Article

Untargeted In Silico Compound Classification—A Novel Metabolomics Method to Assess the Chemodiversity in Bryophytes

Kristian Peters [1,2,3,*], Gerd Balcke [4], Niklas Kleinenkuhnen [5,6], Hendrik Treutler [1,7] and Steffen Neumann [1,2]

1   Bioinformatics & Scientific Data, Leibniz Institute of Plant Biochemistry, Weinberg 3,
     06120 Halle (Saale), Germany; hendrik.treutler@gmail.com (H.T.); sneumann@ipb-halle.de (S.N.)
2   German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e,
     04103 Leipzig, Germany
3   Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg,
     06108 Halle (Saale), Germany
4   Cell and Metabolic Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany;
     gbalcke@ipb-halle.de
5   Max Planck Research Group Chromatin and Ageing, Max Planck Institute for Biology of Ageing,
     Joseph-Stelzmann-Str. 9b, 50931 Cologne, Germany; NKleinenkuhnen@age.mpg.de
6   MS-Platform, Cluster of Excellence on Plant Sciences, Botanical Institute (CEPLAS), University of Cologne,
     50931 Cologne, Germany
7   Datameer GmbH, Magdeburger Straße 23, 06112 Halle (Saale), Germany
*   Correspondence: kpeters@ipb-halle.de; Tel.: +49-345-5582-1475

Abstract: In plant ecology, biochemical analyses of bryophytes and vascular plants are often conducted on dried herbarium specimen as species typically grow in distant and inaccessible locations. Here, we present an automated in silico compound classification framework to annotate metabolites using an untargeted data independent acquisition (DIA)–LC/MS–QToF-sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH) ecometabolomics analytical method. We perform a comparative investigation of the chemical diversity at the global level and the composition of metabolite families in ten different species of bryophytes using fresh samples collected on-site and dried specimen stored in a herbarium for half a year. Shannon and Pielou's diversity indices, hierarchical clustering analysis (HCA), sparse partial least squares discriminant analysis (sPLS-DA), distance-based redundancy analysis (dbRDA), ANOVA with post-hoc Tukey honestly significant difference (HSD) test, and the Fisher's exact test were used to determine differences in the richness and composition of metabolite families, with regard to herbarium conditions, ecological characteristics, and species. We functionally annotated metabolite families to biochemical processes related to the structural integrity of membranes and cell walls (proto-lignin, glycerophospholipids, carbohydrates), chemical defense (polyphenols, steroids), reactive oxygen species (ROS) protection (alkaloids, amino acids, flavonoids), nutrition (nitrogen- and phosphate-containing glycerophospholipids), and photosynthesis. Changes in the composition of metabolite families also explained variance related to ecological functioning like physiological adaptations of bryophytes to dry environments (proteins, peptides, flavonoids, terpenes), light availability (flavonoids, terpenes, carbohydrates), temperature (flavonoids), and biotic interactions (steroids, terpenes). The results from this study allow to construct chemical traits that can be attributed to biogeochemistry, habitat conditions, environmental changes and biotic interactions. Our classification framework accelerates the complex annotation process in metabolomics and can be used to simplify biochemical patterns. We show that compound classification is a powerful tool that allows to explore relationships in both molecular biology by "zooming in" and in ecology by "zooming out". The insights revealed by our framework allow to construct new research hypotheses and to enable detailed follow-up studies.

Keywords: compound classes; classification; chemical ecology; metabolomics; chemodiversity; biodiversity; chemical traits; mosses; bryophytes; data-independent acquisition

1. Introduction

The identification of chemical constituents underlying specific ecological and molecular functioning has become a powerful tool in plant biology [1–3]. It has enabled us to pinpoint compounds that exert specific bioactivities, play vital roles in molecular pathways, are produced as adaptations to environmental or climatic changes, or describe molecular interactions with other organisms [4]. For several decades, the technology of metabolomics has been applied, e.g., LC/MS now allows the detection of the majority of semi-polar compounds in plant samples [5]. However, it is only recently that metabolomics has been applied in ecology [1,2]. In ecological metabolomics, metabolite fingerprinting strategies are most often applied, where characteristic sets and subsets of metabolite features such as compounds, fragments, and adducts are related to ecological functioning either within plants at molecular scales or at more coarse ecological scales without specifically targeting some molecules [4]. With untargeted ecometabolomics, the emphasis is now shifting from finding metabolite fingerprints underlying a specific functioning to a more detailed investigation of metabolites at broader levels [6].

In the field of ecometabolomics, typically non-model species are studied that have important functions in ecosystems but those species are biochemically not well characterized. Despite the fact that only approximately 24,000 species of bryophytes are known today [7], they have evolved a high diversity of bioactive compounds to interact with the environment and with other organisms. As they quickly respond to environmental changes and pollution to sustain their homeostasis, bryophytes are considered distinctive bioindicators [8–10] and can be keystone components in some ecosystems [11]. For instance, bryophytes often colonize early successional stages and by collecting debris, storing water and solidifying soil they pave the way for other organisms [12]. Bryophyte cushions can act as nurseries for a variety of larval stages of insects and serve as indirect food reserves for birds and insect feeding animals [13,14]. Bryophytic species interact with the environment and other organisms by producing a large diversity of "infochemicals" that mediate various kinds of ecological functioning [15]. For example, male plants of the mosses *Bryum argenteum* and *Ceratodon purpureus* were found to emit volatile compounds that attract micro-arthropods, such as springtails and mites that may transport sperm released by the mosses antheridia to neighboring female plants, enhancing sporophyte production [16,17]. To date, the mechanistic functions, including genetic factors, of the majority of these volatile compounds and other secondary metabolites have not been explored in detail [18,19].

In plant ecology, experiments are often conducted in the field, which may involve the sampling of species in remote and inaccessible locations where it is often impossible to carry the equipment to shock-freeze plant material in order to preserve it for later biochemical analyses. As a result, plant material is collected and transported in ambient conditions, and eventually stored in herbarium collections. This is especially true for bryophytes, where the majority of specimens are collected in herbaria as species determination often requires subsequent microscopic investigation. When bryophytes are stored as herbarium specimen, they undergo physiological changes which can be described as the combined effects of drought, oxidative stress, light inhibition, and apoptosis. While some liverworts, such as the genus *Scapania* lose their oil bodies in just a few minutes or hours when drying, for the bryophytes *Tortula muralis* and *Riccia macrocarpa* it has been reported that they are able to survive prolonged periods of drought and sustain desiccation and low relative humidity for 10–20 years [20–22]. While desiccation tolerance is a widespread trait in bryophytes, only few vascular species (approximately 70 fern and 60 angiosperms) are known that can withstand similar conditions [23,24]. Physiologically, upon dehydration species like liverworts of the genus *Scapania* experience a breakdown of cellular membranes, such as those surrounding the oil bodies. By contrast, desiccation-tolerant bryophytes possess protective mechanisms and produce specific compounds in order to maintain cellular integrity during both de- and re-hydration [24,25]. For instance, some bryophytes, such as *Polytrichum* spp., produce hydroxylated cinnamic acids that form lignin-like structures ("proto-lignin") in conducting cell wall tissues that protect them from dehydration and

reinforce cellular integrity [26,27]. Proto-lignin-like compounds have not been found so far in liverworts such as *Scapania* which may explain their susceptibility to desiccation. Physiological changes under herbarium conditions also include oxidative stress. Bryophytes have been found to produce a variety of compounds like flavonoids, anthocyanins, alkaloids, terpenes, (bis)bibenzyls covering a large diversity of compound families that protect them from reactive oxygen species (ROS) [28,29]. As the biochemical adaptations are very complex, little is known about the role of particular secondary metabolites and corresponding genes so far [18,30,31].

Recently, technological advances in high resolution LC/MS–MS-based metabolomics enabled the quantitative assessment of signals obtained from MS1 and simultaneous acquisition of associated MS/MS spectra. While MS1 peak intensities are typically used to compare the levels of secondary metabolites between groups, MS/MS fragment spectra often allow to obtain insights into structural information of the respective analytes. With different strategies, techniques such as data dependent acquisition (DDA) and data independent acquisition (DIA) rapidly trigger collision-induced dissociation allowing the collection of thousands of MS/MS spectra in one chromatographic analysis [32]. Both techniques first perform a full MS1 scan. In DDA, the intensity of the MS1 precursors is ranked to fragment only the most abundant precursor ions, which can bias the analysis as lower abundant signals remain unfragmented. The precursor ions and the acquired fragmented spectra are inherently linked, resulting in monoisotopically pure mass spectra. By contrast, DIA partitions all MS1 precursors in the full mass range in predefined isolation windows and acquires multiple MS/MS fragment spectra irrespective of the ion intensity [33]. Mass and retention time related features are extracted by specialized software such as MS-Dial [34] and XCMS [35]. As the readouts are multiplexed in DIA, direct links between the precursor ions and the fragment spectra are missing. This mandates subsequent chromatographic signal deconvolution and increases the complexity of the bioinformatic processing to increase spectral purity [36,37]. Studies comparing DDA and DIA approaches revealed considerably higher spectral coverage for DIA [38].

Recent improvements allow to extract thousands to a few ten thousands of MS1 features with their associated MS/MS spectra from a single chromatographic run [39]. However, typically 90% or more of the spectral metabolomics data in ecometabolomics studies contain compounds that never had been characterized before. This includes so called "known unknowns" (molecules with undefined structure that have been measured before) and "unknown unknowns" (molecules for which neither their identity nor their chemical structure is known) [40,41]. This is particular true for non-model organisms that are biochemically not well characterized, due to few analytical standards available, and very sparse spectral annotations in reference libraries [42]. Exploring the multitude of unknown compounds (the so called "dark matter" [43]) remains a significant challenge as the *de novo* structure elucidation and the identification is a very elaborate and a time-consuming process and often requires the use of additional analytical measurements. Lately, approaches have emerged that are performed in silico which can predict chemical structures based on molecular fingerprints or fragmentation patterns via computational methods [44,45]. However, they can only provide putative structures for a comparably small number of metabolites.

A viable alternative is the annotation of mass spectrometric features at the level of metabolite families. As the complexity and information of metabolic profiles can be overwhelming, studies with ecological context are often aiming at a more global description of metabolic patterns—a method called metabolite fingerprinting [46]. Even though compound classification provides "only" a generalized identity instead of explaining molecular identity in detail [47–49], in ecology this abstracted information can be more powerful.

Formerly, the determination of compound classes required the use of targeted analytical methods. Although there are classical extraction-based, chromatographic, electrophoretic, and structure-indicative spectroscopic methods (i.e., NMR) available, applying these methods often take biochemists months of work to characterize the different sets

of compound classes separately and mandate extensive purification of individual compounds [50]. Some of these methods may also be coupled to a mass spectrometer, but only allow for targeted analyses [51].

In the last few years, advances in computing power and the availability of libraries with a sufficient amount of spectral information made it possible to perform classification in silico using computational methods. These are much faster than traditional biochemical methods but their accuracy greatly depend on the quality of the acquired MS/MS fragment spectra and on the used algorithms. The development of a chemical ontology and the tool ClassyFire, which identifies compound classes from molecules with known structure, greatly facilitated these efforts [52,53]. There are numerous approaches described, such as MS2LDA, which decomposes fragment data to annotate molecular substructures [54], ClassyFirePredict, which can classify chemical motifs from spectral data of the Global Natural Products Social Molecular Networking platform (GNPS) using an artificial neural network [48], GNPS molecular networking, which is a community-based effort to annotate structural libraries [55], MetFamily, which performs hierarchical clustering on fragment spectra and allows annotation of different clusters of metabolite families [47], compMS2Miner, which calculates a consensus score for compound classes using external software [56], MS-Dial using a method to assign spectra to chemical motifs [36], a hybrid approach combining CSI:FingerID with ClassyFire [57], and CANOPUS, which can classify unknown fragment spectra using a deep neural network [49]. Even though many different compound classification algorithms have been developed, a general framework is still missing on how to apply these approaches to specific research questions.

Traditionally, the fields of molecular biology and ecology operate at distinct spatiotemporal scales with different resolution [4]. Metabolomics is a key technology that cannot only be applied in both of these contrasting research fields, but also used to connect these fields by providing a comprehensive view on any systemic organismal process without prior knowledge [49]. However, in molecular biology a systemic overview on molecular networks can be extremely complex, whereas in ecology the complexity of metabolite patterns can be overwhelming. Compound classification is an ideal tool to simplify these overly complex molecular networks and metabolite patterns. In molecular biology, discriminating compound classes are first determined and broadly assigned to molecular traits. In an additional step, by focusing only on the relevant compound classes, molecular entities are assigned to specific molecular networks while simultaneously accelerating the annotation processes. This approach allows to "zoom in" and further mechanistically relate compounds or compound classes to specific cellular and molecular processes [58]. By contrast, in ecology, compound classification allows for comparative approaches of different species, for relating ecological traits to species' biochemistry, or for exploring the responses of species to different conditions [6]. This is being achieved by "zooming out"—relating patterns in compounds or compound classes to ecological functioning at more coarse scales using multivariate statistics such as ordination methods [4].

