471																																			
39	C	481.3/482																																			
		495.3/486																																			



PCA based on fragmentation pattern of phloroglucinols (corresponding to the table above) (left: scores plot, right loadings plot) colored by correlation behavior: red= correlates with hyperforin related compounds, blue = correlates with lower prenylated phloroglucinols

Appendix 6.1 Stacked $^1\text{H-NMR}$ spectra of *Hypericum* species

Appendix 6.2 PCA of of $^1\text{H-NMR}$ aromatic range

PCA of $^1\text{H-NMR}$ aromatic range ($\delta > 5$ ppm): scores plots are coloured by species and show a) PC1 and PC2 and b) PC5 and PC6. The corresponding loadings plots highlighted with discriminant chemical shifts.

Appendix 6.3 Peak table of major compounds based on the UHPLC-MS analysis of 21 *Hypericum* species

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P1	191.1/13	191.056	0.22	C ₇ H ₁₁ O ₆	-0.53	173.0458 (72), 127.0404 (85), 85.0298 (100)	citric acid
P2	173/14	173.0455	0.23	C ₇ H ₉ O ₅	-0.096	-	shikimic acid (11)
P3	341.1/15	341.1088	0.25	C ₁₂ H ₂₁ O ₁₁	-0.365	179.0563 (100), 161.0458 (30), 143.0353 (34), 119.0352 (21), 113.0247 (38)	sucrose (12)
P4	353.1/44	353.0877	0.73	C ₁₆ H ₁₇ O ₉	-0.44	191.0563 (100), 179.0351 (31), 135.0455 (6)	chlorogenic acid (10)
P5	289.1/146	289.0713	2.43	C ₁₅ H ₁₃ O ₆	-1.562	245.0821 (100), 205.0508 (32), 179.0352 (15)	epicatechin (14a)
P6	209.1/195	209.0818	3.25	C ₁₁ H ₁₃ O ₄	-0.489	194.0587 (38), 181.0872 (15), 166.0274 (820), 139.0404 (852), 137.0975 (53), 111.0454 (100)	n.a.
P7	407.1/233	407.0977	3.88	C ₁₉ H ₁₉ O ₁₀	-1.646	317.0669 (18), 287.0563 (100)	n.a.
P8	557/235	557.0242	3.92	C ₂₁ H ₁₇ O ₁₆ S	-0.105	477.0681 (100), 380.9927 (72), 301.0359 (19)	quercetin-3-O-glucuronide-sulphate
P9	289.1/244	289.0715	4.07	C ₁₅ H ₁₃ O ₆	-1.008	245.0821 (100), 205.0508 (35), 179.0352 (17)	catechin (14)
P10	161/258	161.0246	4.3	C ₉ H ₅ O ₃	1.134		umbelliferone ^a
P11	421.1/259	421.0781	4.32	C ₁₉ H ₁₇ O ₁₁	1.129	403.0675 (26), 331.0462 (89%), 301.0356 (100), 259.0250 (5)	mangiferin (46)
P12	407.1/263	407.0986	4.38	C ₁₉ H ₁₉ O ₁₀	0.663	245.0458 (100), 201.0560 (22), 159.0455 (8)	n.a.
P13	223.1/265	223.0976	4.42	C ₁₂ H ₁₅ O ₄	-0.055	223.0978 (149), 208.0743 (28), 166.0275 (17), 151.1132 (40), 139.0404 (48), 125.0611 (100)	n.a.
P14	445.2/267	445.2082	4.45	C ₂₁ H ₃₃ O ₁₀	0.718	385.1870 (100), 283.1191 (5)	n.a.
P15	497.1/273	497.1085	4.55	C ₂₅ H ₂₁ O ₁₁	-0.874	375.0725 (100), 365.0672 (9), 261.0408 (45), 243.0302 (8)	n.a.
P16	477.1/279	477.0672	4.65	C ₂₁ H ₁₇ O ₁₃	-0.616	301.0357 (100)	miquelianin ^a (50)
P17	447.1/280	447.0935	4.67	C ₂₁ H ₁₉ O ₁₁	0.415	369.0621 (8), 357.0620 (42), 327.0514 (100) C ₁₇ H ₁₁ O ₇	n.a.
P18	489.2/285	489.1035	4.75	C ₂₃ H ₂₁ O ₁₂	-0.632	429.0833 (100), 393.0623 (18), 369.0622 (36), 357.0622 (16), 339.0516 (28), 327.0516 (39)	n.a.
P19	391.1/287	391.1034	4.78	C ₁₉ H ₁₉ O ₉	-0.09	331.0828 (12), 283.0616 (18), 271.0616 (16), 253.0509 (100), 229.0509 (95)	n.a.
P20	575.1/288	575.1044	4.8	C ₂₆ H ₂₃ O ₁₅	0.342	531.1149 (100), 471.0937 (13)	n.a.
P21	505.1/289	505.0984	4.82	C ₂₃ H ₂₁ O ₁₃	-0.641	463.0891 (18), 445.0785 (25), 301.0354 (62), 300.0282 (100)	acetyl quercetin-3-O-hexoside I
P22	431.1/291	431	4.85	C ₂₁ H ₁₉ O ₁₀	0.046	341.0672 (10), 311.0565 (100)	n.a.
P23	449.1/291	449.1093	4.85	C ₂₁ H ₂₁ O ₁₁	0.791	287.0564 (100), 151.0041 (4)	n.a.
P24	549.1/294	549.1243	4.9	C ₂₅ H ₂₅ O ₁₄	-1.163	507.1150 (69), 489.1045 (64), 447.0939 (29), 303.0514 (17), 261.0408 (100), 243.0302 (30)	n.a.
P25	391.1/297	391.0674	4.95	C ₁₈ H ₁₅ O ₁₀	0.819	373.0568 (6), 331.0461 (28), 301.0354 (100)	n.a.
P26	449.1/299	449.1093	4.98	C ₂₁ H ₂₁ O ₁₁	0.814	323.0775 (13), 303.0512 (100), 285.0407 (87), 151.0041 (25)	astilbin (19)
P27	463.1/300	463.0878	5.00	C ₂₁ H ₁₉ O ₁₂	-0.905	343.0462 (4), 301.0354 (100), 300.0285 (48)	hyperoside ^a (7) isoquercetin ^a (9)
P28	609.1/302	609.1458	5.03	C ₂₇ H ₂₉ O ₁₆	-0.538	343.0465 (8), 301.0355 (100), 271.0253 (8)	rutin ^a (6)
P29	449.2/306	449	5.10	C ₂₁ H ₂₁ O ₁₁	-0.255	275.0563 (100), 257.0457 (8)	n.a.
P30	433.1/310	433.0773	5.17	C ₂₀ H ₁₇ O ₁₁	-0.726	301.0356 (100)	avicularin (51)
P31	527.1/312	527.119	5.20	C ₂₆ H ₂₃ O ₁₂	-1.004	467.0990 (12), 405.0833 (37), 303.0514 (10), 285.0409 (74), 267.0302 (17), 261.0408 (100), 243.0303 (13)	n.a.

