plant biology

RESEARCH ARTICLE

Ontogeny and organ-specific steroidal glycoside diversity is associated with differential expression of steroidal glycoside pathway genes in two *Solanum dulcamara* leaf chemotypes

R. A. Anaia[1,2,3] (iD), I. Chiocchio[4], R. Sontowski[1,2,5], B. Swinkels[1,3], F. Vergara[1,2] & N. M. van Dam[1,2,5]

1 Molecular Interaction Ecology, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
2 Institute of Biodiversity, Friedrich Schiller University, Jena, Germany
3 Plant and Animal Biology, Radboud Institute for Biological and Environmental Sciences, Radboud University, Nijmegen, the Netherlands
4 Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy
5 Leibniz Institute for Vegetable and Ornamental Crops (IGZ), Großbeeren, Germany

ABSTRACT

- Solanaceous plants, such as *Solanum dulcamara*, produce steroidal glycosides (SGs). Leaf SG profiles vary among *S. dulcamara* individuals, leading to distinct phytochemical phenotypes ('chemotypes') and intraspecific phytochemical diversity ('chemodiversity'). However, if and how SG chemodiversity varies among organs and across ontogeny, and how this relates to SG metabolism gene expression is unknown.
- Among organs and across ontogeny, *S. dulcamara* plants with saturated (S) and unsaturated (U) SG leaf chemotypes were selected and clonally propagated. Roots, stems and leaves were harvested from vegetative and flowering plants. Extracts were analysed using untargeted LC–MS. Expression of candidate genes in SG metabolism (*SdGAME9*, *SdGAME4*, *SdGAME25*, *SdS5αR2* and *SdDPS*) was analysed using RT-qPCRs.
- Our analyses showed that SG chemodiversity varies among organs and across ontogeny in *S. dulcamara*; SG richness (D_{mg}) was higher in flowering than vegetative plants. In vegetative plants, D_{mg} was higher for leaves than for roots.
- Lack of *SdGAME25* expression in *U*-chemotype leaves, while readily expressed in roots and stems, suggests a pivotal role for *SdGAME25* in differentiation of leaf chemotypes in vegetative and flowering plants. By acting as an ontogeny-dependent chemotypic switch, differential regulation of *SdGAME25* enables adaptive allocation of SGs, thereby increasing SG chemodiversity in leaves. This indicates that differential expression and/or regulation of glycoalkaloid metabolism genes, rather than their presence or absence, explains observed chemotypic variation in SG chemodiversity among organs and across ontogeny.

INTRODUCTION

Plants chemically defend themselves by employing plant-specialized metabolites (PSMs). Specifically, the *Solanum* produce PSMs called steroidal glycosides (SGs), which serve as chemical defence against herbivores (Calf *et al.* 2018) and pathogens (Sonawane *et al.* 2018). SGs consist of a steroidal aglycone (SA) that is conjugated to a glycoside moiety. This class includes steroidal saponin glycosides (SSGs) and their nitrogen-containing analogues, steroidal glycoalkaloids (SGAs). As a class, SGs are structurally highly diverse, among others, because of variations in saturation of the steroidal aglycone and types and numbers of sugar molecules in the glycoside chain (Zhao *et al.* 2021). *Solanum* species, and individuals within a species, may have distinct SG profiles. In *S. dulcamara*, structural variation in SGs is based on variations in the aglycone, and their decoration, which give rise to intraspecific 'phytochemical diversity' (hereafter 'chemodiversity'). These structural variations are introduced by chemical modifications,

e.g. hydroxylation, acetylation and glycosylation, of the aglycone. These modifications may result from spontaneous or enzyme-catalysed reactions (Wang *et al.* 2019). Furthermore, the steroidal aglycones may vary in number of double bonds and rings. This can be expressed as the ring-double bond equivalent (RDBE), which is equal to the sum of the number of rings and double bonds in a molecular system (Heinig & Aharoni 2014). For instance, spirostanes consist of six rings (A – F), while furostanes have five-membered ring system (A – E) in which the F-ring is opened into an alkyl chain.

Over the last decade, great advances have been made in identification and characterization of genes involved in biosynthesis of SGs, including SSGs (Cheng *et al.* 2023) and SGAs (Itkin *et al.* 2013; Cárdenas *et al.* 2015; Sonawane *et al.* 2020), as well as in that of their precursor, cholesterol (Itkin *et al.* 2013; Sonawane *et al.* 2017). The majority of genes involved in SGA production are clustered on two chromosomes, which are syntenic across *S. lycopersicum* and *S. tuberosum* (Itkin *et al.* 2013), and thus likely wild relatives. Genes involved in production of SGA

and SGG are commonly referred to as *GLYCOALKALOID METABOLISM* (*GAME*) genes. In *Solanum* spp., SGA biosynthesis is regulated by a transcription factor called Jasmonate-Responsive Ethylene Response Factor 4 (JRE4) or GAME9 (Cárdenas *et al.* 2016; Nakayasu *et al.* 2018). A key enzyme in SGA biosynthesis is GAME4, which catalyses the first dedicated step in SGA production (Itkin *et al.* 2013; Paudel *et al.* 2017). In *S. lycopersicum*, β-Hydroxysteroid Dehydrogenase/3-Ketosteroid Reductase (3βHSD1) or GAME25, and STEROID 5α-REDUCTASE2 (S5αR2) are both involved in reduction of the double bond between C5 and C6 in the B-ring of dehydrotomatidine to produce tomatidine (Akiyama *et al.* 2019; Lee *et al.* 2019; Sonawane *et al.* 2018). After glycosylation of steroidal aglycones, spirostanes, such as (dehydro)tomatine, may potentially be further transformed into solanidanes. Spirostanes and solanidanes are six-ringed steroids with distinct fusion patterns between their E and F rings. Spirostanes are fused by a single quaternary carbon atom (spiro carbon), whereas solanidanes are fused by a single covalent bond between a tertiary carbon and nitrogen atom, in an ortho fusion arrangement (Moss 1998). A 2-oxoglutarate-dependent dioxygenase, DIOXYGENASE FOR POTATO SOLANIDANE SYNTHESIS (DPS), catalyses C-16α hydroxyl of spirostanes, which is considered the first-dedicated step towards solanidane-type SGAs in *S. tuberosum* (Akiyama *et al.* 2021).