In this study, we present a general methodological framework to comprehensively assess the chemical diversity using in silico classification. We apply the MetFamily classifier (available at https://github.com/ipb-halle/MetFamily, accessed on 22 March 2021), which employs a machine learning approach on consensus spectral information of over-represented fragmentation patterns [42,47]. While our preceding study [42] has focused on the classification of the most abundant compound classes present in bryophytes via DDA, in this study we apply the new data annotation methodology using untargeted raw DIA–MS/MS data from LC– sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH)–QToF analyses, which captures the majority of semi-polar compounds in biological samples. To demonstrate the general applicability of our framework, we use this representative data set with bryophytes, which is openly available in the repository MetaboLights as MTBLS851 [59].

Here, we specifically target research questions in chemodiversity and molecular ecology. We investigate metabolite profiles of ten different species of bryophytes and

provide a detailed comparative analysis at the level of compound classes comparing samples, which were collected in a fresh state (shock-frozen on-site as it would be in the lab) and stored in an herbarium for half a year. We show that our untargeted in silico classification framework provides an elaborate overview on broad systemic changes. The insights revealed by our framework enable a more detailed understanding of mechanistic molecular processes and allow to construct new research hypotheses and detailed follow-up studies.

2. Results

Raw metabolite profiles including the MS1 metabolite fingerprinting tables, MS/MS fragment spectra acquired in DIA mode, classification tables, and corresponding data tables, classifiers and meta-data are available as identifier MTBLS851 in the repository MetaboLights [59]. Additional results from metabolite fingerprinting, a comparison between positive and negative ion modes, quality control plots and tables of the MetFamily classification performances are available in the Supplementary Material. Abbreviations of species names used in the plots and the text are described in the section Abbreviations.

2.1. Chemical Diversity Analyses

The chemodiversity was investigated at the level of compound classes using our in silico classification approach which was applied to samples acquired in positive and negative ion mode separately. The data matrices were joined afterwards into a single matrix. Both the number of compounds (chemical richness) and the number of compounds unique to either of the levels was significantly larger in fresh samples than in samples stored in herbarium conditions (Figure 1a,b). This resulted in a significantly larger Shannon diversity index for fresh samples (Figure 1c). The Pielou's evenness was relatively similar in fresh and herbarium samples (Figure 1d).
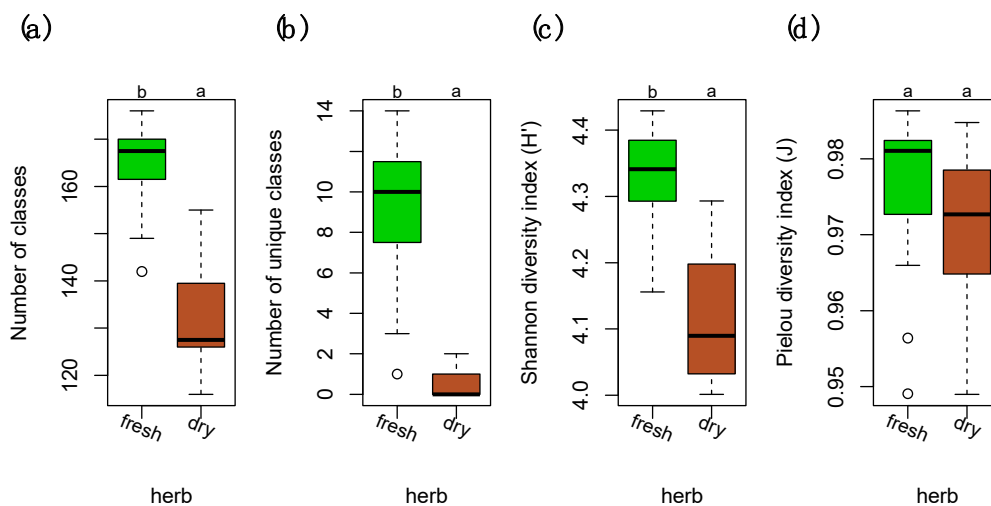

Figure 1. Diversity measures of compound classes in fresh and dry herbarium conditions. (**a**) Number of total compound classes. (**b**) Number of unique compound classes that were exclusively present in either of the levels. (**c**) Shannon diversity indices (*H'*) for the relative richness of compound classes present in fresh vs. dry. (**d**) Pielou's evenness (J) for the homogeneity of the distribution of the numbers of compound in the classes. Differences among groups (letters on the top of the plot) were calculated with performing the Tukey honestly significant difference (HSD) test post hoc on a one-way ANOVA. Different letters show significant differences ($p < 0.05$). n = 40 for each level. Factor levels were colored: green for fresh and brown for herbarium conditions.

Diversity analyses at the species level showed that *Marchantia polymorpha (Mar)* had significantly more unique compound classes than the other species (Figure 2b). The homogeneity in the richness of compound classes was significantly lower in *Mar* when compared

to the other species due to significantly more compounds in the dry condition (Figure 2d). The diversity was otherwise relatively similar in both acrocarpous and pleurocarpous species (Figure 2a–d).
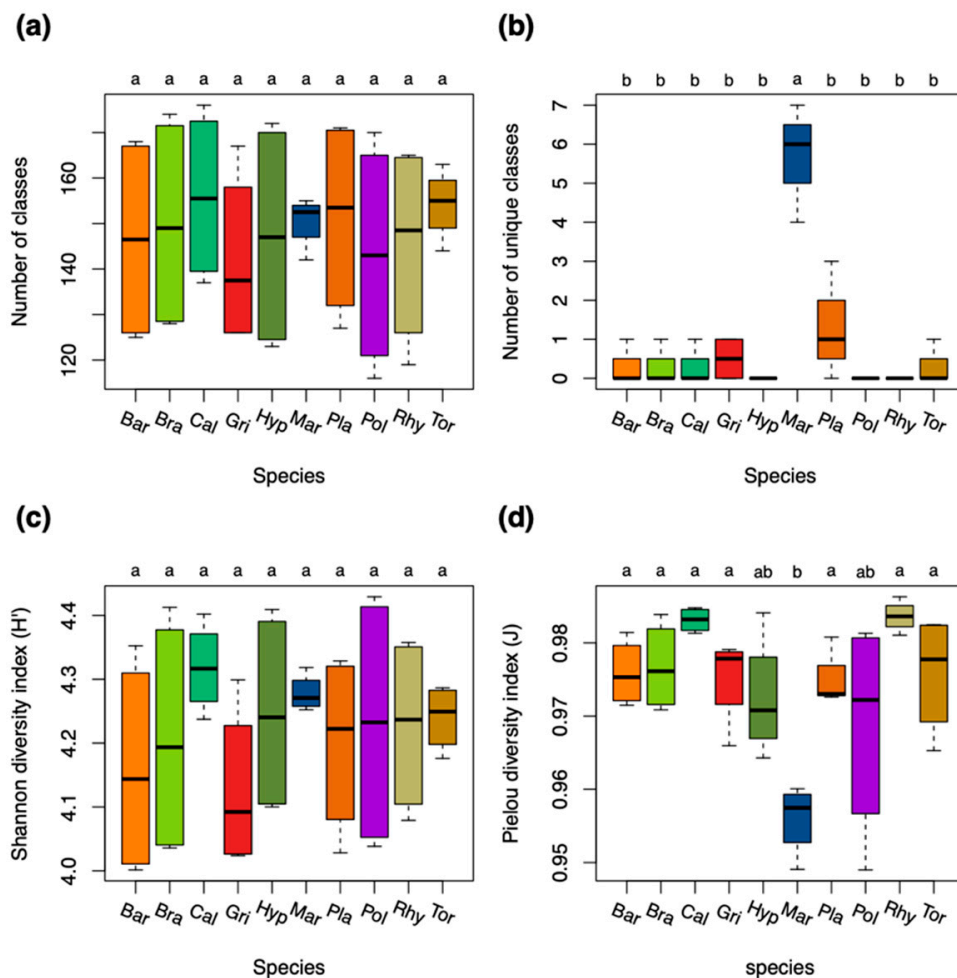

Figure 2. Diversity measures of compound classes for the ten bryophyte species. (**a**) Number of total compound classes. (**b**) Number of unique compound classes that were exclusively present in either of the species. (**c**) Shannon diversity indices (*H'*) for the relative richness of compound classes present in the species. (**d**) Pielou's evenness (J) for the homogeneity of the distribution of compound classes in the species. Differences among groups (letters on the top of the plot) were calculated with performing the Tukey HSD post hoc on a one-way ANOVA. Different letters show significant differences ($p < 0.05$). n = 8 for each species. Species were colored: red and brown colors for acrocarpous species, green and yellow colors for pleurocarpous species, and blue color for liverworts. Species codes are explained in the Abbreviations section.

2.2. Exploring Differences in the Composition of Metabolite Families

Differences in the composition of compound classes comparing fresh and dry herbarium conditions were explored using sunburst plots (Figure 3a,b) and the Fisher's exact test (Figure 3c). In the herbarium samples we found an increase in compounds belonging to the superclass of organic acids derivatives (compounds belonging to phosphate esters); organic nitrogen compounds (compounds belonging to quaternary ammonium salts, amines and organonitrogen compounds); the superclass of lipids and lipid-like molecules (several glycerophosphorylated compound classes like glycerophospholipids, -serines, -ethanolamines, and -cholines, fatty acid esters); organic oxygen compounds (compounds belonging to organic pyrophosphates, organic oxoanionic compounds); nucleosides, nucleotides, and

analogues (compounds belonging to ribonucleotides including purine and pyrimidine nucleotides and nucleosides); and benzenesulfonic acids and derivatives (Figure 3).
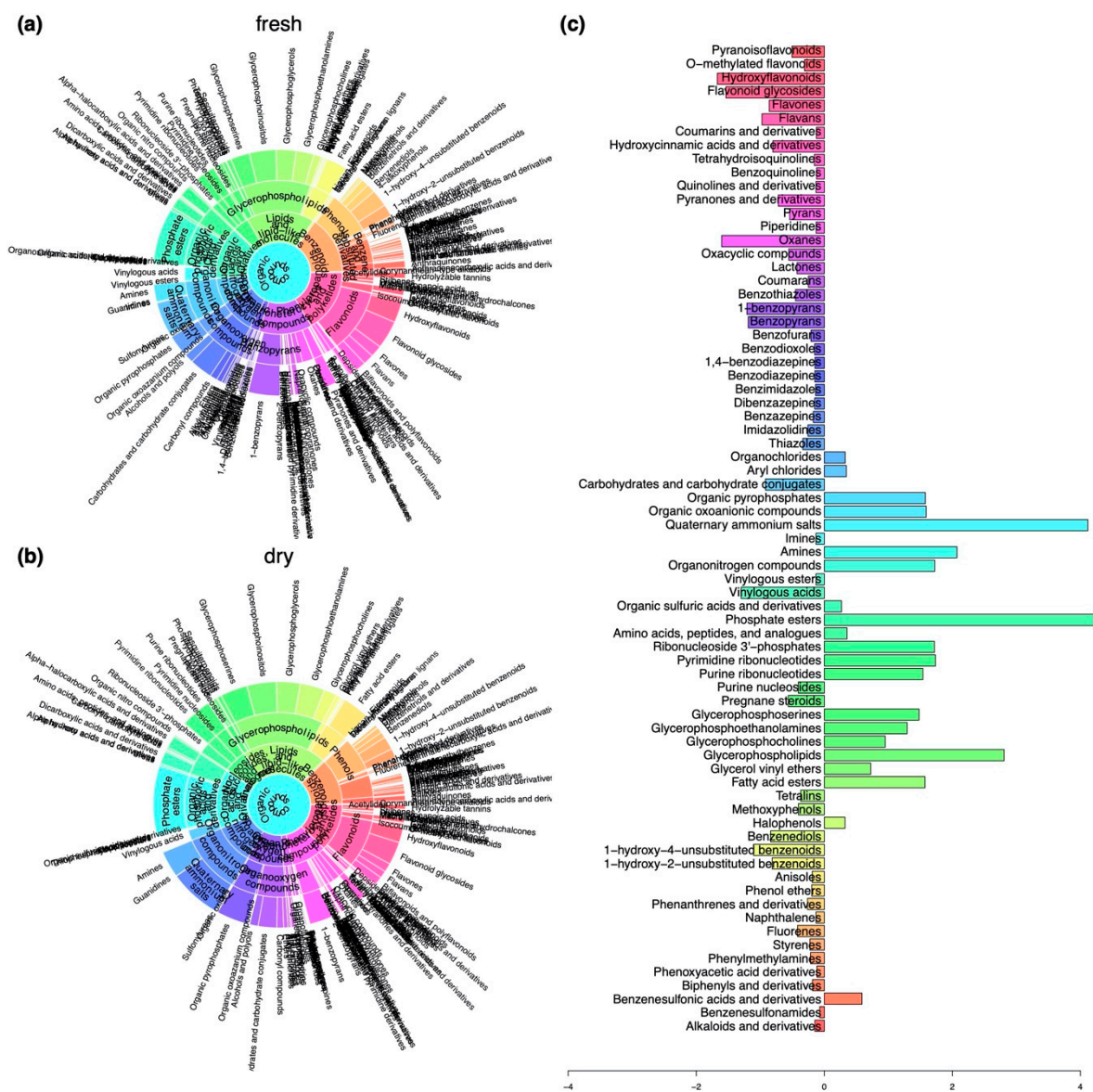

Figure 3. Differences in the composition of compounds belonging to compound classes at different levels of the chemical ontology with regard to herbarium conditions. (**a**) Sunburst plot showing the composition of compounds in samples that were collected in a fresh state. (**b**) Sunburst plot showing the composition of compounds in samples that were stored in the herbarium. (**c**) Changes in relative frequencies of compound classes when comparing fresh with herbarium samples determined using the Fisher's exact test. Negative values indicate a decline and positive values mark an enrichment in the respective compound classes under herbarium conditions.