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P32	539.2/312	539.1196	5.20	C ₂₇ H ₂₃ O ₁₂	0.242	497.1097 (5), 417.0833 (100), 375.0727 (12), 365.0673 (6), 261.0409 (27), 243.0303 (11)	n.a.
P33	591.1/316	591.1361	5.27	C ₂₇ H ₂₇ O ₁₅	0.975	549.1258 (100), 531.1154 (11), 507.1152 (30), 489.1047 (16), 261.0409 (13)	n.a.
P34	447.1/319	447.0926	5.32	C ₂₁ H ₁₉ O ₁₁	-1.531	301.0355 (100), 300.0282 (35)	quercitrin (8)
P35	337/328	337.0021	5.47	C ₁₄ H ₉ O ₈ S	-0.923	337.0025 (10), 321.9789 (100), 257.0456 (71)	1,3-dihydroxy-5-methoxyxanthone-4-sulfonic acid (47)
P36	505.1/332	505.0981	5.53	C ₂₃ H ₂₁ O ₁₃	-1.374	463.0888 (28), 445.0783 (10), 301.0350 (60), 300.0280 (100), 299.0205 (5)	acetyl quercetin-3-O-hexoside II
P37	263.1/337	263.1287	5.62	C ₁₅ H ₁₉ O ₄	-0.541	194.0586 (100), 166.0639 (18), 151.0040 (6)	dimethylallyl-phlorisobutyrophenone ^a (15)
P38	297.1/339	297.1134	5.65	C ₁₈ H ₁₇ O ₄	0.463	228.0429 (100), 200.0482 (25), 145.0298 (4)	n.a.
P39	393.2/346	393.1914	5.77	C ₂₁ H ₂₉ O ₇	-1.085	363.1819 (18), 323.1504 (100), 293.1399 (15), 275.1294 (11), 248.0694 (13), 217.0874 (32)	n.a.
P40	277.1/358	277.1443	5.97	C ₁₆ H ₂₁ O ₄	-0.839	208.0742 (100), 180.0795 (32), 151.0040 (8)	n.a.
P41	611.1/359	611.1406	5.98	C ₃₀ H ₂₇ O ₁₄	-0.112	569.1310 (28), 551.1205 (47), 489.1047 (100), 447.0941 (76), 303.0515 (12), 261.0409 (87), 243.0303 (31)	n.a.
P42	333.2/360	333.1707	6	C ₁₉ H ₂₅ O ₅	-0.112	264.1007 (100), 247.1341 (6), 236.1057 (10), 221.0459 (10), 193.0510 (18)	n.a.
P43	351.2/386	351.1606	6.43	C ₂₂ H ₂₅ O ₄	1.331	282.0900 (100), 254.0952 (10), 214.0274 (32), 186.0325 (8)	n.a.
P44	317.2/391	317.1759	6.52	C ₁₉ H ₂₅ O ₄	0.055	248.1056 (100), 220.1108 (10), 180.0431 (25), 152.0483 (10)	n.a.
P45	361.2/398	361.2028	6.63	C ₂₁ H ₂₉ O ₅	2.029	343.1920 (26), 319.1919 (35), 317.2127 (41), 292.1321 (35), 265.1449 (100), 249.1499 (43), 233.0823 (50)	n.a.
P46	307.2/403	307.1553	6.72	C ₁₇ H ₂₃ O ₅	0.726	289.1447 (32), 238.0847 (100), 197.0821 (12)	n.a.
P47	377.2/403	377.1973	6.72	C ₂₁ H ₂₉ O ₆	0.79	359.1871 (7), 302.1528 (18), 301.1451 (14), 233.0823 (100)	n.a.
P48	283.2/407	283.0613	6.78	C ₁₆ H ₁₁ O ₅	0.435	268.0380 (100)	n.a.
P49		255.1082	6.84	C ₁₆ H ₁₅ O ₃	0.597	211.1130 (100)	n.a.
P50	463.2/413	463.2487	6.88	C ₂₉ H ₃₅ O ₅	-0.534	394.1791 (13), 379.1557 (15), 310.1213 (23), 309.1138 (10), 267.0667 (100), 255.0667 (17), 242.0587 (51), 241.0511 (13)	n.a.
P51	305.2/414	305.1758	6.9	C ₁₈ H ₂₅ O ₄	-0.074	305.1762 (6), 261.1862 (100), 221.1185 (8), 193.1237 (6)	n.a.
P52	375.2/419	375.2183	6.98	C ₂₂ H ₃₁ O ₅	1.686	357.2078 (24), 331.2284 (88), 305.1764 (25), 233.1551 (20), 211.0979 (100), 167.1082 (7)	n.a.
P53	379.2/419	379.1914	6.98	C ₂₄ H ₂₇ O ₄	-0.297	310.1215 (53), 267.0667 (100), 255.0668 (10), 242.0589 (24), 241.0511 (2)	n.a.
P54	331.2/422	331.1916	7.03	C ₂₀ H ₂₇ O ₄	0.264	331.1920 (6), 194.0587 (100), 166.0639 (42), 151.0041 (12)	4-O-geranyl-methylpropanoyl-phloroglucinol (43)
P55	435.2/427	435.2179	7.12	C ₂₇ H ₃₁ O ₅	0.397	298.0852 (100), 283.0618 (85), 281.0827 (11), 270.0904 (23), 255.0668 (11), 228.0428 (9), 227.0355 (39), 215.0355 (10)	n.a.
P56	499.3/428	499.3068	7.13	C ₃₀ H ₄₃ O ₆	0.576	429.2652 (30), 319.1918 (100)	n.a.
P57	361.2/430	361.2017	7.17	C ₂₁ H ₂₉ O ₅	-1.044	292.1322 (11), 249.1499 (100), 233.0823 (31), 181.0874 (89)	n.a.
P58	345.2/436	345.2063	7.27	C ₂₁ H ₂₉ O ₄	-2.296	276.1374 (61), 233.0824 (100), 221.0825 (15), 208.0746 (28)	n.a.
P59	433.2/439	433.2375	7.32	C ₂₈ H ₃₃ O ₄	-2.176	433.2375 (6), 365.1729 (6), 364.1690 (45), 321.1140 (100),	n.a.