Consequently, the observed SG chemodiversity in *S. dulcamara* may be related to the absence or presence of particular genes involved in SG biosynthetic pathways or to differential expression patterns (Calf *et al.* 2019). Interestingly, differences in the relative presence of SGAs with saturated (*S*) or unsaturated (*U*) steroidal aglycones have been related to differences in gastropod preference in preference assays with leaf discs (Calf *et al.* 2018). In both preference assays and a common garden experiment, plants that predominantly produce *S*-type SGAs were more preferred by generalist slugs (*Deroceras reticulatum*) compared to *U*-type SGAs containing accessions (Calf *et al.* 2018, 2019). On the other hand, specialist flea beetles were more abundant on plants with *U*-type SGA profiles, and avoided plants rich in SSGs rather than SGAs (Calf *et al.* 2019). It was postulated that SG leaf chemotypes in *S. dulcamara* may be heritable (Schreiber & Rönsch 1965; Willuhn 1966; Calf 2019). This suggests that SG chemodiversity in *S. dulcamara* may be driven by differential selection pressures exerted by different herbivore communities (Wetzel & Whitehead 2020; Petrén *et al.* 2023a, 2023b; Thon *et al.* 2024).

Although SG chemodiversity in *S. dulcamara* leaves (Calf *et al.* 2018) and roots (Chiocchio *et al.* 2023) is well described, little is known about ontogenetic, organ- and chemotype-specific SG variation in *S. dulcamara*, especially in relation to expression of candidate genes in SG biosynthesis. It is known from other PSM and families, e.g. glucosinolates in Brassicaceae, that leaf and root profiles differ considerably within individual plants (van Leur *et al.* 2006; Tsunoda *et al.* 2017). onsequently, leaf glucosinolate profiles are more distinct than these of roots in two different *Barbarea vulgaris* leaf chemotypes (van Leur *et al.* 2008). Recently, we profiled SG chemodiversity in embryonic and adventitious roots of *S. dulcamara*, using liquid chromatography coupled to mass spectrometry (LC–MS). We found that both root types have distinct SG profiles (Chiocchio *et al.* 2023). This suggests that there may be

additional levels of intraspecific and intra-individual chemical diversity than in leaves. Based on these root analyses, we proposed a SG classification system for mass spectra, based on structural differences among steroidal aglycones (Chiocchio *et al.* 2023). Such classification systems allow studying SGs in terms of chemodiversity, since we can apply (species) diversity indices, such as Margalef's richness (Margalef 1958) and Pielou's evenness (Pielou 1966), to PSMs as previously suggested (Hilker 2014; Marion *et al.* 2015; Kessler & Kalske 2018; Wetzel & Whitehead 2020; Petrén *et al.* 2023a, 2023b; Thon *et al.* 2024). For SGs, chemodiversity can be measured in terms of chemical richness and evenness, by considering the existence of distinct 'steroidal aglycone species' and the number of associated glycosides per unique aglycone species. Quantifying chemical richness among organs and individual leaf chemotypes in conjunction with expression of relevant genes, enhances our understanding of how chemodiversity is regulated within plants as well as in ontogenetic development. Moreover, it provides new hypotheses on potential selection processes that have shaped evolution of different levels of chemodiversity.

Here, we analysed whether SG profiles are organ-specific in vegetative and flowering *S. dulcamara* full-sibs of two contrasting leaf chemotypes. We asked whether the SGA leaf chemotype (hereafter 'chemotype') is constant over plant organs and ontogeny. Furthermore, we investigated whether the previously defined SGA chemotypes are characterized by a broader SG diversity using chemical profiling across ontogenetic stages and organs within a plant. Thereafter, we asked whether the detected organ, ontogeny and chemotype-specific variation in SG chemodiversity is related to the differential expression of candidate genes in SG metabolism. To do so, we prepared an F1 population by crossing two Dutch *S. dulcamara* accessions: 'Zandvoort Dry' (ZD04) and 'Texel Wet' (TW12) described by Calf *et al.* (2018). Accession ZD04 produces unsaturated SGAs in leaves, while accession TW12 predominantly produces saturated SGAs. The F1 progeny (TW12 × ZD04) was chemotyped by LC–MS, after which siblings with *S*- or *U*-chemotypes were selected for further analyses. The selected plants were chemically profiled for SG chemodiversity in adventitious roots, stems and leaves of vegetative and flowering *S. dulcamara*. In the same tissues, we assessed expression of candidate genes in SG biosynthesis, including homologues of *GAME9* (Cárdenas *et al.* 2016), *GAME4* (Itkin *et al.* 2013; Paudel *et al.* 2017), *GAME25* (Lee *et al.* 2019; Sonawane *et al.* 2018). *S5αR2* (Akiyama *et al.* 2019) and *DPS* (Akiyama *et al.* 2021) using RT-qPCR. Together, these analyses provide new insights into regulation of chemotypic, organ- and ontogeny-specific SG chemodiversity in two main *S. dulcamara* chemotypes.