We determined decreases in the numbers of compounds present in the superclasses of organoheterocyclic compounds (oxanes, pyrans, pyranones, and derivatives, oxacyclic compounds, benzopyrans, benzothiazoles, benzodiazepines, (di)benzazepines, benzofurans, carbohydrates, and carbohydrate conjugates, thiazoles, lactones), benzenoids (benzenediols, unsubstituted benzenoids, styrenes, tetralins, methoxyphenols), phenylpropanoids and polyketides (hydroxyflavonoids, flavonoid glycosides, flavans, flavones, pyranoisoflavonoids, hydroxycinnamic acids and derivatives, coumarins and coumarins and their derivatives), organic acids and derivatives (vinylogous acids), and lipids and lipid-like molecules (pregnane steroids) (Figure 3).

A variable selection using sparse partial least squares discriminant analysis (sPLS–DA) was performed to evaluate how the differences in compound classes between fresh and dry conditions were constituted in the different species (Figure 4). In short, the multivariate sPLS–DA found similar results as the univariate Fisher's exact test (Figure 3c). From the selected compound classes, fresh samples of *Mar* were clustered together with the herbarium samples of the other species (Figure 4). Only the classes of alkaloids and derivatives, pregnane steroids, benzothiazoles, and lactones were selected that were more present in fresh vs. herbarium samples in *Mar* (Figure 4). *Plagiomnium undulatum (Pla)* and *Polytrichum strictum (Pol)* were more dissimilar with regard to richness in compound classes when compared to the other species.
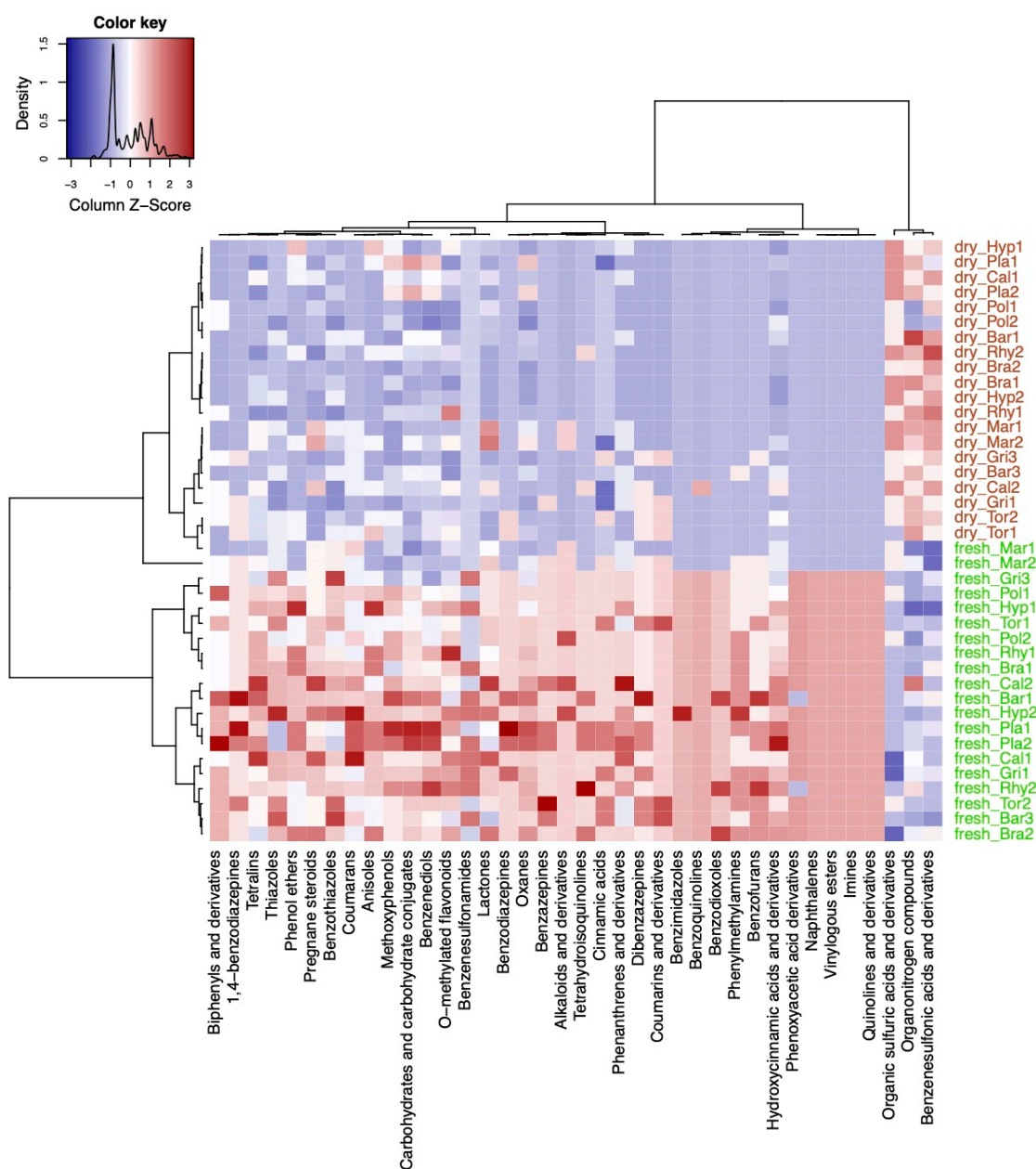

Figure 4. Results from the variable selection with sparse partial least squares discriminant analysis (sPLS-DA). Heatmap of the selected variables by the sPLS-DA model using the compound classification table. Compound classes that were enriched are shown in red color and classes that were decreased are shown in blue color in the plots. Sample names starting with either fresh or dry indicate fresh and herbarium samples, respectively. The species codes are further explained in the abbreviations section. Factor levels in the column were colored: green for fresh and brown for herbarium conditions.

sPLS-DA selected the classes of organic sulfuric acids, organonitrogen compounds, and benzenesulfonic acids and derivatives to be significantly enriched in herbarium samples except for *M. polymorpha* (Figure 4).

Next, we compared differences in numbers of compound classes in the different species (Figure 5). The sPLS-DA model selected five compound classes which were unique for *Mar*: corynanthean–type alkaloids, dibenzylbutane lignans, pyridoindoles, depsipeptides, carbazoles (Figure 5). The compound classes of linear diarylheptanoids, benzyloxycarbonyls, and tetrahydrofuran lignans were also found in *Pla* and *Pol* and *Rhy* (*Rhytidiadelphus squarrosus*). The compound classes of vinylogous acids, hydroxyflavonoids, ethers, organoheterocyclic compounds, heteroaromatic compounds, flavones, and pyranones and derivatives were enriched in fresh and dry samples of *Pla*, and in fresh samples of the pleurocarpous species *Rhy*, *Hyp* (*Hypnum cupressiforme*), *Bra* (*Brachythecium rutabulum*), and *Cal* (*Calliergonella cuspidata*). Hydroxysteroids, oxosteroids, gamma butyrolactones, alcohols and polyols were enriched in *Cal* (fresh and dry) and *Bra* (fresh only), and *Mar* (fresh only) when compared to the other species. Compounds belonging to beta lactams, naphthols and derivatives, and naphthalenecarboxylic acids and derivatives were enriched in pleurocarpous species. The acrocarpous species were less distinct than liverworts and pleurocarpous species and rather characterized by relatively low numbers in the aforementioned compound classes. *Pol* had enriched phosphosphingolipids, imidazolidines, and thiazolidines when compared to other acrocarpous species. In general, dibenzazepines, coumarins and derivatives, and isoflavans were more likely associated with acrocarpous species (Figure 5).

2.3. Comparison of Chemotaxonomic Analysis with Phylogeny

A phylogenetic tree of the bryophytes was extracted from the Open Tree of Life project (Figure 6a) and compared with chemotaxonomic trees reconstructed from the metabolite fingerprinting data (Figure 6b), compound classification data in negative ion mode (Figure 6c), and compound classification data in positive ion mode (Figure 6d). The chemotaxonomic information from metabolite fingerprinting allowed a good reconstruction of phylogenetic relationships (Figure 6a). Liverworts, pleurocarpous and acrocarpous bryophytes were clustered in separate clades, with *Mar* being the most dissimilar species. The chemotaxonomic clustering of the in silico classification data in negative ion mode (Figure 6c) resulted in *Pol* to be the most dissimilar species with regard to the richness and composition of its compound classes. Acrocarpous and pleurocarpous species were clustered apart from each other. In contrast to the phylogenetic tree, *Mar* was clustered together with the acrocarpous species. The chemotaxonomic tree calculated from the classification data in positive ion mode (Figure 6d) resulted in a heterogeneous distribution of species in the tree which was in contrast to the phylogenetic tree. *Mar* was still the most dissimilar species, but acrocarpous and pleurocarpous species were intermixed.