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P60	419.2/442	419.2233	7.37	C ₂₇ H ₃₁ O ₄	1.186	309.1141 (12), 296.1061 (16), 295.0985 (8), 253.0514 (12), 350.1530 (16), 282.0900 (100), 281.0825 (94), 254.0953 (17), 227.0354 (32), 214.0275 (64), 186.0327 (34)	n.a.
P61	461.3/445	461.2904	7.42	C ₂₇ H ₄₁ O ₆	-1.045	392.2213 (26), 333.1716 (11), 235.0981 (100)	n.a.
P62	385.2/446	385.2387	7.43	C ₂₄ H ₃₃ O ₄	0.849	316.1685 (18), 248.1056 (92), 247.0980 (100), 220.1108 (10), 193.0510 (27), 180.0432 (43), 152.0483 (21)	n.a.
P63	359.2/448	359.2224	7.47	C ₂₂ H ₃₁ O ₄	-1.01	359.2237 (10), 315.2338 (11), 290.1530 (42), 247.0980 (100), 235.0981 (14), 222.0902 (25)	n.a.
P64	347.2/449	347.1862	7.48	C ₂₀ H ₂₇ O ₅	-0.568	347.1868 (7), 329.1762 (7), 303.1968 (99), 277.1446 (21), 195.0663 (45), 194.0587 (100), 152.0119 (20), 151.0768 (8)	n.a.
P65	447.3/451	447.2529	7.52	C ₂₉ H ₃₅ O ₄	-2.533	447.2529 (12), 403.2656 (8), 378.1848 (8), 310.1210 (34), 309.1141 (100), 267.0670 (70), 255.0670 (18), 242.0590 (43), 241.0515 (12)	n.a.
P66	443.3/454	443.2804	7.57	C ₂₇ H ₃₉ O ₅	0.141	399.2914 (25), 375.2147 (9), 374.2107 (100), 317.1764 (11), 303 (6), 248.1058 (10), 247.0981 (9), 236.1045 (9), 235.0981 (80), 222.0902 (17)	n.a.
P67	521.1/455	521.0884	7.58	C ₃₀ H ₁₇ O ₉	1.218	521.0890 (100), 477.0993 (16)	protopseudohypericin (18)
P68	375.2/456	375.2174	7.6	C ₂₂ H ₃₁ O ₅	-0.712	331.2285 (24), 313.2177 (7), 306.1477 (100), 291.1245 (25), 279.1607 (8), 250.0850 (47), 248.1048 (8), 247.0981 (74), 233.0824 (6), 191.0353 (2)	n.a.
P69	429.3/457	429.2645	7.62	C ₂₆ H ₃₇ O ₅	-0.414	222.0895 (27), 221.0825 (100)	n.a.
P70	311.1/459	311.1289	7.65	C ₁₉ H ₁₉ O ₄	0.089	267.1394 (55), 242.0587 (100)	n.a.
P71	483.3/461	483.3109	7.68	C ₃₀ H ₄₃ O ₅	-1.464	413.2708 (45), 303.1972 (100), 259.2074 (24), 193.1239 (7)	n.a.
P72	501.3/461	501.3006	7.68	C ₃₃ H ₄₁ O ₄	-0.944	432.2313 (100)	n.a.
P73	373.2/462	373.2388	7.7	C ₂₃ H ₃₃ O ₄	0.957	373.2391 (839), 329.2491 (100), 304.1685 (28), 289.1451 (52), 249.1137 (29), 221.0823 (12)	n.a.
P74	447.3/463	447.2529	7.72	C ₂₉ H ₃₅ O ₄	-2.533	447.2529 (14), 378.1848 (7), 310.1210 (12), 309.1141 (100), 267.0670 (22), 255.0670 (7), 242.0590 (11)	n.a.
P75	519.1/464	519.0772	7.73	C ₃₀ H ₁₅ O ₉	0.125	519.0733 (100), 503.0422 (8), 487.0470 (16), 475.0471 (6)	pseudohypericin (4)
P76	399.3/466	399.2536	7.77	C ₂₅ H ₃₅ O ₄	-1.184	399.2549 (100), 355.2650 (30), 329.2128 (83), 261.229 (39), 261.1139 (10), 247.0981 (41), 207.0667 (18)	n.a.
P77	497.3/466	497.3269	7.77	C ₃₁ H ₄₅ O ₅	-0.619	413.2707 (38), 303.1972 (100), 259.2074 (23)	n.a.
P78	467.3/469	467.3169	7.82	C ₃₀ H ₄₃ O ₄	0.571	398.2469 (100), 383.2235 (5)	hyperifirin (40)
P79	487.3/469	487.2852	7.82	C ₃₂ H ₃₉ O ₄	-0.293	349.1452 (100), 296.1060 (13), 282.0903 (10), 227.0354 (67)	n.a.
P80	443.3/470	443.2794	7.83	C ₂₇ H ₃₉ O ₅	-2.047	236.1051 (26), 235.0983 (100), 145.1772 (6)	chinsenin isomer (49)
P81	501.3/471	501.3007	7.85	C ₃₃ H ₄₁ O ₄	-0.585	501.3028 (17), 389.1770 (10), 363.1612 (100), 312.1140 (65), 309.1141 (14), 296.1060 (14), 253.0513 (22)	benzoylphloroglucinol
P82	291.2/474	291.1601	7.9	C ₁₇ H ₂₃ O ₄	-0.318	247.1707 (10), 223.0975 (6), 222.0899 (100), 166.0276 (2)	n.a.
P83	481.3/481	481.3319	7.91	C ₃₁ H ₄₅ O ₄	-0.879	412.2628	hyperpolyphyllirin (39)