MATERIAL AND METHODS

Plant material, chemotyping and experimental design

Hybrid *S. dulcamara* (TW12 × ZD04) seeds were germinated as described in Chiocchio *et al.* (2023). Briefly, seeds were placed onto wet glass beads (1 mm Ø) in plastic boxes. Seeds were cold stratified in the dark at 4 °C for 2 weeks. Then, the seed-containing plastic boxes were put in a climate chamber (L:D 16 h:8 h, 20 °C day/17 °C night, with light of 500 µmol·m^{-2}·s^{-1}) to induce germination (Chiocchio

et al. 2023). Emerged seedlings with two similarly sized cotyledons were transplanted to trays (QuickPot™ 24R, Ø 7.5 × 7.0 cm; Groß Kreutz, Germany) filled with a 1:1 (v/v) autoclaved soil (Floradur B pot clay medium coarse; Floragard Vetriebs, Germany) and sand (0/2 washed; Rösl Rohstoffe, Germany) mixture. When seedlings had a second set of true leaves, leaf samples were taken for metabolite extraction and SG chemotyping, as described below. For chemotyping, extracted ion chromatograms (EICs) were produced for *m/z* 414.3 and *m/z* 416.3 and the resulting EICs were inspected. Plants were assigned to the saturated (*S*) SG chemotype when peak *m/z*-fragment 416.3 was present in the chromatograms, while plants were assigned to the unsaturated (*U*) SG chemotype when *m/z* 414.3 was observed in the absence of *m/z* 416.3. Based on available plant material, three *S*-chemo-genotypes and four *U*-chemo-genotypes were selected for further experimentation.

To generate sufficient plant material for experimentation, multiple stem cuttings were taken from the chemotyped stock plants, as described by Calf *et al.* (2018). Stem cuttings were potted (11 × 11 × 12 cm) in 1 l pots containing a well-watered 1:1 soil:sand mixture supplied with 4 g·l⁻¹ Osmocote Pro 8-9M (ICL Boulby, Cleveland, UK), and kept under greenhouse conditions (17–25 °C, RH: ±65%) with light supplemented to 280 µmol·m⁻²·s⁻¹ with high-pressure sodium lamps. Healthy, vigorously growing plants were selected for sampling 6 and 11 weeks after transplantation. At 6 weeks old, all plants were vegetative, with vegetative meristems and no inflorescences (vegetative stage), while 11-week-old plants were flowering, with open, pollen-producing flowers (flowering). Leaves and adventitious roots (hereafter referred to as roots) were harvested from vegetative and flowering plants. In addition, stems of flowering plants were harvested. This was not possible for vegetative plants, as the remaining stems were used to generate new plants by clonal propagation, as described above. At harvest, plants were carefully removed from their pots and remaining soil removed under running tap water. Then roots were washed using deionized water and gently tapped dry with paper tissues. Simultaneously, five fully-expanded leaves, counted from the first fully expanded (~2-cm wide) leaf from the stem apex (Viswanathan & Thaler 2004), were harvested using sharp scissors. The leaves were stacked and midveins were removed with scissors. The stem segment on which the first five fully expanded leaves were present was cut into multiple pieces and sampled. Material for every harvested sample was divided and separately collected into two 15-ml Falcon tubes, one for chemical and one for gene expression analyses, and immediately flash-frozen in liquid nitrogen. Scissors were cleaned with 70% EtOH and tissue paper between harvests of different organ samples to avoid cross-contamination.

Sample processing and extraction of endogenous semi-polar metabolites for metabolomic analysis

First, samples were freeze-dried under vacuum to constant weight for 3 days in a freeze drier (FreeZone Plus 12; Labconco, Kansas City, MI, USA) at −80 °C. Thereafter, dried samples were ground using a ball mill (Mixer Mill MM 400; Retsch) containing two metal beads (5 mm Ø; 50 Hz, 3 × 10 s). Ground samples were stored in 2-ml Safe-Lock® tubes (Eppendorf, Hamburg, Germany) at room temperature. Aliquots of 20 ± 1 mg (leaves and stems) and 10 ± 1 mg (roots) were

weighed into 2 ml round-bottom Eppendorf tubes for metabolite extraction. Some root samples of vegetative plants weighed <10 mg; in those cases, extraction buffer volume was adjusted proportionally to mass of the sample. The samples were extracted using the protocol described in Chiocchio *et al.* (2023). Briefly, samples were extracted twice in 1 ml (leaves and stems) and 0.5 ml (roots) 3:1 methanol:acetate buffer (pH 4.8) in 2 ml reaction tubes containing metal beads by shaking in a ball mill (Mixer Mill MM 400; Retsch) at 50 Hz for 5 min. Thereafter, samples were centrifuged for 15 min at 15,000 *g* at 4 °C. Supernatants (~0.8 ml) of both extraction steps were combined and centrifuged for 10 min at 15,000 *g*. Dilutions of 1:10 (leaves, stems) and 1:5 (roots) were prepared by pipetting aliquots into amber 1 ml HPLC vials containing extraction buffer.