2.4. Trait Analysis

Relationships of ecological characteristics were investigated using distance-based redundancy analysis (dbRDA) with regard to the metabolite fingerprinting and in silico classification data. Ecological characteristics like Ellenberg indices and capsule positions were acquired from literature for the used bryophyte species. Tables are available in [60,61]. The dbRDA model using the metabolite fingerprinting data explained a variation of 10.54% on the first two axes of the RDA ordination (Figure 7a). The model found that the metabolism of pleurocarpous species was associated with inclined capsule positions, procumbent stems, plagiotropic growth form, and relatively large spore diameters (a mean of 18.75 μm for pleurocarpous vs. 15.1 μm for acrocarpous species). The acrocarpous *Pla* was clustered near the pleurocarpous species due to having similarly sized spores (28 μm large). Metabolite features of acrocarpous species were associated with erect to inclined capsule positions, and those of *Mar* were separated by the characteristic vegetative regeneration via gemmae (Figure 7a).
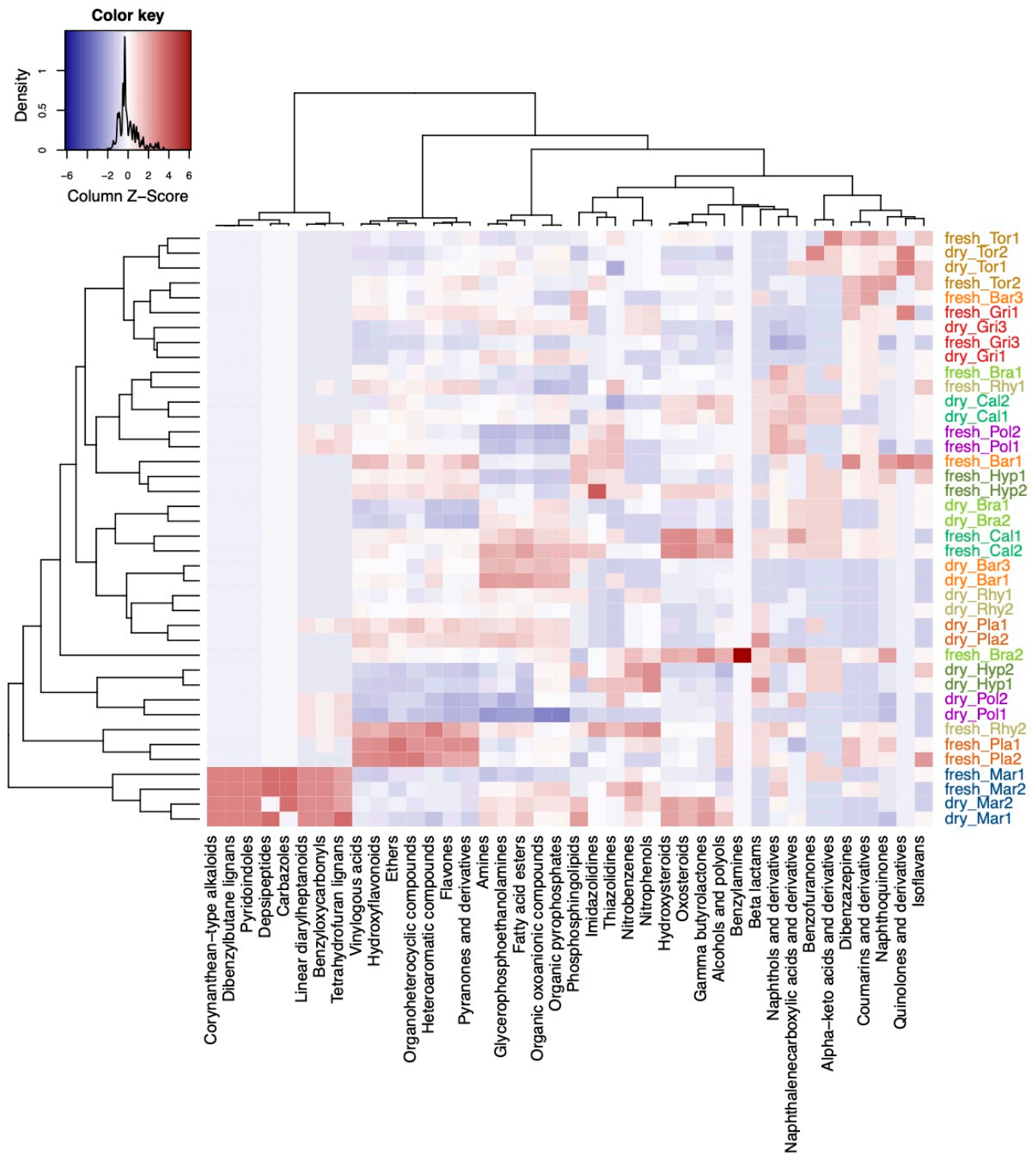
Figure 5. Differences in the composition of compounds belonging to compound classes at different levels of the chemical ontology with regard to the different species. Shown is a heatmap of the selected variables by the sPLS-DA model. Compound classes that were enriched in the species are shown in red color and classes that were decreased are shown in blue color in the plots. Sample names starting with either fresh or dry indicate fresh and herbarium samples, respectively. The species codes are further explained in the Abbreviations section. Species were colored: red and brown colors for acrocarpous species, green and yellow colors for pleurocarpous species, and blue color for liverworts.
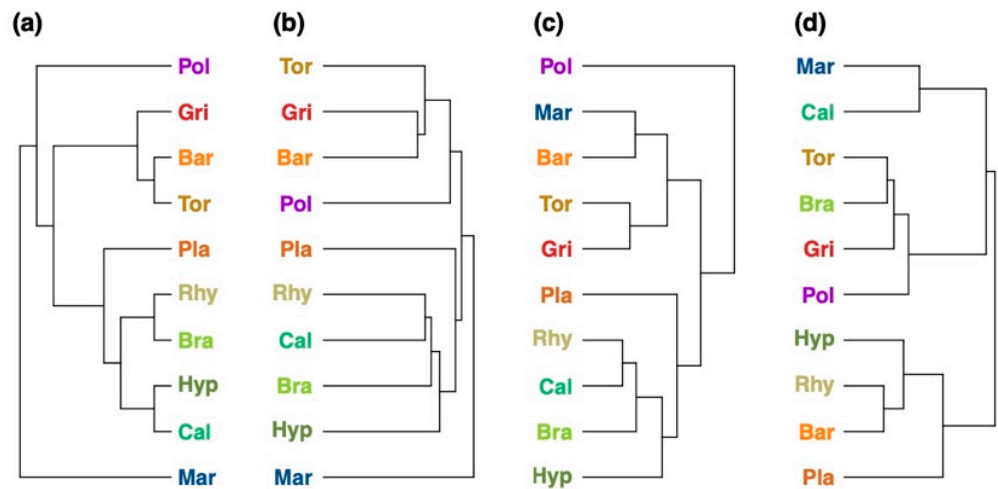
Figure 6. Comparison of the (**a**) phylogenetic tree with the (**b**–**d**) chemotaxonomic trees of the 10 bryophyte species reconstructed by (**b**) metabolite fingerprinting, (**c**) in silico classification performed in negative ion mode, and (**d**) in silico classification performed in positive ion mode. The phylogenetic tree showing the phylogenetic relationships of the ten bryophyte species. Comparison metrics: (**b**) Mantel statistic: r = 0.321, cophenetic correlation coefficient: c = 0.653, Robinson–Foulds metric: rf = 0.5; (**c**) Mantel statistic: r = 0.005, cophenetic correlation coefficient: c = 0.042, Robinson–Foulds metric: rf = 0.75; (**d**) Mantel statistic: r = 0.151, cophenetic correlation coefficient: c = 0.0, Robinson–Foulds metric: rf = 1.0. Species names were colored: red and brown colors for acrocarpous species, green and yellow colors for pleurocarpous species, and blue color for liverworts. The species codes are explained in the Abbreviations section.
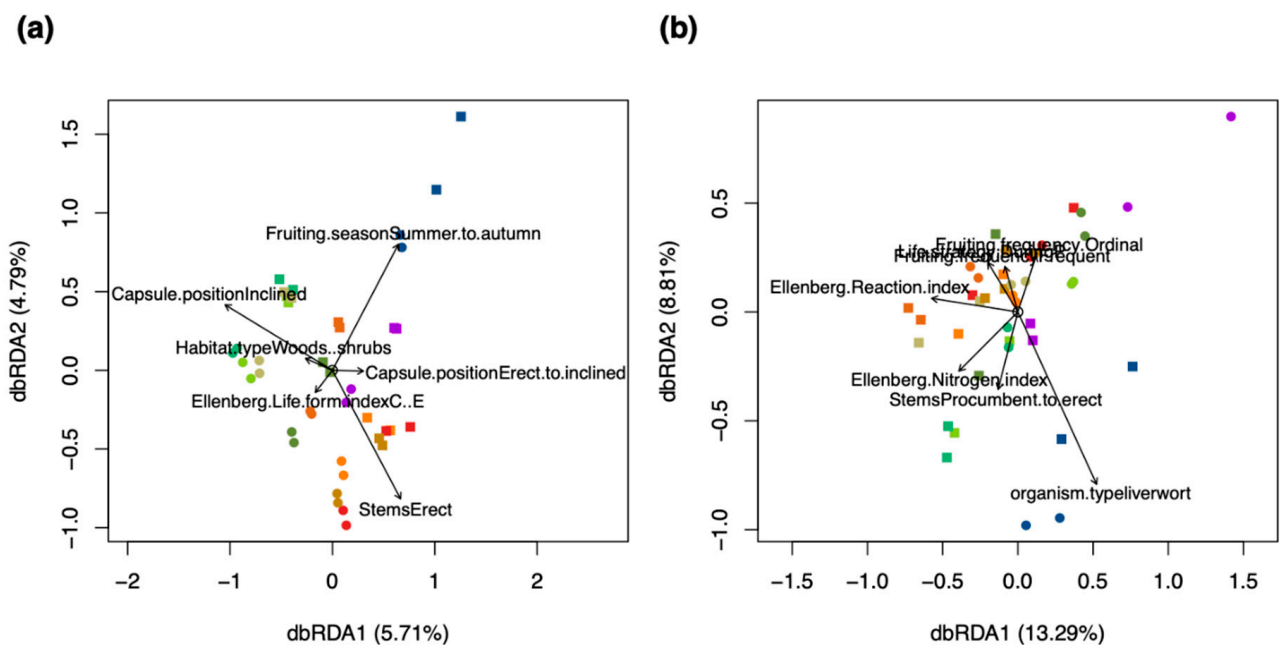

Figure 7. Relationships of ecological characteristics with (**a**) the metabolite fingerprinting data, and (**b**) in silico classification data. The lengths of the arrows represent the explanatory power of the variable with regard to the data. The positions of the samples relative to the direction of the arrow represent the relationship of the samples with the variable. The scores were colored according to the ten species used in the study: red and brown colors for acrocarpous species, green and yellow colors for pleurocarpous species, and a blue color for liverworts. A round shape of the scores represents samples of dry herbarium conditions and a square represents samples of fresh conditions. n = 80.

Investigating the dbRDA comparing the in silico classification data with ecological characteristics revealed a model that explained greater variation (21.86%) than metabolite fingerprinting on the feature level (10.54%) (Figure 7a,b). The composition of compound classes in the different species was found to be related to environmental characteristics like the Ellenberg moisture, reaction and nitrogen indices (Figure 7b). The discrimination of species was larger in the metabolite fingerprinting data, where also species-specific characteristics like growth forms and capsule positions were more important than environmental characteristics (Figure 7a,b).

3. Discussion

The annotation of fragment spectra in LC/MS–MS data is still a highly challenging task [62]. Annotation is typically performed by searching spectra in spectral libraries, such as MassBank, GNPS, or the WEIZMASS library [55,63–65]; or by using in silico annotation tools like MetFrag, CFM-ID or SIRIUS [44,66,67]. However, metabolite structure databases usually cover only few compounds present in samples of biological non-model species. A viable alternative are in silico classifiers that perform a less specific approach and avoid complex biochemical preparation procedures [54,68,69]. In this study, we used the MetFamily classifier, which recognizes structural features in fragment spectra instead of predicting the structure of entire molecules. The classifier has been trained on sets of usable high-quality spectra from the MassBank of North America (MoNA) database and is currently representing the 677 compound classes covered by MoNA [63].

Chemical classification can be performed at different levels of the chemical ontology [53]. When interpreting the results from the MetFamily classifier, it returns a ranked list of candidate classes per spectrum similarly to the results of in silico annotation tools like MetFrag or SIRIUS [44,66]. The correct class may not be ranked first and could be hidden in the list of alternative parents. For instance, a flavonoid glycoside could primarily be classified as a glycoside, or a diterpenoid primarily as a phenylpropanoid. Both classifications are correct but may not reflect the desired classification level from a biochemist's perspective. Here, interpretation of the results can be facilitated when restricting the classification to a specific level of the chemical ontology, i.e., at the level of superclasses, classes or subclasses; which is still meaningful in ecometabolomics. The results of our classifier can be improved by increasing the number of high-quality spectra for training the machine learning model covering a wider range of compound classes. The scoring algorithm of the classifier can also be adapted by integrating skilled biochemists in the classification process. Furthermore, methods, such as deep learning and boosted regression trees, are promising to improve the accuracy of the model [49]. However, the classification of truly novel substances will remain a challenge due to missing reference spectra.

Our chemical classification framework revealed several insights into the chemodiversity of bryophytes and molecular changes under herbarium conditions. The lower chemodiversity of metabolite profiles show that many secondary metabolites degrade due to breakdown. The compound classes of quaternary ammonium salts, amines, and other organonitrogen compounds were enriched in herbarium samples indicating the degradation of amino acids, peptides, and analogues as products of a non-functional respiration metabolism and general catabolic processes [70–72]. The classes of quaternary ammonium salts are also structurally related to many functional proteins and nitrogen containing glycerophospholipids, which constitute metabolites produced by bryophytes and lichens against desiccation and as a response to reactive oxygen species [28,73]. We further detected an increase in phosphate esters and nucleosides, nucleotides and analogues, which are likely involved in phosphorylation in the fatty-acid metabolism in order to retain membrane structure [31].

It has often been described that bryophytes produce carbohydrates that protect them from osmotic stress and desiccation; thus, carbohydrates are involved in the glycolysis and gluconeogenesis, the synthesis of proto-lignin to stabilize cell walls, the citrate cycle, and photosynthesis and related sugar metabolisms [74,75]. Our results show a downregulation

of carbohydrates under herbarium conditions, which suggests that carbohydrates detected by LC/MS only play minor roles with protection from desiccation and degradation [73].

Flavonoids have been reported to be one of the most diverse groups of natural products with regard to the response of bryophytes to environmental changes, including the protection from excessive light, temperature and drought stress, and various biotic interactions [76,77]. When comparing fresh with herbarium samples, we detected significantly fewer flavonoids (including the subclasses hydroxyflavonoids, flavonoid glycosides, flavones, flavans, and hydrocinnamic acids). The decrease can be explained by the generally low stability of flavonoids. Under severe stress, bryophytes may further stimulate the degradation of flavonoids to scavenge free radicals [78]. As flavonoids have been described to regulate protein stability, their absence would also explain the many quaternary ammonium salts and fragments of amino-acids under herbarium conditions.

In the herbarium samples, we detected the absence of many steroids and their derivatives. They perform important roles in plant development, are precursors in the synthesis of phytohormones, and are involved in defenses against pathogens and in the degradation of xenobiotics (e.g., pollution) [9,79–81]. In *Marchantia*, bibenzyls and bisbybenzyls are the most abundant steroidal products. They are considered important biomarkers with regard to biotic and environmental interactions and the development of pharmaceutical and bio-control agents [18,28,29,82]. Under herbarium conditions, many of these highly bioactive products degrade or fragment into smaller components. We further detected that many benzenic compounds were lost from herbarium samples. Bryophytes produce structurally complex benzenoids, including various phenolics, lignans, phenylpropanoids, and terpenes. Many of these compounds have been found to be located in the oil bodies of liverworts [83], explaining the functional roles related to biotic and abiotic stress responses in living tissues [84,85].

In ecometabolomics, metabolite fingerprinting is now widely used to reveal metabolic patterns that drive ecological functioning. Here, we present a computational framework to annotate metabolite families using DIA–LC/MS–MS data. We were able to assign MS/MS fragment spectra to 91.5–95.7% of MS1 features (in negative and positive ion mode, respectively) and, thus, capturing the predominant part of semi-polar metabolites in plant samples. Our analytical methodology allows for a near complete picture of the composition of secondary metabolites in the tested bryophyte species. Despite the fact that classification simplifies information on the characteristics of compounds, the results from our ordination with dbRDA show that when correlating ecological characteristics with the classification data, the level of explained variance is improving (Figure 7). However, as the intra-specific variability of species was very large, environmental responses of bryophytes explained more variance in the composition of compound classes than species-specific and life-history characteristics. This is likely due to the fact that bryophytes produce a magnitude of secondary metabolites in order to maintain homeostasis to balance for a lack of specialized anatomical structures in response to environmental conditions [10]. Our results are in line with other findings that bryophytes produce low amounts of nitrogen-containing compounds, which has been attributed to an evolutionary advantage in nitrogen-poor environments [10,86,87]. Here, bryophytes restrict nitrogen-containing products to structurally complex metabolites like benzenoids (phenolics), phenylpropanoids (terpenes), amino acids, and alkaloids [70,88].