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P84	413.3/476	413.2698	7.93	C ₂₆ H ₃₇ O ₄	0.066	369.2807 (8), 344.1999 (46), 301.1450 (100), 289.1451 (11), 276.1373 (15), 233.0824 (10)	deprenylhyperpolyphyllirin (38)
P85	327.2/480	327.0877	8	C ₁₈ H ₁₅ O ₆	0.882	327.0877 (53), 312.0642 (21), 283.0251 (23), 272.0329 (100)	n.a.
P86	361.2/480	361.2383	8	C ₂₂ H ₃₃ O ₄	-0.423	361.2385 (10), 303.1602 (10), 301.2174 (100), 289.1447 (10)	n.a.
P87	427.3/482	427.2859	8.03	C ₂₇ H ₃₉ O ₄	1.21	383.2964 (17), 290.1528 (37), 289.1452 (16), 247.0980 (17), 222.0902 (31)	n.a.
P88	365.2/483	365.1757	8.05	C ₂₃ H ₂₅ O ₄	-0.418	321.1866 (99), 229.0502 (28), 228.0431 (100), 227.0357 (37), 200.0483 (7)	n.a.
P89	495.3/483	495.3488	8.05	C ₃₂ H ₄₇ O ₄	1.75	-	adhyperpolyphyllirin
P90	601.4/484	601.3532	8.07	C ₃₈ H ₄₉ O ₆	-0.303	479.3176 (100)	n.a.
P91	415.3/485	415.2485	8.08	C ₂₅ H ₃₅ O ₅	-1.15	371.2597 (12), 345.2076 (25), 263.1286 (68), 262.1215 (93), 219.0666 (100), 207.0666 (24), 194.0588 (47)	n.a.
P92	515.3/485	515.3168	8.08	C ₃₄ H ₄₃ O ₄	0.149	437.2709 (6), 267.0669 (100), 255.0590 (11), 242.0590 (14), 241.0513 (7)	n.a.
P93	331.2/488	331.1913	8.13	C ₂₀ H ₂₇ O ₄	-0.491	331.1919 (35), 313.1813 (10), 287.2020 (100), 262.1214 (10), 261.1500 (16), 207.0666 (23), 194.0587 (8)	3-geranyl-methylpropanoyl-phloroglucinol ^a (17)
P94	443.3/489	443.2801	8.15	C ₂₇ H ₃₉ O ₅	-0.513	399.2914 (8), 374.2107 (35), 356.2001 (16), 317.1765 (43)	chinesin isomer
P95	467.3/491	467.3161	8.18	C ₃₀ H ₄₃ O ₄	-1.162	423.3276 (13), 355.1920 (10), 329.1763 (100), 287.1293 (65), 275.1294 (12), 262.1215 (13), 219.0666 (16)	n.a.
P96	569.4/492	569.3633	8.2	C ₃₈ H ₄₉ O ₄	-0.656	431.2244 (37), 309.1309 (14), 309.1143 (100), 309.0949 (11)	benzoylphloroglucinol
P97	345.2/494	345.2074	8.23	C ₂₁ H ₂₉ O ₄	0.832	327.1972 (13), 301.2178 (100), 261.1501 (17), 221.0823 (27)	3-geranyl-methylbutanoyl-phloroglucinol ^a
P98	535.4/496	535, 3783	8.27	C ₃₅ H ₅₁ O ₄	-1.799	467.3137 (31), 466.3106 (100), 397.2397 (36), 395.2608 (13), 384.2292 (14), 383.2242 (68), 315.1612 (54), 313.1819 (38)	hyperforin (1)
P99	549.4/498	549.3946	8.3	C ₃₆ H ₅₃ O ₄	-0.589	431.3289 (32), 480.3257 (100), 411.2550 (33), 398.2441 (15), 397.2393 (65), 395.2608 (13), 329.1765 (851), 313.1816 (42)	adhyperforin (2)
P100	505.1/499	505.0939	8.32	C ₃₀ H ₁₇ O ₈	2.018	505.0939 (100), 461.1042 (12)	protohypericin (18)
P101	331.2/500	331.1908	8.33	C ₂₀ H ₂₇ O ₄	-2.121	331.1921 (8), 287.2022 (64), 195.0662 (15), 194.0588 (100), 152.0120 (14)	2-O-geranyl-methylpropanoyl-phloroglucinol (41)
P102	443.3/503	443.2807	8.38	C ₂₇ H ₃₉ O ₅	1.021	374.2106 (35), 356.2001 (26), 317.1765 (36), 304.1683 (10), 303.1609 (10), 249.1137 (45), 247.0981 (13), 236.1050 (30), 235.0980 (100), 222.0902 (30)	chinesin isomer
P103	345.2/506	345.2068	8.43	C ₂₁ H ₂₉ O ₄	-0.877	345.2078 (13), 301.2179 (59), 261.1502 (5), 209.0875 (17), 208.0745 (100), 152.0120 (21)	2-O-geranyl-1-methylbutanoyl-phloroglucinol (42)
P104	415.2/506	415	8.43	C ₂₅ H ₃₅ O ₅	-1.391	357.2078 (32), 279.1242 (75), 278.1165 (12), 263.1293 (10), 219.0666 (37), 208.0743 (100), 207.0668 (30)	n.a.
P105	433.2/507	433.238	8.45	C ₂₈ H ₃₃ O ₄	-1.068	433.2394 (100), 389.2494 (54), 355 (1922 (61), 321.1138 (40), 309.1138 (38), 295.0982 (26), 287.2023 (27), 253.0512 (21), 145.0300 (19)	n.a.
P106	637.4/508	637.4261	8.47	C ₄₃ H ₅₇ O ₄	-0.147	500.2950 (100), 485.2718 (12), 431.2241 (71)	n.a.
P107	279.2/511	279.2328	8.52	C ₁₈ H ₃₁ O ₂	-0.55		Fatty acid
P108	503.1/512	503.0778	8.53	C ₃₀ H ₁₅ O ₈	1.032	503.0785 (100), 461.0677 (5), 459.0886 (12)	hypericin (3)

No	Feature	[M-H] ⁻	Rt (min)	Molecular Formula [M-H] ⁻	Δ ppm	Fragments (Intensity)	Identification
P109	443.3/513	443.2803	8.55	C ₂₇ H ₃₉ O ₅	0.118	375.2141 (18), 374.2104 (100) 359.2239 (16), 359.1870 (32) 331.1564 (20), 2470980 (57), 236.1046 (25), 235.0977 (94.58) 234.0901 (62), 233.0825 (32)	chinesin isomer
P110	603.4/513	603.4411	8.55	C ₄₀ H ₅₉ O ₄	-1.232	466.3103 (100), 451.2869 (12), 397.2395 (83)	n.a.
P111	399.3/528	399.2542	8.8	C ₂₅ H ₃₅ O ₄	0.268	399.2551 (9), 355.2653 (21), 263.1279 (28), 262.1217 (100), 261.1141 (26), 219.0668 (84), 207.0668 (20), 194.0590 (42)	n.a.
P112	413.3/531	413.2692	8.85	C ₂₆ H ₃₇ O ₄	-1.217	413.2707 (6), 369.2808 (22), 277.1434 (29), 276.1372 (100), 275.1296 (28), 233.0824 (89), 221.0824 (21), 208.0745 (43)	n.a.
P113	497.3/541	497.3276	9.02	C ₃₁ H ₄₅ O ₅	0.688	469.3330 (18), 453.3381 (25), 401.2703 (100)	n.a.
P114	567.4/543	567.3694	9.05	C ₃₅ H ₅₁ O ₆	0.577	539.3746 (7), 497.3271 (7), 471.3118 (100), 453.3008 (7), 415.3221 (6)	deoxy-hydroperoxy-furohyperforin
P115	581.4/549	581.3849	9.15	C ₃₆ H ₅₃ O ₆	0.254	553.3909 (8), 497.3280 (18), 486.3314 (29), 485.3281 (100), 467.3175 (17), 429.3381 (14), 383.2598 (9)	hydroperoxy fruoadhyperforin
P116	551.4/558	551.3748	9.3	C ₃₅ H ₅₁ O ₅	1.038	523.3801 (11), 507.3851 (15), 455.3173 (100), 437.3064 (5), 329.2855 (7)	n.a.