Chemotyping and metabolomic profiling using UPLC-qToF-MS

Metabolomic profiling of semi-polar metabolites was conducted as described by Chiocchio *et al.* (2023). Extracts were injected into a UPLC–MS (Dionex UltiMate 3000; Thermo Fisher Scientific, Waltham, USA) equipped with a C18 analytical column (Acclaim TM RSLC 120; 2.1 × 150 mm, 2.2 µm particle size, 120 Å pore size). The column was maintained at a constant temperature of 40°C. Mobile phases used consisted of water/formic acid (0.05% v/v, solvent A), and acetonitrile/-formic acid (0.05% v/v, solvent B). The flow rate was 400 µl·min⁻¹. The multi-step gradient for solvent B was: 0–1 min 5%, 1–4 min 28%, 4–10 min 36%, 10–12 min 95%, 12–14 min 95%, 14–18 min 5%. The chromatograph was equipped with an autosampler that maintained samples at a constant temperature of 4°C and injected sample volumes of 1 µl (leaves and stems) or 2 µl (roots). The chromatograph was coupled with a maXis impact HD MS-qToF (Bruker Daltonics, Bremen, Germany) operated in positive polarity. ESI source conditions were: end plate offset = 500 V, capillary = 4500 V, nebulizer = 2.5 bar, dry gas = 11 L·min⁻¹, dry temperature = 220 °C. Transfer line conditions were: funnels 1 and 2 = RF 300 Vpp, isCD energy = 0 eV, hexapole = 60 Vpp, quadrupole ion energy = 5 eV, low mass = 50 *m/z*, collision cell energy = 10 eV, collision RF = 500 Vpp, transfer time = 60 µs, pre-pulse storage = 5 µs. The mass spectrometer was operated in full scan (MS1) mode with a mass range of 50–1500 *m/z* and a spectral acquisition rate of 3 Hz. Masses were calibrated using sodium formate (10 mM) clusters, prepared by combining 250 ml isopropanol, 1 ml formic acid, and 5 ml 1 M sodium hydroxide. The mixture was adjusted to a final volume of 500 mL with water.

Selection of candidate *GAME* genes and primer design

The *GAME9* transcribes a transcription factor that regulates *GAME* and upstream mevalonate pathway genes (Cárdenas *et al.* 2016) and its expression was used as an indicator of overall SGA biosynthetic activity. *GAME4* transcribes a cytochrome P450 that is active at the bifurcation step for biosynthesis of SGAs and SSGs (Paudel *et al.* 2017) and its expression was used as proxy for the influx of SA precursors into the SGA pathway. Expression of *GAME25* (Lee *et al.* 2019; Sonawane *et al.* 2018) and *S5αR2* (Akiyama *et al.* 2019) were used as proxies for

conversion of unsaturated steroidal aglycones into saturated aglycones by GAME25 and S5αR2. Lastly, *DPS* expression was used as proxy for potential conversion of spirostanes into solanidane-type SGAs (Akiyama *et al.* 2021). In addition, primers for reference genes, *SdEXP* (Expressed sequence) and *SdSAND* (a SAND family gene), were selected from their use in literature (Calf *et al.* 2019, 2020). For primer design, cDNA sequences of abovedescribed genes-of-interest (GOIs) functionally characterized in *S. tuberosum* and *S. lycopersicum* were queried against a *S. dulcamara* transcriptome (D'Agostino *et al.* 2013) using BLAST. All alignments with identity >40% were further examined by determining and translating their longest open reading frame (ORF) using the Expasy 'Translate' tool (Gasteiger *et al.* 2003). Then, amino acid sequences were aligned with the protein sequence of the GOI using CLUSTAL Omega (Sievers *et al.* 2011). This multiple sequence alignment was visually inspected to select the best homologue among the selected contigs. Selected *S. dulcamara* homologues of the GOI were fed into the NCBI Primer BLAST (Altschul *et al.* 1990) by uploading the relevant FASTA sequences and the *S. dulcamara* transcriptome (D'Agostino *et al.* 2013). The search parameters were left unchanged, except for PCR product size (75–200, Min–Max), Tm (58.0–60.0–62.0, Min–Opt–Max), primer GC content (45.0–65.0%, Min to Max) and Max Poly-X (4). In addition, Primer3 was used to generate additional candidate primer pairs (Untergasser *et al.* 2012). The designed primer pairs were tested for specificity by evaluating the NCBI Primer BLAST results and using the BLAST tool of Solgenomics. Subsequently, the OligoAnalyzer tool (https://eu.idtdna.com/pages/tools/oligoanalyzer) of Integrated DNA Technologies (Coralville, IA, USA) was utilized to assess thermodynamic stability of any predicted secondary structures formed by the most selective primer pairs. The parameters of the analysis tool were adjusted to fine-tune it for use on RNA sequences meant for PCR. A cut-off of -9 kcal·mol^{-1} was used for the ΔG value, where any structure predicted with a ΔG below -9 kcal mol^{-1} resulted in rejection of the corresponding primer pair. The following gene specific primers were used for RT-qPCRs of GOIs: *SdGAME9*-F: GTGGTGTGTGAGG AAAACGC, *SdGAME9*-R: CTCGGATCTTGTAAGCGGCT; *SdGAME4*-F: ACGGGTTCTTCTGTAGCAGC, *SdGAME4*-R: TCTCGGCGATTAACAGCTCC; *SdGAME25*-F: TCTTGGCG TCCGATGAATCC, *SdGAME25*-R: ACAGCACACCAACGAG AGAG; *SdS5αR2*-F: GACCCGAATAAGACCAGCCC, *SdS5αR2*-R: TACCCTCTTCGCCTCCACTT; *SdDPS*-F: TGGTTTTAGA GAGTCTTGGGCT, *SdDPS*-R: CCACCATCTGTGTGGCTACC; *SdSAND*-F: TGCTTACACATGTCTTCCACTTGC, *SdSAND*-R: AAACAGGACCCCTGAGTCAGTTAC and *SdEXP*-F: CTAA GAACGCTGGACCTAATGACAAG, *SdEXP*-R: AAAGTCGA TTTAGCTTTCTCTGCATATTTC.