As with many ecometabolomics studies, the amount of molecular information can be overwhelming. Our results show that classification can be used to improve and facilitate the ecological interpretation of the physiological and biochemical responses of bryophytes to biotic and abiotic environmental conditions [6]. Patterns in the diversity of metabolite families and related physiological characteristics can be used to define chemical traits that further explain variations in species distributions and niches [25], which are described by a close relationship between bryophyte functional categories and microhabitat conditions [89]. The trait analyses from this study also suggest that indicator values and certain life-history characteristics are not suited to correlate molecular responses of bryophytes to

environmental conditions. We suggest to use a combination of morphological and anatomical traits, such as shoot mass per area, photosynthetic nitrogen use efficiency, specific stem diameter, height, or leaf area, which have been described to perform better for vascular plants [87,90].

We propose to use our in silico classification framework in future studies of ecometabolomics to evaluate chemical traits (to evaluate those compound classes that have major ecological impact or serve as bioindicators for certain ecological relationships). For instance, using the flavonoid richness as a chemical trait to quantitatively assess the exposition to sunlight of the sampled individuals in the habitat instead of the generalized Ellenberg indicator for light, which is an average literature value for the species. We expect that when using these novel chemical traits, ecological explanation can be improved [91–93]. Further research and detailed comparative studies are needed to precisely assess the composition and richness of metabolite families in bryophytes and other plants. As in many ecometabolomics studies DDA–LC/MS–MS is already part of the analytical pipeline, our classification framework can be directly applied to these data and, thus, enabling new insights through the evaluation of chemical traits based on metabolite families. Our framework can also be applied retroactively to existing data sets, where the full extent of DDA–LC/MS–MS may have not been explored at the global level and, thus, fostering the re-use of ecometabolomics data [94].

Chemotaxonomic analyses have been considered to be a viable alternative over complicated DNA sequencing of bryophytes due to the lack of detailed genetic maps and the presence of repetitive sequences [85]. We encourage to use our framework to explore inter-specific variability in the composition of metabolite families of species to differentiate phylogenetically relevant metabolites or compound classes in different species ("phylomarkers") that can be used to determine cryptic species without requiring expensive DNA barcoding methods [85,95,96]. The presence or absence of particular biomarkers or compound classes has become an important tool for the identification of lichens and has been applied for almost 50 years [96–100]. Comparing the chemical diversity of different bryophyte species may also shed some more light on the chemical evolution of phylogenetically closely (i.e., cryptic species) or distantly related species, may reveal previously unknown connections in molecular networks, and can accelerate the finding of phylogenetically distinct biomarkers [95]. The results from this study showed that reconstructing the chemotaxonomic tree from classification data in negative ion mode was more consistent with phylogeny than when using classification data in positive ion mode. The results from the trait analysis indicate that in negative ion mode more species-specific metabolites were acquired by our analytical setup, whereas more metabolites regarding environmental adaptation (e.g., correlations of compound classes to Ellenberg Reaction index) were acquired in positive ion mode, which highlights the need to acquire both ionization modes in experiments.

In plant ecology, biochemical analyses are often performed on herbarium plant material as experiments are often conducted in the field. The results from this study show that the spectrum of compounds found in herbarium material was considerably dissimilar to fresh material. As a result, chemodiversity analyses conducted solely on herbarium specimen may be biased and the validity of results may be restricted to the classes which are found in herbarium samples. Our results show that the differences between fresh and herbarium samples are highly species-specific. Further research is needed to explore how large the spectrum of compound classes is changed in different species when sampled fresh vs. using herbarium specimen. Once the changes in the composition of compound classes in different species have been determined, regression methods like nonlinear local polynomial (LOESS) based normalization or support vector regression (SVR) may be applied in order to correct for the bias [101–103].

Our novel classification framework can provide a broad overview on the chemical characteristics of any kind of biological species by assessing the chemodiversity and the composition of compound classes. By "zooming out" biochemical patterns can be related to

ecological functioning at more coarse scales. As most bryophytes are typically non-model species, they represent an ideal use case in the research field of ecometabolomics as they contain a large portion of completely novel compounds. Our method can be applied at different levels of the chemical ontology, i.e., at the level of superclasses, classes or subclasses, potentially revealing ecological patterns at different scales. We also kept our framework flexible so that it is possible to use different spectral classifiers (e.g., among others CFM-ID, GNPS, CANOPUS [49,57,67]).

We further recommend to apply our framework in classic metabolomics studies. For instance, it is well known that plants produce more flavonoids when grown in sunlight compared to growing in shade. Automatically selecting the relevant compounds belonging to the chemical class of flavonoids would greatly facilitate the work of biochemists as they could focus the identification only on the selected spectra of a particular metabolite family. In comparative studies, classification would also allow to directly investigate the role of particular metabolite families, helping to characterize physiological processes more quickly at broader levels, to uncover patterns in metabolite fingerprints, and to simplify complex molecular patterns in order to find associated and expressed genes [58]. Thus, our classification framework is perfectly suited for hypotheses generation and to create follow-up studies by focusing on particular compounds or compound classes.

4. Materials and Methods

4.1. Sampling

Samples of the ten bryophyte species Bar (Barbula unguiculata Hedw.), Bra (Brachythecium rutabulum (Hedw.) Schimp.), Cal (Calliergonella cuspidata (Hedw.) Loeske), Gri (Grimmia pulvinata (Hedw.) Sm.), Hyp (Hypnum cupressiforme Hedw.), Mar (Marchantia polymorpha L.), Pla (Plagiomnium undulatum (Hedw.) T.J. Kop.), Pol (Polytrichum strictum Menzies ex Brid.), Rhy (Rhytidiadelphus squarrosus (Hedw.) Warnst.), and Tor (Tortula muralis Hedw.) were collected at random locations in the area of the Botanical Gardens of the Martin Luther University of Halle-Wittenberg. Sampling was performed in spring on 18 May 2017at stable and sunny conditions after mid-day between 1:00 p.m. and 3:00 p.m.

Four composite samples of the sterile gametophytes (multiple individuals) of each species were taken, leading to a total of $4 \times 10 = 40$ samples. Only above-ground parts of gametophytes, such as leaves, branches, stems, or thalloid parts were taken for sampling. Visible archegonial and antheridial heads and any belowground parts such as rhizoids and rooting stems were removed with a sterile tweezer. One half of the sample collection ($2 \times 10 = 20$ samples) were put in Eppendorf tubes and instantly frozen on dry ice. The other remaining fresh samples ($2 \times 10 = 20$ samples) were put in non-chlorinated tea bags (to allow air-drying and to avoid possible chemical side-reactions with the tea bags) and stored in an herbarium at Leibniz Institute of Plant Biochemistry. On 14 November 2017, two composite samples of each herbarium-dried species ($2 \times 10 = 20$ samples) were put in Eppendorf tubes and stored at -80 °C. On 30 January 2018, samples were prepared for the analytical analysis. Metabolomics analyses were performed based on protocols of [39,104]. They are described in detail further below.

4.2. Metabolite Extraction

A total of 25 mg of dry and 50 mg fresh tissue were weighed under cryogenic conditions into wall-reinforced cryo-tubes of 1.6 mL volume (Precellys Steel Kit 2.8 mm, Peqlab Biotechnologie GmbH, Erlangen, Germany) filled with 3 steel beads (3 mm), 1 steel bead (5 mm), and 200 mg glass beads. Then, 900 µL of a cold mixture of dichloromethane/ethanol (-80 °C) was added first followed by 150 µL of 30 mM HCl (4 °C). Cell rupture/metabolite extraction was assisted by FastPrep bead beating (3×20 s, speed 6.5 m/s, FastPrep24, MP Biomedicals LLC, Santa Ana, CA, USA) followed by phase separation by centrifugation at $20,000 \times g$ (2 min, 4 °C). After removal of the upper aqueous phase, 100 µL 30 mM HCl

were added and bead milling and centrifugation were repeated. The aqueous phase was completely removed.

Semi-polar and apolar metabolites were collected from the denser organic phase, from which the extract was aspirated from the bottom using a narrow bore pipet. Exhaustive extraction was achieved by addition of 500 μL fresh tetrahydrofuran (THF) to the cell debris, bead milling, and centrifugation. The second organic extract was aspirated from top, unified with the aliquot of the first organic extract and dried in a stream of nitrogen gas before storage at −80 °C. Semi-polar metabolites dissolved in 180 μL 80% MeOH were centrifuged and the supernatant was collected for LC/MS analysis within 24 h.

4.3. Metabolite Separation

Separation of medium polar metabolites was performed on a Nucleoshell RP18 (2.1 × 150 mm, particle size 2.1 μm, Macherey & Nagel, GmbH, Düren, Germany) using a Waters ACQUITY UPLC System, equipped with an ACQUITY Binary Solvent Manager and ACQUITY Sample Manager (20 μL sample loop, partial loop injection mode, 5 μL injection volume, Waters GmbH, Eschborn, Germany). Eluents A and B were aqueous 0.3 mmol/L NH_4HCOO (adjusted to pH 3.5 with formic acid) and acetonitrile, respectively. Elution was performed isocratically for 2 min at 5% eluent B, from 2 to 19 min with linear gradient to 95% B, from 19–21 min isocratically at 95% B, and from 21.01 min to 24 min at 5% B. The flow rate was set to 400 μL/min and the column temperature was maintained at 40 °C. Metabolites were detected by positive and negative electrospray ionization and mass spectrometry. The analytical Quality Control (QC) involved batches with QC-standard mixtures ("MM10") and blank samples to control memory effects and to ensure retention time stability. The MM10 mixture included the following compounds which give peaks in negative and positive ion modes: DL-a-Phenylglycine (151.06333), 6-Furfurylaminopurine (215.08071), Rutin (610.15339), o-Anisic acid (152.04734), Phlorizin Dihydrat (436.13695), 3-Indolylacetonitrile (156.06875), N-(3-Indolylacetyl)-L-valine (274.13174), Biochanin A (284.06847), Carnosol (330.18311), Abietic acid (302.22458). Additionally, the column was washed with 2-propanol to remove also hydrophobic residues after each batch. Afterwards the column was washed back with 100% acetonitrile for column storage.

4.4. Untargeted Mass Spectrometry

Untargeted mass spectrometric analysis of small molecules was performed with MS–ToF–SWATH–MS/MS (TripleToF 5600, both Sciex GmbH, Darmstadt, Germany) operating in negative or positive ion mode and controlled by Analyst 1.6 TF software (Sciex GmbH, Darmstadt, Germany). The source operation parameters were as following: ion spray voltage, −4500 V/+5500 V; nebulizing gas, 60 psi; source temperature, 600 °C; drying gas, 70 psi; curtain gas, 35 psi. TripleToF instrument tuning and internal mass calibration were performed every 5 samples with the calibrant delivery system applying APCI negative or positive tuning solution, respectively (Sciex GmbH, Darmstadt, Germany).

For the MS1 measurements, ToF masses were scanned between 65 and 1250 Dalton with an accumulation time of 50 ms and a Q2 energy of 10 V (−10 V for negative ion mode), which is the vendor recommended setting for improved ion transmission in MS1 [39]. The MS2-SWATH-experiments were divided into segments of 20 ms accumulation time. Together the SWATH experiments covered the entire mass range from 65 to 1250 Dalton in 48 separate scan experiments, which allowed a cycle time of 1.1 s. As chromatographic peak widths varied depending on concentration and analyte, the core shell technology was used, which normally produces peaks in the range of very few seconds (<20 s). For the narrowest peaks 6–10 duty cycles per peak were managed. For an average peak at least 10 cycles per peak were achieved with our instrument [104]. Throughout all MS/MS scans a declustering potential of 35 (or −35 V) was applied. Collision energies for all SWATH–MS/MS were set to 35 V (−35) and a collision energy spread of ±25 V, maximum sensitivity scanning, and elsewise default instrument settings [104].

4.5. Raw Data Acquisition

Centroid raw data were converted to the mzML format with the msconvert tool of the ProteoWizard software suite [105]. The converted raw data were uploaded to MetaboLights and metadata were recorded in compliance with the minimum information guidelines for Metabolomics studies [106,107]. Profiles in positive and negative ion mode were recorded and used for further data analyses.

Prior to data analyses, a matrix with two factors was constructed containing the levels for herbarium (fresh, dry) and the species (Bar, Bra, Cal, Gri, Hyp, Mar, Pla, Pol, Rhy, Tor).