^a identified with reference compound

Appendix 6.4 Antibacterial assay results of different *Hypericum* species

species	Abbreviation	<i>B. subtilis</i>				<i>A. fischeri</i>			
		c1 (50 µg/ml)	Std. Dev. C1	c2 (500 µg/ml)	Std. Dev. C2	c1 (50 µg/ml)	Std. Dev. C1	c2 (500 µg/ml)	Std. Dev. C2
<i>H. androsaemum</i>	KEW_44_a	14%	17%	99%	3%	-13%	8%	-71%	13%
	KEW_44_b	-52%	67%	-88%	23%	-33%	2%	-89%	3%
	KEW_44_c	18%	39%	-79%	36%	-32%	2%	-90%	6%
	IPK_27_a	-21%	41%	-32%	33%	-2%	4%	-58%	8%
	IPK_27_b	-27%	40%	-53%	38%	-39%	8%	-120%	7%
<i>H. attenuatum</i>	IPK_27_c	18%	21%	99%	2%	7%	6%	-56%	6%
	IPK_2_a	-8%	45%	66%	14%	1%	3%	13%	5%
	IPK_2_b	-63%	43%	-117%	27%	-1%	2%	12%	4%
<i>H. barbatum</i>	IPK_2_c	-74%	42%	-133%	33%	-38%	4%	-3%	5%
	KEW_23_a	-1%	21%	100%	6%	-29%	4%	15%	4%
	KEW_23_b	-43%	72%	68%	15%	-12%	5%	33%	1%
<i>H. calycinum</i>	KEW_23_c	-10%	37%	-128%	39%	-34%	3%	7%	3%
	IPK_4_a	20%	19%	101%	7%	3%	2%	-60%	4%
	IPK_4_b	-18%	28%	100%	7%	-9%	7%	-69%	4%
<i>H. canariense</i>	IPK_4_c	6%	14%	103%	3%	-36%	7%	-49%	9%
	KEW_4_a	96%	1%	103%	2%	3%	3%	71%	3%
	KEW_4_b	114%	20%	104%	3%	-9%	2%	67%	2%
<i>H. coris</i>	KEW_4_c	55%	2%	99%	2%	-10%	2%	56%	3%
	IPK_33_a	97%	0%	93%	7%	6%	1%	11%	6%
	IPK_33_b	-6%	29%	93%	13%	-13%	2%	-44%	8%
	IPK_33_c	95%	2%	91%	3%	-53%	7%	2%	19%
	KEW_46_a	97%	1%	95%	5%	-5%	5%	50%	3%
<i>H. curvisepalum</i>	KEW_46_b	-21%	31%	96%	3%	-40%	8%	58%	2%
	KEW_33_a	97%	0%	101%	2%	-22%	6%	-52%	7%
	KEW_33_b	96%	1%	104%	8%	-10%	4%	60%	8%
<i>H. elodes</i>	KEW_2_a	-52%	52%	99%	4%	-4%	4%	-44%	7%
	KEW_22_b	94%	1%	101%	3%	-92%	16%	100%	0%
	KEW_22_c	98%	0%	89%	10%	-28%	1%	100%	0%

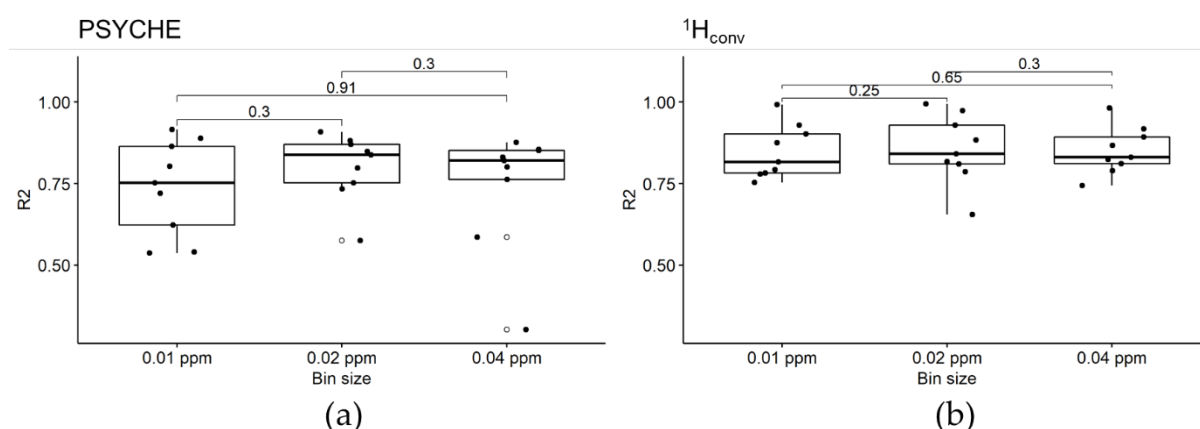
species	Abbreviation	<i>B. subtilis</i>				<i>A. fischeri</i>			
		c1 (50 µg/ml)	Std. Dev. C1	c2 (500 µg/ml)	Std. Dev. C2	c1 (50 µg/ml)	Std. Dev. C1	c2 (500 µg/ml)	Std. Dev. C2
<i>H. grandifolium</i>	IPK_3_a	-1%	16%	95%	8%	-49%	8%	-77%	4%
	IPK_3_b	-32%	48%	87%	15%	1%	2%	-70%	7%
	IPK_3_c	-28%	16%	98%	11%	-3%	9%	-79%	5%
<i>H. hirsutum</i>	IPK_34_a	9%	24%	91%	6%	-11%	3%	-78%	3%
	IPK_34_b	-9%	20%	99%	4%	-2%	9%	-23%	9%
	IPK_34_c	-23%	31%	99%	2%	-13%	8%	-65%	8%
<i>H. hookerianum</i>	KEW_33_c	24%	30%	97%	3%	-33%	5%	-108%	9%
	KEW_2_b	15%	10%	99%	10%	-13%	8%	-88%	8%
<i>H. humifusum</i>	KEW_19_a	95%	0%	103%	2%	-4%	2%	-20%	10%
	KEW_19_b	93%	1%	99%	2%	-38%	4%	14%	10%
	KEW_19_c	27%	18%	94%	3%	-10%	3%	-5%	8%
<i>H. inodorum</i>	IPK_5_a	-32%	50%	-43%	27%	-12%	3%	-59%	3%
	IPK_5_b	1%	19%	-115%	36%	-3%	6%	-65%	3%
<i>H. lagarocladum</i>	KEW_37_a	95%	1%	105%	5%	-2%	7%	-68%	12%
	KEW_37_b	-20%	35%	100%	29%	-51%	7%	-95%	7%
	KEW_37_c	30%	26%	91%	10%	-13%	2%	-47%	2%
<i>H. maculatum</i>	IPK_28_a	-54%	70%	7%	28%	3%	3%	20%	8%
	IPK_28_b	-55%	50%	-88%	36%	4%	7%	-8%	7%
	IPK_28_c	-45%	17%	-130%	34%	2%	6%	-5%	6%
<i>H. orientale</i>	IPK_35_a	-3%	35%	103%	4%	10%	5%	81%	6%
	IPK_35_b	17%	12%	82%	13%	-7%	2%	63%	2%
	IPK35_c	9%	21%	85%	13%	-36%	6%	80%	5%
<i>H. perforatum</i>	IPK24_a	22%	11%	62%	37%	-20%	4%	-3%	3%
	KEW9_a	9%	12%	-68%	45%	-42%	4%	-9%	2%
	KEW9_b	11%	15%	-76%	25%	-21%	5%	-10%	4%
	KEW9_c	-9%	33%	-112%	47%	0%	12%	8%	2%
	KEW10_a	27%	15%	96%	7%	-14%	6%	5%	4%
	KEW10_b	-22%	45%	-154%	29%	-13%	10%	8%	7%
	KEW12_a	62%	7%	103%	6%	-4%	2%	21%	11%
	KEW26_a	82%	5%	104%	9%	1%	6%	20%	2%
	KEW12_b	5%	10%	10%	29%	-5%	8%	13%	13%
	KEW26_b	11%	29%	27%	37%	-16%	3%	8%	3%
	KEW12_c	0%	19%	-34%	19%	-14%	1%	16%	5%
KEW26_c	-65%	56%	-136%	37%	-25%	3%	3%	6%	
<i>H. polyphyllum</i>	IPK10_a	53%	22%	106%	7%	2%	2%	33%	7%
	IPK10_b	18%	13%	-15%	22%	1%	8%	5%	2%
	IPK10_c	-32%	45%	62%	18%	-3%	5%	29%	5%
<i>H. pulchrum</i>	KEW27_a	95%	1%	107%	3%	-6%	4%	-10%	3%
	KEW28_a	13%	9%	101%	3%	-5%	1%	-46%	6%
	KEW27_b	-82%	41%	86%	8%	-24%	4%	-20%	4%
	KEW28_b	8%	35%	98%	4%	-16%	1%	-53%	4%
	KEW27_c	-7%	48%	97%	2%	-26%	3%	-39%	3%
	KEW28_c	-23%	45%	96%	3%	-21%	3%	-55%	5%
<i>H. reflexum</i>	KEW5_a	101%	4%	101%	3%	-3%	4%	59%	1%
<i>H. tetrapterum</i>	KEW45_a	61%	4%	94%	5%	1%	5%	21%	2%
	KEW15_b	20%	14%	99%	7%	-28%	8%	-41%	2%