RNA extraction and gene expression analysis using RT-qPCRs

Fresh plant material was stored in 15 ml Falcon tubes at $-80\,^{\circ}$C until sample processing. Frozen plant tissues were ground to a fine powder under liquid nitrogen using a mortar and pestle. The ground samples were stored in 1.5 ml Eppendorf tubes at $-80\,^{\circ}$C. Total RNA was extracted from ground plant material according to a protocol adapted from Oñate-Sánchez & Vicente-Carbajosa (2008). Extracted RNA samples were treated with DNAase I (Thermo Fisher Scientific)

according to the manufacturer's instructions. RNA integrity was visually inspected using gel electrophoresis. To check for RNA quality, absorbance ratios 260/230 and 260/280 nm were measured using a P330 NanoPhotometer® (IMPLEN, Munich, Germany) and quality checks were passed at absorbance ratios of in the ranges ~2–2.2 and ~1.8–2, respectively. Thereafter, 2 µg of DNA-free RNA were transferred to a new 0.2 ml PCR tube containing 24 µl autoclaved ddH$_2$O. Subsequently, 1 µl 50 µM Oligo dT 20 was added, after which the mixture was spun down. Thereafter, 4 µl of 5× RT buffer, 2 µl 10 mM dNTP Mix and 1 µl RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific) were added, and the tube spun down again. Samples were incubated in a thermocycler (Techne, Stone, UK) for 60 min at 42 °C, 15 min at 50 °C and 15 min at 70 °C. Each sample was measured in triplicate following RT-qPCR procedures on the CFX384 Real-time system (Bio-Rad, Munich, Germany), using 1 µl cDNA, 10 µl Dream-Taq polymerase (DreamTaq Green PCR Master Mix 2×; Thermo Fisher Scientific), 0.5 µl 10 µM forward and reverse primers and 8 µl autoclaved ddH$_2$O in a total volume of 20 µl per reaction. The qPCR conditions were: 10 s at 95 °C and 30 s at 60 °C for 40 cycles.

Data processing and statistical analyses

All statistical analyses and data visualizations were performed and produced in R (version 4.3.1; R Core Team 2021), except when explicitly mentioned otherwise. Data visualizations were performed using R package *ggplot2* (Wickham 2016). In general, Linear Mixed Models (LMMs) were built using packages *lme4* and *glmmTMB* (Brooks *et al.* 2017). Model fit and residual diagnostics were checked using the *performance* (Bates *et al.* 2015) and *DHARMa* (Hartig 2022) packages.

A peak-intensity table was produced by simultaneous pre-processing all LC–MS data in MetaboScape 5 (Bruker Daltonics, Bremen, Germany). The resulting table was sum-normalized, log-transformed and mean-centred, after which principal components analysis (PCA) was performed using MetaboAnalystR 4.0 (Pang *et al.* 2021).

A mass-difference network (MDN) was inferred using the *MetNet* package (Naake & Fernie 2019). Briefly, commonly observed neutral losses (NLs) in SGs (Heinig & Aharoni 2014) including hydroxylation, glycosylation, acetylation and malonylation, were used to cluster nodes (which represent ions of specific *m/z*) associated with SGs. Then, a retention time-corrected adjacency matrix based on structural information was built. Undirected network graphs were produced from the structural adjacency matrix and exported to .graphml format using R package *igraph* (Csardi & Nepusz 2006). Singleton nodes were removed from the MDN. Then, the MDN was visualized using Cytoscape version 3.8.2 (Shannon *et al.* 2003), and relative intensities in LC–MS were mapped onto its nodes. Mass spectra were manually exported from Bruker Data Analysis (v. 5.2 Bruker Daltonics) and wre plotted using SciDAVis version 2.4.0. Chemical structures of putative metabolites were drawn using ACD/ChemSketch version 2020.1.2 (Advanced Chemistry Development, Toronto, Ontario, Canada).

The MDN was manually inspected for *m/z* signals associated with steroidal aglycones, and glycosylated steroidal aglycones. Then, ion chromatograms (EICs) were manually extracted using Bruker Data Analysis (v.5.2; Bruker Daltonics) for 12 *m/z*

values associated with steroidal aglycons (Fig. 2c; hereafter 'SA species'): 412.32, 414.35, 416.35, 428.32, 430.33, 432.35, 415.32, 417.33, 433.33, 434.36, 446.33 and 448.34. Mass spectra associated with peaks in EICs were inspected for neutral losses indicative of glycosylation with pentoses, deoxyhexoses, or hexoses. Only peaks with a glycosylation signature were retained and counted. When multiple of the aforementioned SA species were found in the same mass spectrum, only the heaviest fragments were counted. The resulting matrix, containing the counted number of SGs per SA species for every chromatogram, was used to calculate diversity indices including Margalef's richness (Margalef 1958; D_{mg}) and Pielou's evenness (Pielou 1966; J) which were defined as:

$$\text{Margalef's richness} = D_{mg} = \frac{(S-1)}{\ln(N)}$$

where S is number of unique SA species, and N is total number of SG molecules detected in a chromatogram; and

$$\text{Pielou's evenness} = J = \frac{H'}{H'_{max}} = \frac{H'}{\ln(S)} = -\frac{\sum_i^S p_i \ln(p_i)}{\ln(S)}$$

with $p_i = n_i/N$, where n_i is the number of SG molecules belonging to the i_{th} SA species, N is the total number of SG molecules detected in a chromatogram, and S is number of unique SA species (i.e. J is equal to the ratio of the sample-specific Shannon-index (H'; Shannon 1948) and its maximum value (H'_{max}).