4.6. Peak Detection

Chromatographic peak detection was performed in R 4.0.2 (available at https://cran.r-project.org, accessed on 8 January 2021) with the package XCMS 3.11.7 [35] separately for positive and negative ion modes. The following settings were used performing the centWave algorithm: ppm = 25, mzCenterFun = mean, prefilter = c(3, 100), peakwidth = c(10, 20), snthresh = 5, integrate = 2. Chromatographic peaks were then grouped using the factor matrix as pheno-data. Grouping was based on time dimension peak densities using PeakDensityParam with the parameters: minFraction = 0.5 and bw = 20. Retention time correction was performed using the parameters minFraction = 0.5, smooth = loess, span = 0.2 and family = Gaussian. Peaks were grouped again afterwards. In order to assess the performance of the peak detection, a technical validation procedure (quality control) was performed. Deviations in m/z and retention times were evaluated in the metabolite profiles of the bryophytes and the MM10 samples as part of the analytical quality control. Figures S11–S15 in the Supplementary Material show these quality control plots.

Peak detection of MS/MS peaks in all isolation windows (buckets) of the DIA-mode was used with the same parameters as the MS1 peak detection using the findChromPeaksIsolationWindow function of XCMS. Only MS/MS spectra with precursors between retention times of 60 to 1200 s were retained (using the filterRT function). Peaks before 60 s would contain numerous salt signals and non-retained hydrophilic metabolites, which were too difficult to separate without exact retention time information. Late eluting peaks (>1200 s) contained lipids with very broad elution profiles and were therefore excluded as well. For each MS1-level peak the according MS/MS spectra were reconstructed using the function reconstructChromPeakSpectra with the parameter minCor = 0.5. The MS/MS spectral information was converted to a peak list using the fData function.

The MS1-level peak table was constructed with the function featureValues and the parameters method = medret and value = into were applied. Only those peaks were kept for which MS/MS spectral information were reconstructed (91.5% for negative mode, 95.7% for positive mode). Overlapping spectra were merged using the function combineSpectra using the parameters mzd = 0.05, ppm = 50, and intensityFun = "max". All spectra present in a sample were saved in MSP files, which were used later for the in silico classification. The peak table was log transformed and missing values were imputed with zeros. A binary matrix to be used for compositional analyses was created containing the presence and absence values of compounds. A compound was deemed as present when its abundance was above 0.1% the abundance of the largest peak.

The above steps were implemented for positive and negative ion mode separately. The data matrices were joined afterwards into a single matrix.

4.7. In Silico Classification

In silico classification was performed using the MetFamily classifier [42,47] for positive and negative ion modes separately. The MetFamily classifier was created using a machine learning approach on a training set of approximately 57,000 MS/MS fragment spectra in negative and positive ion modes with known structures from the MassBank of North America (MoNA) [63] enriched with terms from the ChemOnt ontology (version 2.1) [53]. In a first step, the training data were cleaned in such a way that only fragment spectra of

usable quality are being used for training. This step resulted in a total of 11,328 fragment spectra for negative ion mode and 21,908 for positive ion mode. The MetFamily algorithm first decomposed the spectral information of the spectra into neutral losses. Then, an internal fragment matrix was constructed containing one row for each compound and one column for each neutral loss as described in [47]. The resulting sparse fragment matrix was assembled from the consensus spectra with an absolute m/z error of 0.01 Dalton and a relative m/z error of 20 ppm. In order to guarantee a minimum level of representativeness, metabolite families were restricted to (1) contain at least ten different fragment spectra; (2) selected subtrees must have at least one fragment present in at least 75% of the spectra within the cluster; (3) each cluster must contain at least five compounds; and (4) when criterion 2 was not fulfilled for all other clusters which contained the current cluster. Consensus spectra were learned from the common features of all spectra belonging to the remaining metabolite families (where each metabolite family was equivalent to a compound class as defined by a unique ChemOnt term and represented by a particular cluster in the hierarchical clustering tree). Finally, a binary classifier was constructed for each of the metabolite families. A *p*-value was calculated using the background scores distribution in accordance to [47]. A stratified repeated subset validation procedure was employed with ten data set compositions with 70% training data set and 30% test data set. The MetFamily classifier obtained consensus spectra for a total of 363 compound classes in negative mode and 451 compound classes in positive mode.

When the learned MetFamily classifier is applied to an unknown spectrum, it provides scores for the primary compound class and scores for alternative parents in case it detected spectral features, which are also present in other classes. The usage and methodology is further explained further in [42,47].

The MetFamily classifier was applied on any spectra present in each sample with the following parameters: msms.intensity.threshold = 10, minimumIntensityOfMaximalMS2-peak = 0.01%, minimumProportionOfMS2peaks = 0.05, mzDeviationAbsolute_grouping = 0.01, mzDeviationInPPM_grouping = 25, doPrecursorDeisotoping = TRUE, mzDeviationAbsolute_precursorDeisotoping = 0.001, mzDeviationInPPM_precursorDeisotoping = 10, maximumRtDifference = 0.02, doMs2PeakGroupDeisotoping = TRUE, mzDeviationAbsolute_ms2PeakGroupDeisotoping = 0.01, mzDeviationInPPM_ms2PeakGroupDeisotoping = 10, proportionOfMatchingPeaks_ms2PeakGroupDeisotoping = 0.9, mzDeviationAbsolute_mapping = 0.01, minimumNumberOfMS2PeaksPerGroup = 1, neutralLossesPrecursorToFragments = TRUE, neutralLossesFragmentsToFragments = FALSE. The classification scores of primary compound class and alternative parents were saved for all compounds present in samples in separate matrices for positive and negative ion modes. Tables containing the classification performances on the metabolite families present in the samples are available in the Supplementary Material as Tables S1 and S2. For the subsequent analyses, only the information on the primary compound class were used.

4.8. Statistical and Diversity Analyses

Statistical analyses were carried out with R 4.0.2 and the following additional packages: BiocParallel, doParallel, RColorBrewer, multtest, MSnbase, vegan, multcomp, Hmisc, gplots, lme4, lmerTest, glmnet, VennDiagram, dtwclust, pROC, PRROC, mltest, multiROC, ape, dendextend, cba and phangorn. Prior to the statistical analyses, peak and classification tables were log transformed to reach a semi-normal distribution. Missing values were imputed with zeros. No other normalization on the data were performed.

To get an overview on the influence of study factors and separation of samples, Principal Component Analysis (PCA) was carried out using the function prcomp. Variance partitioning from the package vegan (function varpart) was calculated for the study factors species and herbarium to evaluate the amount of variation explained in the data tables for the study factors [108]. Variable selection of metabolite features that were related to either particular species' or herbarium conditions was performed using Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) using the function splsda from the mixOmics

package [109]. As with discriminant analysis the number of components relate to the number of interactions of factor levels and the explained variations of the factor levels scatter on the corresponding component axes, the parameter ncomp was chosen 2 for herbarium, and 10 for species, respectively. A vector of variables to keep (parameter keepx) was constructed using the amount of variation explained from the variance partitioning multiplied by the number of total metabolite features in the data table. A heatmap was constructed with the selected variables using the function cim of the package mixOmics and variables extracted that correspond to the levels of either species or herbarium. In order to evaluate the performance of the fitted model, M-fold cross-validation with 10 folds using the function perf was performed and the following additional performance measures were calculated: multi-class area under curve (AUC) (function ml_test of the package mltest), the R^2 of the fitted vs. the entire model, and Receiver Operating Characteristic (ROC) and PR (Precision and Recall) curves using the functions plot.roc and ci.se from the pROC package and the function pr.curve from the PRROC package [110–113]. The area under curve (AUC) was calculated from the ROC and the area under precision recall curve (AUC-PR) from the PR curve, respectively. Summaries of model performances are available in Supplementary Material.

Diversity analyses were applied on the data of the metabolite fingerprinting (peak table) and on the compositional data of the classification approach (classification table). Analyses were performed calculating the compound richness (sum of compounds per each sample), the Shannon diversity index (H') on compounds in each sample and the Pielou's evenness (J) based on the H' divided by the log transform of the richness [114,115]. The following Equation (1) was used to calculate the Shannon diversity index H':

$$H' = \sum_{i=1}^{t} p_i \ \ln(p_i) \tag{1}$$

The Pielou's evenness J was calculated according to Equation (2):

$$J = \frac{H'}{\log S} \tag{2}$$

where the relative species/compound abundance p_i is calculated as following: $p_i = \frac{N_i}{N}$; N_i is the abundance of the i-th species/compound and N is the total abundance of all species/compounds; and S is the species richness. These diversity measures can be used as sample sizes were equal between levels [116].

Unique compounds were determined when a compound was present only in one species and not the others, and for fresh vs. dry, respectively. To compare abundances between the different species and the herbarium conditions, boxplots were generated using the function boxplot in R. The Tukey honestly significant difference (HSD) test on a one-way ANOVA was performed post hoc using the multcomp package to test the factor levels for significant differences.

Sunburst plots were constructed from the compositional data tables mainly using the functions draw.sector and arctext from the packages circlize, plotrix, and gplots [117]. These plots show an overview of the richness of compounds with regard to the ontology of the corresponding compound classes. In order to compare sunburst plots and find significant differences, the Fisher's exact test was used creating contingency tables for all compound classes separately using the function fisher.test in R [118,119]. *p*-values below 0.005 were treated as significantly. A simple barplot was constructed for the significant classes showing the enrichment or reduction of their relative abundances. The frequency of the relative abundances f was calculated using Equation (3), where k is the respective compound class and n the total number of classified compound classes:

$$f_k = \frac{dry_k}{\sum_{i=1}^{n} dry_i} - \frac{fresh_k}{\sum_{i=1}^{n} fresh_i} \tag{3}$$

A phylogenetic tree based on taxonomic information of bryophytes was extracted from the Open Tree of Life project with the identifier opentree12.3@mrcaott541ott1066 [120]. The supertree was pruned to only contain phylogenetic information for the used species in the study using the R package ape. Dendrograms using the chemotaxonomic information of the feature and the classification tables were created separately by constructing a cophenetic distance matrix using Bray-Curtis dissimilarity and hierarchical clustering was performed using Ward.D as agglomeration method. The two topmost nodes containing the acrocarpous and pleurocarpous groups in the dendrograms were swabbed using the function reorder of the package dendextend to improve visual representation. In order to compare the phylogenetic tree with the chemotaxonomic information of the metabolite fingerprinting and the classification, Procrustes analysis and the Mantel test (r) were performed together with calculating Pearson's correlation (c) and the Robinson–Foulds metric (rf).

In order to investigate the relation of ecological characteristics with the peak and classification tables, distance-based redundancy analysis (dbRDA) using the function dbrda of the package vegan was carried out. The ecological characteristics described in [60,61] were used. Bray-Curtis was chosen as distance measure. Ordinal and categorical ecological characteristics were transformed to the presence–absence matrices for the ordination. Significant characteristics were chosen using the function ordistep with forward variable selection. The model performance was evaluated using the goodness of fit statistic (squared correlation coefficient) by applying the function envfit post-hoc on the dbRDA ordination. Similarly, using the envfit function, arrows were fitted on the ordination plot where the length of the arrow is proportional to the correlation between species ordination and the respective characteristics.

4.9. Data Records

Raw data and meta-data were uploaded to the data repository MetaboLights and are available under the identifier MTBLS851. The vignettes and data to recreate the plots are available in MetaboLights (https://www.ebi.ac.uk/metabolights/MTBLS851, accessed on 22 March 2021) and in Zenodo (https://zenodo.org/record/4627092, doi: 10.5281/zenodo.4627092, accessed on 22 March 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

acrocarpous	Phylogenetic clade of bryophytes having mostly an erect stem and sporophytes originate at the top of stems
ANOVA	Analysis of Variance
AUC	Area Under receiver operating characteristic Curve
AUC-PR	Area Under Precision Recall Curve
Bar	Species code for *Barbula unguiculata* Hedw., ott1027526
Bra	Species code for *Brachythecium rutabulum* (Hedw.) Schimp., ott734784
Bryophytina	Subdivision of plants on which the phylogenetic tree was constructed, ott471195
Cal	Species code for *Calliergonella cuspidata* (Hedw.) Loeske, ott405418
dbRDA	Distance-based Redundancy Analysis
DIA	Data Independent Acquisition
dry	Factor level: Samples which were taken under herbarium conditions
fresh	Factor level: Samples which were taken under fresh conditions
Gri	Species code for *Grimmia pulvinata* (Hedw.) Sm., ott604969
Hyp	Species code for *Hypnum cupressiforme* Hedw., ott160023
LC/MS	Liquid chromatography coupled to a Mass Spectrometer
Mar	Species code for *Marchantia polymorpha* L., ott56596
MS1	Mass-spectral analysis
MS/MS	Tandem mass-spectral analysis where two MS are performed resulting in fragment spectra
ott	Phylogenetic identifier for species and phylogenetic groups
PCA	Principal Component Analysis
Pla	Species code for *Plagiomnium undulatum* (Hedw.) T.J. Kop., ott744339
pleurocarpous	Phylogenetic clade of bryophytes having mostly a branched prostrate stem and sporophytes are produced in branches
Pol	Species code for *Polytrichum strictum* Menzies ex Brid., ott101288
PR	Precision and Recall
R	R statistical programming language
R^2	R-squared
Rhy	Species code for *Rhytidiadelphus squarrosus* (Hedw.) Warnst., ott734791
ROC	Receiver Operating Characteristic
sPLS-DA	Sparse Partial Least Squares Discriminant Analysis
SWATH	Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra
Tor	Species code for *Tortula muralis* Hedw., ott318814
Tukey HSD	Honestly significant difference test
XCMS	Framework to perform detection of chromatographic peaks