Appendix 7.1 Quantitative correlation of bins of different signals

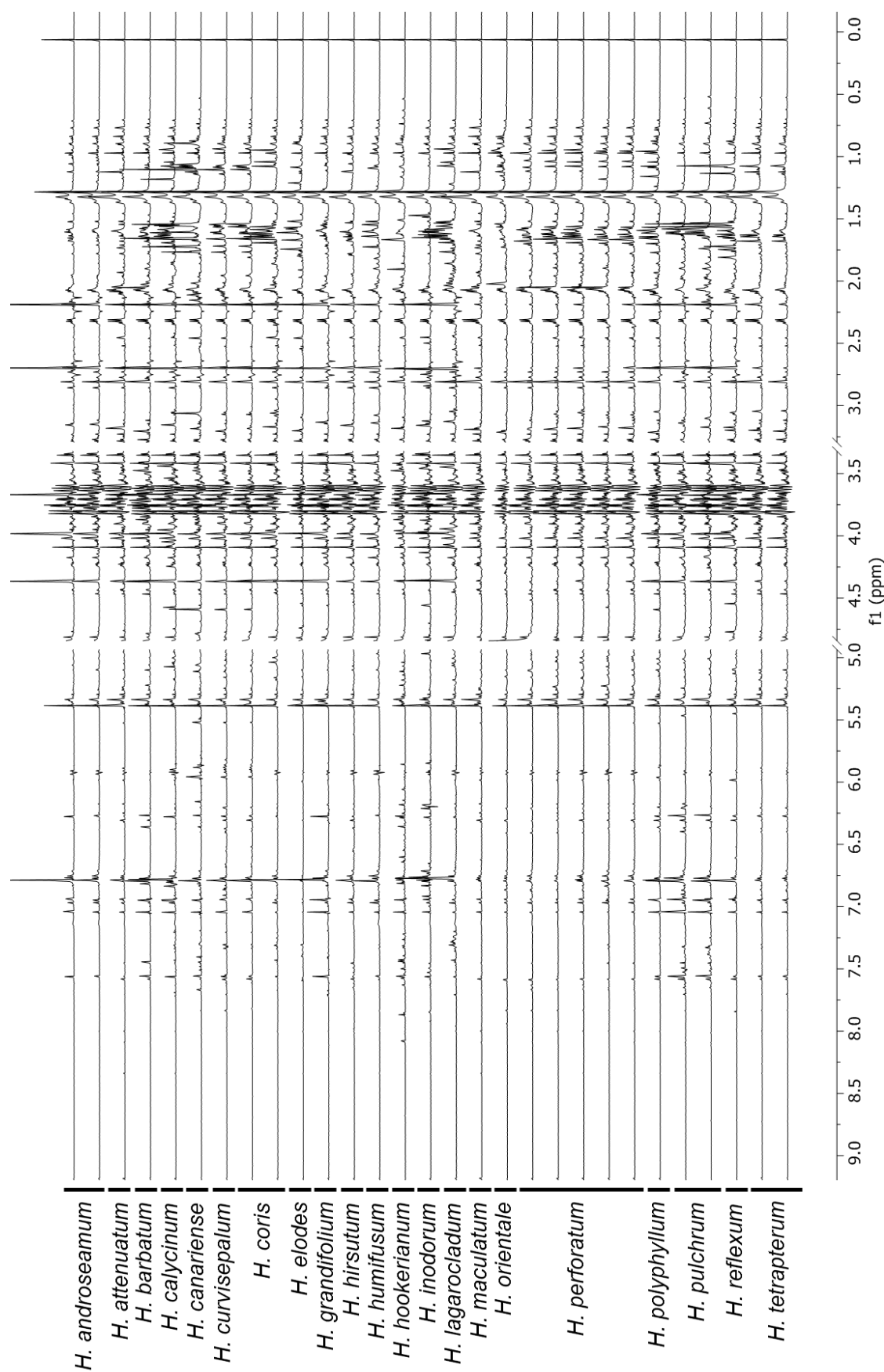
Summary of the coefficient of determination (R^2) of linear regression between integral (^1H qNMR) and bin values of experiments with different bin sizes.