The LMMs were built using SG richness and evenness as response variables. Fixed effects were modelled as the interaction between 'chemotype' and 'organ, while 'plant individual' nested within 'genotype' was modelled as random effect. The Wald test was used to test the significance of predictors in models, with SG richness and evenness as responses. For post-hoc testing, the estimated marginal means (EMM) method was applied to calculate EMMs for Margalef's richness (D_{emm}) and Pielou's evenness (J_{emm}) for treatment groups using the R package *emmeans* (Lenth 2024). TIC of SG-associated features (TIC_{SG}) was calculated by taking the sum of the intensities of ions in the MDN (excluding signals with m/z 329.32). Pearson correlation coefficient was used to investigate the relationship between TIC_{SG} and SG chemodiversity.

In order to partition the observed variation in SG counts for every experimental factor, two separate generalized linear models (GLM) of the Poisson family were built using the R package *gASCA* (Franceschi 2022). For the first GLM, organ-type, chemotype and their interaction were specified as fixed effects. For the second GLM, sample-type (defined as different organ–chemotype combinations), ontogeny and their interaction were specified as fixed effects. Models were subsequently used for ANOVA-simultaneous component analysis (ASCA). The decomposition was validated using a permutation-based approach (n = 1000). Variables with higher R^2_{pseudo} in the specified models compared to their respective null-model were selected for decomposition in the presented GLM-ASCA models.

For RT-qPCR data, generalized linear mixed models (GLMMs) of the Poisson family were built in a Bayesian framework using the R package *MCMC.qpcr* (Matz *et al.* 2013). For *SdDPS*, *SdGAME4*, *SdGAME9*, *SdGAME25*, and *SdS5αR2*, amplification efficiencies were calculated based on dilution series (1–1000×), and for *SdEXP* and *SdSAND* efficiencies of 2 were assumed and were used to inform the model. First, raw Cq values from RT-qPCR were transformed into counts using the function *ct2counts()*. The MCMC chain in *mcmc.qpcr()* was set to 110,000 iterations, with a thinning interval of 100, and the initial 10000 iterations were discarded. Model convergence was inspected by checking the column 'eff.samp' in the model summary for values smaller than the difference between the total number of iterations and the number of discarded iterations, divided by the thinning interval (Matz *et al.* 2013).

To check for global effects, a naïve model containing all RT-qPCR data was built. First, log₂-transformed data were extracted from the naïve model. Then, Manhattan distance indices were calculated from normalized data and were used as input for PERMANOVA and PCoA, using the functions *adonis2()* and *cmdscale()* from the R packages *vegan* and *stats*, respectively. Thereafter, separate soft-normalization models were built per ontogenetic group using the *mcmc.qpcr()* function. Precisely, the interaction between the factors "target gene", "chemotype", and "organ" was specified as fixed effect, while "sample-type" (defined as organ–chemotype combinations) and "genotype" were specified as crossed random effects. *SdEXP* and *SdSAND* were used as additional reference genes in the soft-normalization models. False-discovery rate adjusted P-values were calculated for pairwise comparisons of organ–chemotype combinations per ontogenetic stage.

RESULTS

Ontogenetic, chemotypic and organ-specific variation in semi-polar metabolites in *S. dulcamara*

To study whether overall phytochemical diversity of semi-polar metabolites varies among different ontogenetic stages, organs and chemotypes, PCA was performed using 2906 features that eluted in the retention time range 0.75–11.00 min. The samples from vegetative plants clustered together and separated from those of flowering plants on PC1, which explained 26.3% of the observed variance (Fig. 1a; symbols). Additionally, the samples taken from the different organs separated on PC2, which explained 20% of the observed variance (Fig. 1a; colours). Furthermore, leaf samples from the *S*- and *U*-chemotypes in the vegetative stage separated on PC3, which explained 8% of the observed variance (Fig. 1b). This was different in flowering plants: half of the leaf and stem samples of *S*-chemotype clustered with *U*-chemotype samples (Fig. 1c). In both vegetative (Fig. 1b) and flowering (Fig. 1c) plants, the root samples of *U*- and *S*-chemotypes clustered close together.

Mass-difference networking annotates features associated with chemotypic steroidal glycoside (SG) variation in *S. dulcamara*