References

1. Sardans, J.; Peñuelas, J.; Rivas-Ubach, A. Ecological metabolomics: Overview of current developments and future challenges. *Chemoecology* **2011**, *21*, 191–225. [CrossRef]
2. Van Dam, N.M.; van der Meijden, E. A Role for Metabolomics in Plant Ecology. In *Annual Plant Reviews Volume 43*; Hall, R.D., Ed.; Wiley-Blackwell: Oxford, UK, 2011; pp. 87–107, ISBN 978-1-4443-3995-6.
3. *Metabolomics Tools for Natural Product Discovery*; Roessner, U., Dias, D.A., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2013; Volume 1055, ISBN 978-1-62703-576-7.
4. Peters, K.; Worrich, A.; Weinhold, A.; Alka, O.; Balcke, G.; Birkemeyer, C.; Bruelheide, H.; Calf, O.; Dietz, S.; Dührkop, K.; et al. Current Challenges in Plant Eco-Metabolomics. *Int. J. Mol. Sci.* **2018**, *19*, 1385. [CrossRef]
5. Fiehn, O. Metabolomics—The link between genotypes and phenotypes. In *Functional Genomics*; Town, C., Ed.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 155–171, ISBN 978-94-010-3903-1.
6. Uthe, H.; van Dam, N.M.; Hervé, M.R.; Sorokina, M.; Peters, K.; Weinhold, A. A practical guide to implementing metabolomics in plant ecology and biodiversity research. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2020; p. S0065229620300732.
7. Goffinet, B.; Shaw, A.J. *Bryophyte Biology*; Cambridge University Press: Cambridge, UK, 2009; ISBN 978-0-511-45773-9.
8. Shi, X.-M.; Song, L.; Liu, W.-Y.; Lu, H.-Z.; Qi, J.-H.; Li, S.; Chen, X.; Wu, J.-F.; Liu, S.; Wu, C.-S. Epiphytic bryophytes as bio-indicators of atmospheric nitrogen deposition in a subtropical montane cloud forest: Response patterns, mechanism, and critical load. *Environ. Pollut.* **2017**, *229*, 932–941. [CrossRef] [PubMed]
9. Bennett, J. Bioindicators & Biomonitors, Principles, Concepts and Applications. *Sci. Total Environ.* **2004**, *328*, 295. [CrossRef]

10. Roos, R.E.; Zuijlen, K.; Birkemoe, T.; Klanderud, K.; Lang, S.I.; Bokhorst, S.; Wardle, D.A.; Asplund, J. Contrasting drivers of community-level trait variation for vascular plants, lichens and bryophytes across an elevational gradient. *Funct. Ecol.* **2019**, *33*, 2430–2446. [CrossRef]

11. Cornelissen, J.H.C.; Lang, S.I.; Soudzilovskaia, N.A.; During, H.J. Comparative Cryptogam Ecology: A Review of Bryophyte and Lichen Traits that Drive Biogeochemistry. *Ann. Bot.* **2007**, *99*, 987–1001. [CrossRef] [PubMed]

12. Streitberger, M.; Schmidt, C.; Fartmann, T. Contrasting response of vascular plant and bryophyte species assemblages to a soil-disturbing ecosystem engineer in calcareous grasslands. *Ecol. Eng.* **2017**, *99*, 391–399. [CrossRef]

13. Wulf, P.; Pearson, R.G. Mossy stones gather more bugs: Moss as habitat, nurseries and refugia for tropical stream invertebrates. *Hydrobiologia* **2017**, *790*, 167–182. [CrossRef]

14. Vanderpoorten, A.; Goffinet, B. *Introduction to Bryophytes*; Cambridge University Press: Cambridge, UK, 2009; ISBN 978-0-511-65080-2.

15. Müller, C.; Caspers, B.A.; Gadau, J.; Kaiser, S. The Power of Infochemicals in Mediating Individualized Niches. *Trends Ecol. Evol.* **2020**, *35*, 981–989. [CrossRef] [PubMed]

16. Cronberg, N. Microarthropods Mediate Sperm Transfer in Mosses. *Science* **2006**, *313*, 1255. [CrossRef] [PubMed]

17. Rosenstiel, T.N.; Shortlidge, E.E.; Melnychenko, A.N.; Pankow, J.F.; Eppley, S.M. Sex-specific volatile compounds influence microarthropod-mediated fertilization of moss. *Nature* **2012**, *489*, 431–433. [CrossRef] [PubMed]

18. Asakawa, Y.; Ludwiczuk, A.; Nagashima, F. Phytochemical and biological studies of bryophytes. *Phytochemistry* **2013**, *91*, 52–80. [CrossRef] [PubMed]

19. *Chemical Constituents of Bryophytes: Bio- and Chemical Diversity, Biological Activity, and Chemosystematics*; Asakawa, Y., Ludwiczuk, A., Nagashima, F., (Eds.); Fortschritte der Chemie organischer Naturstoffe = Progress in the chemistry of organic natural products; Springer: Berlin/Heidelberg, Germany, 2013; ISBN 978-3-7091-1083-6.

20. Bristol, B.M. On the remarkable retention of vitality of moss protonema. *New Phytol.* **1916**, *15*, 137–143. [CrossRef]

21. Maheu, M.J. Régénération du *Barbula muralis*, après quatorze ans de sécheresse, par protonémas foliaires primaires propagulifères et protonémas secondaires bulbigènes. *Bull. Société Bot. Fr.* **1922**, *69*, 330–334. [CrossRef]

22. Breuil-Sée, A. Recorded desiccation-survival times in bryophytes. *J. Bryol.* **1993**, *17*, 679–680. [CrossRef]

23. Oliver, M.J.; Tuba, Z.; Mishler, B.D. The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol.* **2000**, *151*, 85–100. [CrossRef]

24. Oliver, M.J. Desiccation Tolerance in Bryophytes: A Reflection of the Primitive Strategy for Plant Survival in Dehydrating Habitats? *Integr. Comp. Biol.* **2005**, *45*, 788–799. [CrossRef]

25. Garcia, E.L.; Rosenstiel, T.N.; Graves, C.; Shortlidge, E.E.; Eppley, S.M. Distribution drivers and physiological responses in geothermal bryophyte communities. *Am. J. Bot.* **2016**, *103*, 625–634. [CrossRef] [PubMed]

26. Xu, B.; Ohtani, M.; Yamaguchi, M.; Toyooka, K.; Wakazaki, M.; Sato, M.; Kubo, M.; Nakano, Y.; Sano, R.; Hiwatashi, Y.; et al. Contribution of NAC Transcription Factors to Plant Adaptation to Land. *Science* **2014**, *343*, 1505–1508. [CrossRef]

27. Niklas, K.J.; Cobb, E.D.; Matas, A.J. The evolution of hydrophobic cell wall biopolymers: From algae to angiosperms. *J. Exp. Bot.* **2017**, *68*, 5261–5269. [CrossRef] [PubMed]

28. Proctor, M.C.F.; Smirnoff, N. Ecophysiology of photosynthesis in bryophytes: Major roles for oxygen photoreduction and non-photochemical quenching? *Physiol. Plant.* **2011**, *141*, 130–140. [CrossRef] [PubMed]

29. Asakawa, Y.; Ludwiczuk, A. Chemical Constituents of Bryophytes: Structures and Biological Activity. *J. Nat. Prod.* **2018**, *81*, 641–660. [CrossRef] [PubMed]

30. Charron, A.J.; Quatrano, R.S. Between a Rock and a Dry Place: The Water-Stressed Moss. *Mol. Plant* **2009**, *2*, 478–486. [CrossRef] [PubMed]

31. Proctor, M.C.F.; Oliver, M.J.; Wood, A.J.; Alpert, P.; Stark, L.R.; Cleavitt, N.L.; Mishler, B.D. Desiccation-tolerance in bryophytes: A review. *Bryologist* **2007**, *110*, 595–621. [CrossRef]

32. Wang, R.; Yin, Y.; Zhu, Z.-J. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Anal. Bioanal. Chem.* **2019**, *411*, 4349–4357. [CrossRef] [PubMed]

33. Hopfgartner, G.; Tonoli, D.; Varesio, E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal. Bioanal. Chem.* **2012**, *402*, 2587–2596. [CrossRef] [PubMed]

34. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526. [CrossRef] [PubMed]

35. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779–787. [CrossRef] [PubMed]

36. Tsugawa, H.; Nakabayashi, R.; Mori, T.; Yamada, Y.; Takahashi, M.; Rai, A.; Sugiyama, R.; Yamamoto, H.; Nakaya, T.; Yamazaki, M.; et al. A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods* **2019**, *16*, 295–298. [CrossRef] [PubMed]

37. Bonner, R.; Hopfgartner, G. SWATH data independent acquisition mass spectrometry for metabolomics. *TrAC Trends Anal. Chem.* **2019**, *120*, 115278. [CrossRef]

38. Zhu, X.; Chen, Y.; Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MS [All] Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography–Quadrupole Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 1202–1209. [CrossRef]

39. Balcke, G.U.; Bennewitz, S.; Bergau, N.; Athmer, B.; Henning, A.; Majovsky, P.; Jiménez-Gómez, J.M.; Hoehenwarter, W.; Tissier, A. Multi-Omics of Tomato Glandular Trichomes Reveals Distinct Features of Central Carbon Metabolism Supporting High Productivity of Specialized Metabolites. *Plant Cell* **2017**, *29*, 960–983. [CrossRef]

40. Van Dam, N.M.; Bouwmeester, H.J. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. *Trends Plant Sci.* **2016**, *21*, 256–265. [CrossRef] [PubMed]

41. Wishart, D.S. Computational strategies for metabolite identification in metabolomics. *Bioanalysis* **2009**, *1*, 1579–1596. [CrossRef] [PubMed]

42. Peters, K.; Treutler, H.; Döll, S.; Kindt, A.S.; Hankemeier, T.; Neumann, S. Neumann Chemical Diversity and Classification of Secondary Metabolites in Nine Bryophyte Species. *Metabolites* **2019**, *9*, 222. [CrossRef] [PubMed]

43. Da Silva, R.R.; Dorrestein, P.C.; Quinn, R.A. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12549–12550. [CrossRef] [PubMed]

44. Ruttkies, C.; Schymanski, E.L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* **2016**, *8*. [CrossRef] [PubMed]

45. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12580–12585. [CrossRef] [PubMed]

46. Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J. Metabolite fingerprinting: Detecting biological features by independent component analysis. *Bioinformatics* **2004**, *20*, 2447–2454. [CrossRef] [PubMed]

47. Treutler, H.; Tsugawa, H.; Porzel, A.; Gorzolka, K.; Tissier, A.; Neumann, S.; Balcke, G.U. Discovering Regulated Metabolite Families in Untargeted Metabolomics Studies. *Anal. Chem.* **2016**, *88*, 8082–8090. [CrossRef]

48. Rogers, S.; Ong, C.W.; Wandy, J.; Ernst, M.; Ridder, L.; van der Hooft, J.J.J. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. *Faraday Discuss.* **2019**, *218*, 284–302. [CrossRef] [PubMed]

49. Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2020**. [CrossRef]

50. Van der Hooft, J.J.J.; Wandy, J.; Barrett, M.P.; Burgess, K.E.V.; Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13738–13743. [CrossRef]

51. Barbas, C.; Moraes, E.P.; Villaseñor, A. Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. *J. Pharm. Biomed. Anal.* **2011**, *55*, 823–831. [CrossRef] [PubMed]

52. Hastings, J.; Chepelev, L.; Willighagen, E.; Adams, N.; Steinbeck, C.; Dumontier, M. The Chemical Information Ontology: Provenance and Disambiguation for Chemical Data on the Biological Semantic Web. *PLoS ONE* **2011**, *6*, e25513. [CrossRef]

53. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **2016**, *8*. [CrossRef] [PubMed]

54. Van der Hooft, J.J.J.; Wandy, J.; Young, F.; Padmanabhan, S.; Gerasimidis, K.; Burgess, K.E.V.; Barrett, M.P.; Rogers, S. Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics. *Anal. Chem.* **2017**, *89*, 7569–7577. [CrossRef]

55. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [CrossRef] [PubMed]

56. Edmands, W.M.B.; Petrick, L.; Barupal, D.K.; Scalbert, A.; Wilson, M.J.; Wickliffe, J.K.; Rappaport, S.M. compMS2Miner: An Automatable Metabolite Identification, Visualization, and Data-Sharing R Package for High-Resolution LC–MS Data Sets. *Anal. Chem.* **2017**, *89*, 3919–3928. [CrossRef] [PubMed]

57. Blaženović, I.; Kind, T.; Sa, M.R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B.S.; Torbašinović, H.; Lee, T.; Mehta, S.S.; et al. Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal. Chem.* **2019**, *91*, 2155–2162. [CrossRef] [PubMed]

58. Tripathi, A.; Vázquez-Baeza, Y.; Gauglitz, J.M.; Wang, M.; Dührkop, K.; Nothias-Esposito, M.; Acharya, D.D.; Ernst, M.; van der Hooft, J.J.J.; Zhu, Q.; et al. Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat. Chem. Biol.* **2021**, *17*, 146–151. [CrossRef] [PubMed]

59. Haug, K.; Salek, R.M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; et al. MetaboLights—An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **2013**, *41*, D781–D786. [CrossRef]

60. Peters, K.; Gorzolka, K.; Bruelheide, H.; Neumann, S. Seasonal variation of secondary metabolites in nine different bryophytes. *Ecol. Evol.* **2018**, *8*, 9105–9117. [CrossRef] [PubMed]

61. Peters, K.; Gorzolka, K.; Bruelheide, H.; Neumann, S. Computational workflow to study the seasonal variation of secondary metabolites in nine different bryophytes. *Sci. Data* **2018**, *5*, 180179. [CrossRef]

62. Schymanski, E.L.; Ruttkies, C.; Krauss, M.; Brouard, C.; Kind, T.; Dührkop, K.; Allen, F.; Vaniya, A.; Verdegem, D.; Böcker, S.; et al. Critical Assessment of Small Molecule Identification 2016: Automated methods. *J. Cheminform.* **2017**, *9*, 22. [CrossRef]

63. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714. [CrossRef] [PubMed]

64. Shahaf, N.; Rogachev, I.; Heinig, U.; Meir, S.; Malitsky, S.; Battat, M.; Wyner, H.; Zheng, S.; Wehrens, R.; Aharoni, A. The WEIZMASS spectral library for high-confidence metabolite identification. *Nat. Commun.* **2016**, *7*, 12423. [CrossRef] [PubMed]

65. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. [CrossRef]

66. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 299–302. [CrossRef]

67. Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D.S. CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* **2019**, *9*, 72. [CrossRef]

68. Maksimova, V.; Klavina, L.; Bikovens, O.; Zicmanis, A.; Purmalis, O. Structural Characterization and Chemical Classification of Some Bryophytes Found in Latvia. *Chem. Biodivers.* **2013**, *10*, 1284–1294. [CrossRef] [PubMed]

69. Klavina, L. A study on bryophyte chemical composition–search for new applications. *Agron. Res.* **2015**, *13*, 969–978.