Compound	Assignment	δ [ppm] multiplicity (J [Hz])	Bin size [ppm]	R^2 of experiment	
				$^1\text{H}_{\text{conv}}$	PSYCHE
Chlorogenic acid (10)	H-8'	6.31 d (15.8 Hz)	0.01	0.7822	0.7200
			0.02	0.6553	0.8485
			0.04	0.8667	0.801
Chlorogenic acid (10)	H-2'	7.05 d (2.1 Hz)	0.01	0.7537	0.5365
			0.02	0.7865	0.5748
			0.04	0.744	0.5861
Rutin (6)	H-6'''	1.12 d (6.2 Hz)	0.01	0.8166	0.6229
			0.02	0.8176	0.7331
			0.04	0.8311	0.3024
Hyperoside (7)	H-2'	7.83 d (2.2 Hz)	0.01	0.902	0.8884
			0.02	0.9288	0.881
			0.04	0.9178	0.8518
Epicatechin/Catechin (14)	H-6	5.94 d (2.4 Hz)	0.01	0.9925	0.9156
			0.02	0.9740	0.7974
			0.04	0.9824	0.8204
Epicatechin/Catechin (14)	H-2'	6.97 d (1.9 Hz)	0.01	0.7791	0.7526
			0.02	0.8104	0.7525
			0.04	0.7896	0.7627
Hyperforin (1)	H ₃ -12	1.08 d (6.5 Hz)	0.01	0.8753	0.8635
			0.02	0.8842	0.8380
			0.04	0.8237	0.8305
Sucrose (12)	H-3'	4.09 d (8.2 Hz)	0.01	0.9294	0.5399
			0.02	0.9945	0.9080
			0.04	0.8928	0.8542
Shikimic acid (11)	H-4	4.36 m ($\nu_{1/2}$ 4.7 Hz)	0.01	0.7929	0.8025
			0.02	0.8412	0.8706
			0.04	0.8115	0.8762

Appendix 7.2 Boxplots of correlation coefficients for different experiments with varied bin sizes



Boxplot of R^2 with varied bin sizes of (a) PSYCHE and (b) conventional ^1H -NMR ($^1\text{H}_{\text{conv}}$). Significance Test: Wilcoxon-Mann-Whitney paired (shown is the p-value).

Appendix 7.3: Stacked PSYCHE spectra of *Hypericum* species

Declaration on the author contributions

Chapter 2: Characterization of pharmaceutical extracts from 20 *H. perforatum* genotypes

Pauline Stark, Katrin Franke, Sarah Scharfenberg, Andrea Porzel, Jacob Duerichen-Parfitt, Helle Wangenstein, Timothy F. Sharbel and Ludger A. Wessjohann (*in finalization for publication*)

In this work, PS developed the methods, planned and performed the chemical analysis of the plant extracts, interpreted the results and wrote the manuscript. SS contributed to the chemometric analysis of the data. AP supervised the NMR measurements. JDP isolated reference compounds. HW performed AChE and DPPH-radical scavenger assay. KF (project leader) coordinated the work and revised the manuscript. LAW and TFS designed and supervised the project, acquired the funding and revised the manuscript.

Chapter 3: Discovery of key regulators of dark glands development and hypericin biosynthesis in St. John's wort (*Hypericum perforatum*)

Paride Rizzo, Lothar Altschmied, Pauline Stark, Twan Rutten, André Gündel, Sarah Scharfenberg, Katrin Franke, Helmut Bäumlein, Ludger Wessjohann, Marcus Koch, Ljudmilla Borisjuk and Timothy F. Sharbel.

Plant Biotechnology Journal (2019) 17, pp. 2299–2312, doi: 10.1111/pbi.13141.

PR conceived the idea of this study, discovered the G+/- PT phenotype combination, performed the characterization work, extracted RNA, interpreted transcriptomics data set and wrote the manuscript. LA analysed RNAseq data, identified candidate genes, assembled the novel hypericin biosynthesis pathway and contributed to the writing. PS performed metabolomics analysis and interpreted metabolomics results. TR performed confocal microscopy. AG established FTIR microscopy imaging methods. SS performed bioinformatic work. KF and LW were involved in project administration and funding acquisition, coordinated scientific work and contributed to writing and correction of the manuscript. HB contributed to the review and editing of manuscript. MK provided germplasm and improved the manuscript. LB supervised scientific work at IPK. TS supervised scientific work at IPK, raised funding and contributed with corrections.

Chapter 4: Intraspecific variance of *H. perforatum* genotypes – a metabolomics approach

Pauline Stark, Katrin Franke, Sarah Scharfenberg, Paride Rizzo, Ludger A. Wessjohann

In this work, PS carried out the sampling, the chemical analysis of the plant material, the chemometric analysis and the interpretation and visualization of the results. SS contributed to chemometric data analysis pipeline. KF (project leader) raised the funding, coordinated the work

and revised the chapter. PR curated the plant selection and provided genetic metadata. LAW supervised the project and was involved in project administration.

Chapter 5: Chemical comparison of two leaf phenotypes

Pauline Stark, Nina Pöttsch, Sarah Scharfenberg, Paride Rizzo, Katrin Franke

In this work PS carried out the sampling, the chemical analysis, structure isolation and interpretation. Leaf phenotyping and counting tool development was performed by Nina Pöttsch supervised by Sarah Scharfenberg. PR provided plant material and performed ploidy analysis. KF contributed to project administration, review and editing.

Chapter 6: Phytochemical analysis of different *Hypericum* species

Pauline Stark, Caroline Zab, Andrea Porzel, Katrin Franke, Sarah Scharfenberg, Paride Rizzo, Ludger A. Wessjohann

PS planned the experiments and performed the sampling with CZA. PS carried out LC-MS experiments, chemometric data analysis, data curation, interpretation and visualization of the results, and wrote the manuscript. CZA carried out the NMR metabolomics experiments and TLC, supervised by PS and AP. SS performed the activity correlation analysis. PR curated the plants and selected the germplasm collection. KF, PR and LAW designed and wrote the funding projects and did project administration. Review and editing were done by AP, KF, and LAW.