To analyse SG chemodiversity among the two *S. dulcamara* SG chemotypes, mass-difference networks (MDNs) were inferred from LC–MS data of all samples. This approach allows for clustering of features based on specified neutral losses, and for
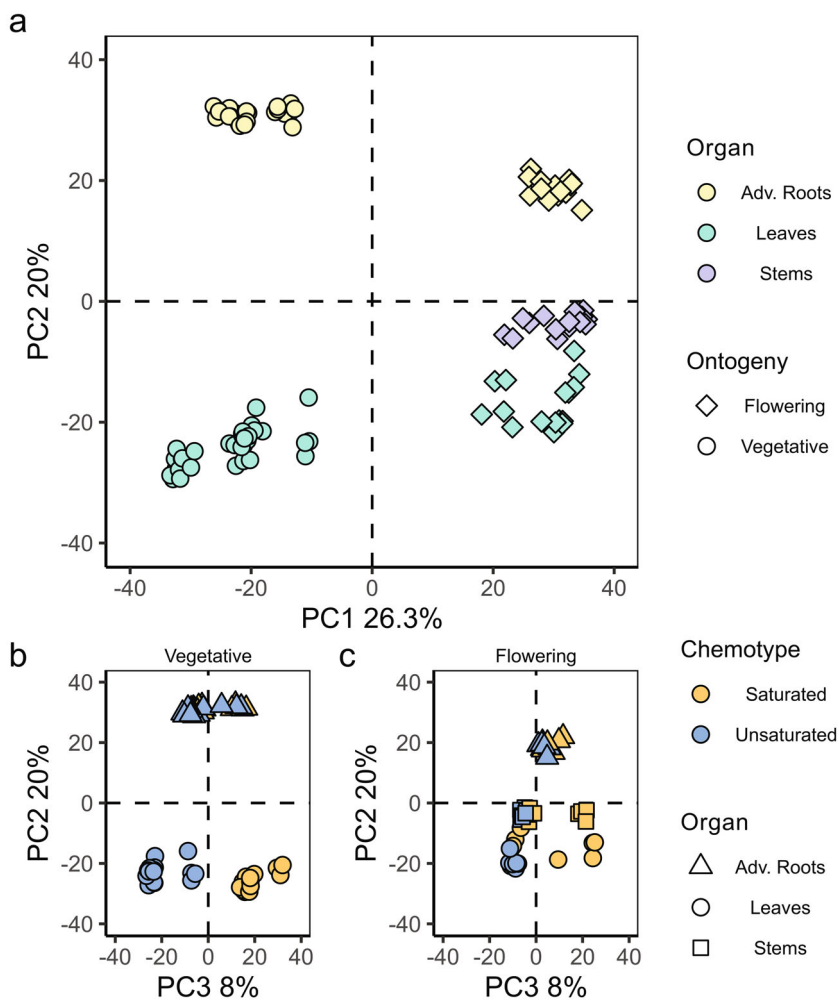
Fig. 1. Plot of principal components analysis (PCA) scores based on 2906 features from liquid chromatography–mass spectrometry (LC–MS) eluting in the retention time range 0.75–11.00 min. Symbols represent individual samples. PCA scores (a) of PC1 and PC2. Symbols represent plant ontogeny (diamonds for vegetative, and circles for flowering plants) while symbol colour represents organ from which extracted tissue originated (yellow: adventitious roots, green: leaves, purple: stems). Plots of PC3 and PC2 for vegetative (b) and flowering (c) plants. Symbol shape represents organ-of-origin (triangle: adventitious roots, dot: leaves, square: stems), while symbol colour represents leaf chemotype (yellow: saturated steroidal glycosides (SGs), blue: unsaturated SGs). PCA loadings are shown in Figure S1.

visualization of features as nodes in which the size is proportional to ion-intensity in the MS (Fig. 2a). Using this approach, features associated with SGs and their in-source fragments could be classified, since they formed a distinct MDN (Fig. 2a, highlighted SG cluster). Zooming in to this SG-associated MDN, we visualized which nodes were associated with each of the chemotypes (Fig. 2b, colour gradient).

The SG-associated network globally separated the *S-* (Fig. 2b, yellow circles) and *U-* (blue circles) chemotypes. The separation among the two chemotypes was based on nodes that differ in their ring-double bond equivalent (RBDE; Fig. 2c). Nodes associated with the *S*-chemotype (yellow) commonly had *m/z* ratios that were 2 Da higher than those associated with the *U*-chemotype (blue). This indicates that the SGs in *S*-chemotypes were overall more saturated, as evidenced by a lower RDBE value (Fig. 2c). The grey nodes represent features that are shared among the two chemotypes, and thus form the common metabolome of all detected semi-polar metabolites (Fig. 2a) and SGs (Fig. 2b) in *S. dulcamara*.

Mass-difference networking annotates features associated with structural steroidal aglycone (SA) variation in *S. dulcamara*