70. Sardans, J.; Gargallo-Garriga, A.; Urban, O.; Klem, K.; Walker, T.W.N.; Holub, P.; Janssens, I.A.; Peñuelas, J. Ecometabolomics for a Better Understanding of Plant Responses and Acclimation to Abiotic Factors Linked to Global Change. *Metabolites* **2020**, *10*, 239. [CrossRef] [PubMed]

71. Araújo, W.L.; Tohge, T.; Ishizaki, K.; Leaver, C.J.; Fernie, A.R. Protein degradation—An alternative respiratory substrate for stressed plants. *Trends Plant Sci.* **2011**, S1360138511001063. [CrossRef]

72. Rico-Reséndiz, F.; Cervantes-Pérez, S.A.; Espinal-Centeno, A.; Dipp-Álvarez, M.; Oropeza-Aburto, A.; Hurtado-Bautista, E.; Cruz-Hernández, A.; Bowman, J.L.; Ishizaki, K.; Arteaga-Vázquez, M.A.; et al. Transcriptional and morpho-physiological responses of *Marchantia polymorpha* upon phosphate starvation. *Int. J. Mol. Sci.* **2020**, *21*, 8354. [CrossRef] [PubMed]

73. Kranner, I.; Beckett, R.; Hochman, A. Desiccation-tolerance in lichens: A review. *Bryologist* **2008**, *111*, 576–593. [CrossRef]

74. Erxleben, A.; Gessler, A.; Vervliet-Scheebaum, M.; Reski, R. Metabolite profiling of the moss Physcomitrella patens reveals evolutionary conservation of osmoprotective substances. *Plant Cell Rep.* **2012**, *31*, 427–436. [CrossRef] [PubMed]

75. Brunetti, C.; George, R.M.; Tattini, M.; Field, K.; Davey, M.P. Metabolomics in plant environmental physiology. *J. Exp. Bot.* **2013**, *64*, 4011–4020. [CrossRef]

76. Yang, L.; Wen, K.S.; Ruan, X.; Zhao, Y.X.; Wei, F.; Wang, Q. Response of Plant Secondary Metabolites to Environmental Factors. *Molecules* **2018**, *23*, 762. [CrossRef] [PubMed]

77. Davies, K.M.; Jibran, R.; Zhou, Y.; Albert, N.W.; Brummell, D.A.; Jordan, B.R.; Bowman, J.L.; Schwinn, K.E. The Evolution of Flavonoid Biosynthesis: A Bryophyte Perspective. *Front. Plant Sci.* **2020**, *11*, 7. [CrossRef] [PubMed]

78. Kumar, S.; Pandey, A.K. Chemistry and Biological Activities of Flavonoids: An Overview. *Sci. World J.* **2013**, *2013*, 1–16. [CrossRef] [PubMed]

79. Lambers, H.; Chapin, F.S.; Pons, T.L. *Plant Physiological Ecology*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2008; ISBN 978-0-387-78340-6.

80. Chopra, R.N.; Bhatla, S.C.; (Eds.) *Bryophyte Development: Physiology and Biochemistry*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2019; ISBN 978-0-429-26056-8.

81. Mithöfer, A.; Boland, W. Plant Defense against Herbivores: Chemical Aspects. *Annu. Rev. Plant Biol.* **2012**, *63*, 431–450. [CrossRef] [PubMed]

82. Nandy, S.; Dey, A. Bibenzyls and bisbybenzyls of bryophytic origin as promising source of novel therapeutics: Pharmacology, synthesis and structure-activity. *DARU J. Pharm. Sci.* **2020**. [CrossRef]

83. Tanaka, M.; Esaki, T.; Kenmoku, H.; Koeduka, T.; Kiyoyama, Y.; Masujima, T.; Asakawa, Y.; Matsui, K. Direct evidence of specific localization of sesquiterpenes and marchantin A in oil body cells of Marchantia polymorpha L. *Phytochemistry* **2016**, *130*, 77–84. [CrossRef]

84. Xie, C.-F.; Lou, H.-X. Secondary Metabolites in Bryophytes: An Ecological Aspect. *Chem. Biodivers.* **2009**, *6*, 303–312. [CrossRef]

85. Commisso, M.; Guarino, F.; Marchi, L.; Muto, A.; Piro, A.; Degola, F. Bryo-Activities: A Review on How Bryophytes Are Contributing to the Arsenal of Natural Bioactive Compounds against Fungi. *Plants* **2021**, *10*, 203. [CrossRef] [PubMed]

86. Lang, S.I.; Cornelissen, J.H.C.; Klahn, T.; van Logtestijn, R.S.P.; Broekman, R.; Schweikert, W.; Aerts, R. An experimental comparison of chemical traits and litter decomposition rates in a diverse range of subarctic bryophyte, lichen and vascular plant species. *J. Ecol.* **2009**, *97*, 886–900. [CrossRef]

87. Wang, Z.; Liu, X.; Bao, W. Higher photosynthetic capacity and different functional trait scaling relationships in erect bryophytes compared with prostrate species. *Oecologia* **2016**, *180*, 359–369. [CrossRef]

88. Züst, T.; Agrawal, A.A. Trade-Offs Between Plant Growth and Defense against Insect Herbivory: An Emerging Mechanistic Synthesis. *Annu. Rev. Plant Biol.* **2017**, *68*, 513–534. [CrossRef] [PubMed]

89. Vieira, C.; Séneca, A.; Sérgio, C.; Ferreira, M.T. Bryophyte taxonomic and functional groups as indicators of fine scale ecological gradients in mountain streams. *Ecol. Indic.* **2012**, *18*, 98–107. [CrossRef]

90. Díaz, S.; Kattge, J.; Cornelissen, J.H.C.; Wright, I.J.; Lavorel, S.; Dray, S.; Reu, B.; Kleyer, M.; Wirth, C.; Colin Prentice, I.; et al. The global spectrum of plant form and function. *Nature* **2016**, *529*, 167–171. [CrossRef]

91. Asner, G.P.; Anderson, C.B.; Martin, R.E.; Tupayachi, R.; Knapp, D.E.; Sinca, F. Landscape biogeochemistry reflected in shifting distributions of chemical traits in the Amazon forest canopy. *Nat. Geosci.* **2015**, *8*, 567–573. [CrossRef]

92. Tsunoda, T.; van Dam, N.M. Root chemical traits and their roles in belowground biotic interactions. *Pedobiologia* **2017**, *65*, 58–67. [CrossRef]

93. Descombes, P.; Kergunteuil, A.; Glauser, G.; Rasmann, S.; Pellissier, L. Plant physical and chemical traits associated with herbivory in situ and under a warming treatment. *J. Ecol.* **2020**, *108*, 733–749. [CrossRef]

94. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]

95. Ludwiczuk, A.; Odrzykoski, I.J.; Asakawa, Y. Identification of cryptic species within liverwort Conocephalum conicum based on the volatile components. *Phytochemistry* **2013**, *95*, 234–241. [CrossRef] [PubMed]

96. Kohn, G.; Demmerle, S.; Vandekerkhove, O.; Hartmann, E.; Beutelmann, P. Distribution and chemotaxonomic significance of acetylenic fatty acids in mosses of the dicranales. *Phytochemistry* **1987**, *26*, 2271–2275. [CrossRef]

97. Brodo, I.M. Interpreting Chemical Variation in Lichens for Systematic Purposes. *Bryologist* **1986**, *89*, 132. [CrossRef]

98. Culberson, W.L. The use of chemistry in the systematics of the lichens. *TAXON* **1969**, *18*, 152–166. [CrossRef]

99. Eisenreich, W.; Knispel, N.; Beck, A. Advanced methods for the study of the chemistry and the metabolism of lichens. *Phytochem. Rev.* **2011**, *10*, 445–456. [CrossRef]

100. Lumbsch, H.T. Analysis of Phenolic Products in Lichens for Identification and Taxonomy. In *Protocols in Lichenology*; Kranner, I.C., Beckett, R.P., Varma, A.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 281–295. ISBN 978-3-540-41139-0.

101. Liu, X.; Li, N.; Liu, S.; Wang, J.; Zhang, N.; Zheng, X.; Leung, K.-S.; Cheng, L. Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Front. Bioeng. Biotechnol.* **2019**, *7*, 358. [CrossRef] [PubMed]

102. Risso, D.; Massa, M.S.; Chiogna, M.; Romualdi, C. A modified LOESS normalization applied to microRNA arrays: A comparative evaluation. *Bioinformatics* **2009**, *25*, 2685–2691. [CrossRef]

103. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]

104. Balcke, G.U.; Bennewitz, S.; Zabel, S.; Tissier, A. Isoprenoid and Metabolite Profiling of Plant Trichomes. In *Plant Isoprenoids*; Rodríguez-Concepción, M., Ed.; Methods in Molecular Biology; Springer: Berlin/Heidelberg, Germany, 2014; Volume 1153, pp. 189–202, ISBN 978-1-4939-0605-5.

105. Adusumilli, R.; Mallick, P. Data Conversion with ProteoWizard msConvert. In *Proteomics*; Comai, L., Katz, J.E., Mallick, P., Eds.; Methods in Molecular Biology; Springer: Berlin/Heidelberg, Germany, 2017; Volume 1550, pp. 339–368. ISBN 978-1-4939-6745-2.

106. Sansone, S.-A.; Rocca-Serra, P.; Field, D.; Maguire, E.; Taylor, C.; Hofmann, O.; Fang, H.; Neumann, S.; Tong, W.; Amaral-Zettler, L.; et al. Toward interoperable bioscience data. *Nat. Genet.* **2012**, *44*, 121–126. [CrossRef]

107. Spicer, R.A.; Salek, R.; Steinbeck, C. Compliance with minimum information guidelines in public metabolomics repositories. *Sci. Data* **2017**, *4*, 170137. [CrossRef] [PubMed]

108. Borcard, D.; Legendre, P.; Drapeau, P. Partialling out the Spatial Component of Ecological Variation. *Ecology* **1992**, *73*, 1045–1055. [CrossRef]

109. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. mixOmics: An R package for omics feature selection and multiple data integration. *PLOS Comput. Biol.* **2017**, *13*, e1005752. [CrossRef] [PubMed]

110. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [CrossRef]

111. Grau, J.; Grosse, I.; Keilwagen, J. PRROC: Computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* **2015**, *31*, 2595–2597. [CrossRef]

112. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

113. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2020**. [CrossRef]

114. Hurlbert, S.H. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* **1971**, *52*, 577–586. [CrossRef] [PubMed]

115. Heip, C.H.R.; Herman, P.M.J.; Soetaert, K. Indices of diversity and evenness. *Océanis* **1998**, *24*, 61–87.

116. McCune, B.; Grace, J.B.; Urban, D.L. *Analysis of Ecological Communities*, 2nd ed.; MjM Software Design: Gleneden Beach, OR, USA, 2002; ISBN 978-0-9721290-0-8.

117. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [CrossRef] [PubMed]

118. Fisher, R.A. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* **1922**, *85*, 87. [CrossRef]

119. Agresti, A. A Survey of Exact Inference for Contingency Tables. *Stat. Sci.* **1992**, *7*, 131–153. [CrossRef]

120. Redelings, B.D.; Holder, M.T. A supertree pipeline for summarizing phylogenetic and taxonomic information for millions of species. *PeerJ* **2017**, *5*, e3058. [CrossRef] [PubMed]