Chapter 7: PSYCHE – a valuable experiment in plant NMR-metabolomics

Pauline Stark, Caroline Zab, Andrea Porzel, Katrin Franke, Paride Rizzo, Ludger A. Wessjohann

Molecules (2020) 25, 1-13e, doi: 10.3390/molecules25215125

PS planned the experiments and performed the sampling with CZ. CZ carried out the NMR metabolomics experiments and TLC, supervised by PS and AP. AP and PS developed project idea. PR curated the plants and selected the germplasm collection. KF, PR and LAW designed and wrote the funding projects and did project administration. Review and editing were done by AP, KF, and LAW.

Curriculum vitae

Personal details

Name Pauline Stark
Date and place of Birth 16th January 1992 in Erfurt
Nationality German

Formal education

2002 – 2010 Königin-Luise-Gymnasium, Erfurt
Degree: General qualification for University entrance

2010 – 2015 Study of food chemistry, Dresden University of Technology, Dresden
Graduation as Diplom food chemist
Title: “Valuable compounds of medically relevant thyme species”
Supervisor: Prof. Dr. Karl Speer, Dr. Gesche Wittpahl

2015 – present PhD candidate (Dr. rer. nat.), Leibniz Institute of Plant Biochemistry, Department of Bioorganic chemistry, Halle (Saale)
Title: “Metabolite Profiling and Chemometric Analysis of *Hypericum*”
Supervisor: Prof. Dr. Ludger Wessjohann
Mentor: Dr. Katrin Franke

Work experience

10/2015 02/2021 Research Assistant, Leibniz Institute of Plant Biochemistry, Department of Bioorganic Chemistry, Halle (Saale)

03/2021 present Responsible Senior Scientist for NMR spectroscopy, Leibniz Institute of Plant Biochemistry, Department of Bioorganic Chemistry, Halle (Saale)

Publications

Publications in peer-reviewed journals

Rizzo P, Altschmied L, **Stark P**, Rutten T, Gundel A, Scharfenberg S, Franke K, Bäumlein H, Wessjohann L, Koch M, Borisjuk L and Sharbel TF. Discovery of key regulators of dark gland development and hypericin biosynthesis in St. John's Wort (*Hypericum perforatum*). *Plant Biotechnol J* **2019**. 17:2299-2312, doi: 10.1111/pbi.13141.

Stark P, Zab C, Porzel A, Franke K, Rizzo P and Wessjohann LA. PSYCHE – A valuable experiment in plant NMR-metabolomics. *Molecules (Basel, Switzerland)* **2020**. 25:1-13, doi: 10.3390/molecules25215125.

Oral presentations

Stark P, Metabolite profiling and the intraspecific variance of *Hypericum perforatum*. 13th Plant Science Student Conference “Saw your future”, Halle (Saale), 6th June 2017.

Stark P, Metabolite profiling of *H. perforatum* reveals new insights into the hypericin biosynthesis pathway. 57. Doktorandenworkshop: „Naturstoffe: Chemie, Biologie und Ökologie“, 10th May 2019, Jena.

Poster presentations

Stark P, Porzel A, Farag M, Fobofou SAT, Franke K, Wessjohann LA; Metabolite Profiling and Chemometrics of St. John's Wort. In: Symposium: “Learning from Nature, learning from our ancestors, from tradition to evidence based medicines.” (14th – 15th April 2016) Leiden, Netherlands.

Stark P, Porzel A, Farag M, Fobofou SAT, Franke K, Wessjohann LA; Metabolite Profiling and chemometrics of *Hypericum perforatum*. In: The 12th Plant Science Student Conference: “Field of Innovation” (4th – 7th July 2016) Gatersleben, Germany.

Stark P, Scharfenberg S, Porzel A, Sharbel T, Franke K, Wessjohann LA; Metabolite Profiling and the intraspecific variance of *Hypericum perforatum*. In: 3rd International Conference on Natural Products Utilization: from Plant to Pharmacy Shelf (ICNPU-2017) (18th – 21th October 2017) Bansko, Bulgaria.

Stark P, Scharfenberg S, Porzel A, Sharbel T, Franke K, Wessjohann LA; Metabolite profiling and the intraspecific variance of *Hypericum perforatum* and *Hypericum maculatum*. In: 3rd European Conference on Natural Products (2nd – 5th September 2018) Frankfurt.

Stark P, Scharfenberg S, Porzel A, Sharbel T, Franke K, Wessjohann LA; Metabolite profiling and the intraspecific and interspecific variance of *Hypericum perforatum* and *Hypericum maculatum*. In: Plant Science Student Conference (19th – 22nd June 2018) Gatersleben.

Stark P, Scharfenberg S, Porzel A, Rizzo P, Franke K, Wessjohann LA; Metabolite Profiling and the intraspecific variance of *Hypericum perforatum*. In: Leibniz Wirkstofftage 2019 (1st – 2nd April 2019) Dresden.

Stark P, Rizzo P, Scharfenberg S, K. Franke K, Wessjohann LA; The intraspecific variance of *Hypericum perforatum* – a metabolomics approach. In: 32. Irseer Naturstofftage (19th – 21st February 2020) Irsee.

Fellowships/ Awards/ Funding aquisition

Best Poster award: 12th Plant Science Student Conference: “Field of Innovation” (4th – 7th July 2016) Gatersleben, Germany.

PSE (Phytochemical Society of Europe) Prize for the Best Poster Presentation: 3rd International Conference on Natural Products Utilization: from Plant to Pharmacy Shelf (ICNPU-2017) (18th – 21th October 2017) Bansko, Bulgaria.

DAAD-RISE Fellowship Supervisor (Interns: Jacob Duerichen-Parfitt from Simon Fraser University, Canada, May August 2018)

Seed Money Funding 2020 by Leibniz Research Alliance “Bioactive Compounds and Biotechnology”, 10.000 €: Stark P, Franke K, IPB (Halle) / Sikorski J, DSMZ (Braunschweig): "*Hypericum* metabolome and rhizosphere microbiome: a cross-kindom factory for bioactive compounds".

Seed Money Funding 2021 by Leibniz Research Alliance “Bioactive Compounds and Biotechnology”, 10.000 €: D’Auria JC, IPK (Gatersleben) / Stark P, IPB (Halle): "Gramine Biosynthesis – Metabolomics and Enzymology (GraB-ME)".

Eidesstattliche Erklärung

Ich erkläre, dass ich die vorliegende, unter der Betreuung von Herrn Prof. Dr. rer. nat. Ludger Wessjohann und Frau Dr. rer. nat. Katrin Franke, angefertigte wissenschaftliche Abschlussarbeit selbstständig verfasst habe.

Andere als die angegebenen Hilfsmittel wurden von mir nicht benutzt. Alle angeführten Zitate wurden kenntlich gemacht.

Halle, den



Pauline Stark