To study the structural diversity of the steroidal aglycones in the two *S. dulcamara* chemotypes, the in-source fragmentation-based MDNs were further inspected for their mass-differences. The presence of nodes with both odd and even *m/z* values in the MDN suggest that the plants contain both SSG- (odd *m/z*) and SGA-type (even *m/z*) SGs (Fig. 2b). Additionally, the MDN shows that steroidal aglycones in both SG classes vary in RDBE levels. Multiple nodes with *m/z* 414.34 and *m/z* 416.35 were detected, associated with the steroidal aglycones of SGAs (steroidal alkamines), and were putatively annotated as solasodine/tomatidenol and soladulcidine/tomatidine (Eich 2008), which have RDBE values of 6 and 7, respectively (Fig. 2c). Two other nodes, with *m/z* 415.32 and *m/z* 417.34, found in top and centre right of the MDN, are associated with the steroidal aglycones of SSGs (steroidal sapogenins), and were putatively annotated as diosgenin/yamogenin
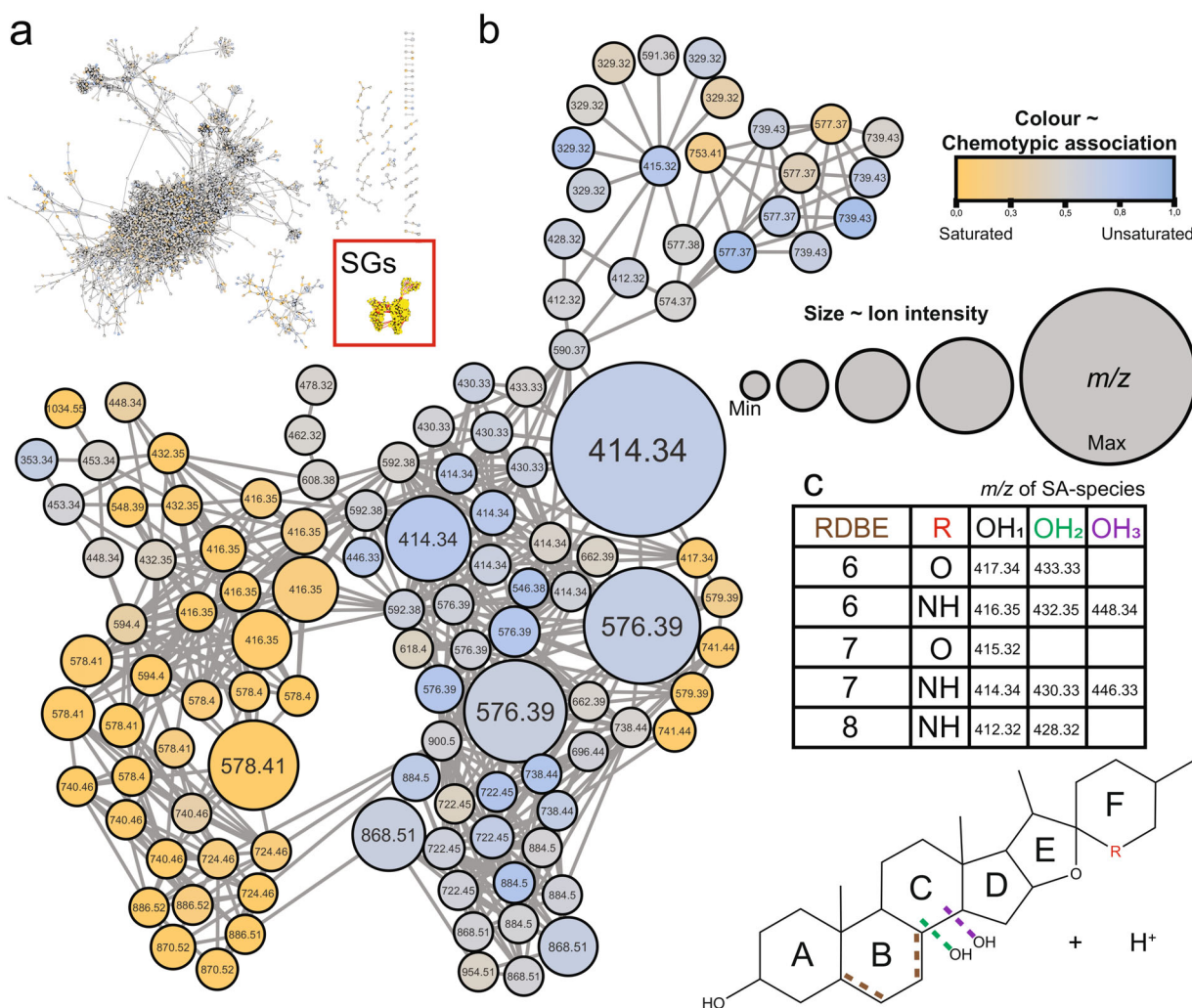
Fig. 2. *Ab initio* inferred mass difference network (MDN) of 118 LC–MS features (a), and subnetwork (b) associated with steroidal glycosides, including steroidal glycoalkaloids (SGAs) and steroidal sapogenin glycosides (SSGs). Numbers in nodes depicts mass-over-charge (*m/z*) signals resulting from in-source fragmentation, while connecting edges represent neutral losses (e.g. biotransformations, such as glycosylation, hydroxylation, malonylation, etc.). Node size is proportional to ion intensity in LC–MS. Node colour represents chemotypic association of node (yellow: saturated SGs, blue: unsaturated SGs; grey: neutral). Nodes are based on LC–MS data generated from all samples analysed. Note that isobaric ions in MDN elute at different retention times. (a) Nodes associated with SGs organize in a subcluster which is (b) coloured by chemotypic association of the node. (c) Proposed structural variation of steroidal aglycones based on *m/z* values in the MDN. RDBE, ring-double bond equivalent; R, NH or O; OH1, 3-hydroxyl; OH2 and OH3, additional hydroxylation of steroidal aglycones.

and dehydrodiosgenin/dehydroyamogenin (Eich 2008), which have RDBE values of 6 and 7, respectively (Fig. 2c).

Furthermore, the MDN revealed variation in the hydroxylation level of steroidal aglycones in *S. dulcamara* (Fig. 2b). For the putative steroidal alkamine aglycones solasodine (*m/z* 414.34) and soladulcidine (*m/z* 416.35), we found nodes associated with their mono-hydroxylated (*m/z* 430.33–432.35) and di-hydroxylated (*m/z* 446.33–448.34; Fig. 2c) analogues, respectively. Interestingly, we detected another pair of SGAs with mono-hydroxylated (*m/z* 412.32–428.32) steroidal aglycones with an RDBE value of 8 (Fig. 2c). For the putative steroidal saponin e, we found only one pair of nodes (*m/z* 417.34–433.33) indicative of hydroxylation (Fig. 2c). Next to the tabulated *m/z* values (Fig. 2c), two additional nodes, with *m/z* 434.36 (RDBE = 5) and *m/z* 453.34, were annotated as potential SGA- and SSG-type steroidal aglycones, respectively, based on their occurrence in the MDN.

Mass-difference networking annotates mass spectra associated with organ-specific and ontogenetic SG variation in *S. dulcamara*

To study the organ- and ontogeny-specific SG distribution in the two *S. dulcamara* chemotypes, nodes in the MDN were visualized as pie charts showing the relative intensity of the node by organ (Fig. 3) or ontogeny (Fig. 4). Using this visualization method, putative SGs exclusively detected in a chemotype (Fig. 2b), organ-type (Fig. 3) or ontogenetic stage (Fig. 4) were annotated.

Features associated with saturated SGAs in leaves were the main drivers separating the two clusters of nodes according to chemotype (Figs 2b and 3). A saturated SGA tetraose I (Fig. 3, Figure S2, *m/z* values: 416.35, 528.77, 578.40, 740.46 and 1034.55) and two saturated SGA triose II (Fig. 3, Figure S2, *m/z* values: 416.35, 454.75, 578.40, 724.46 and
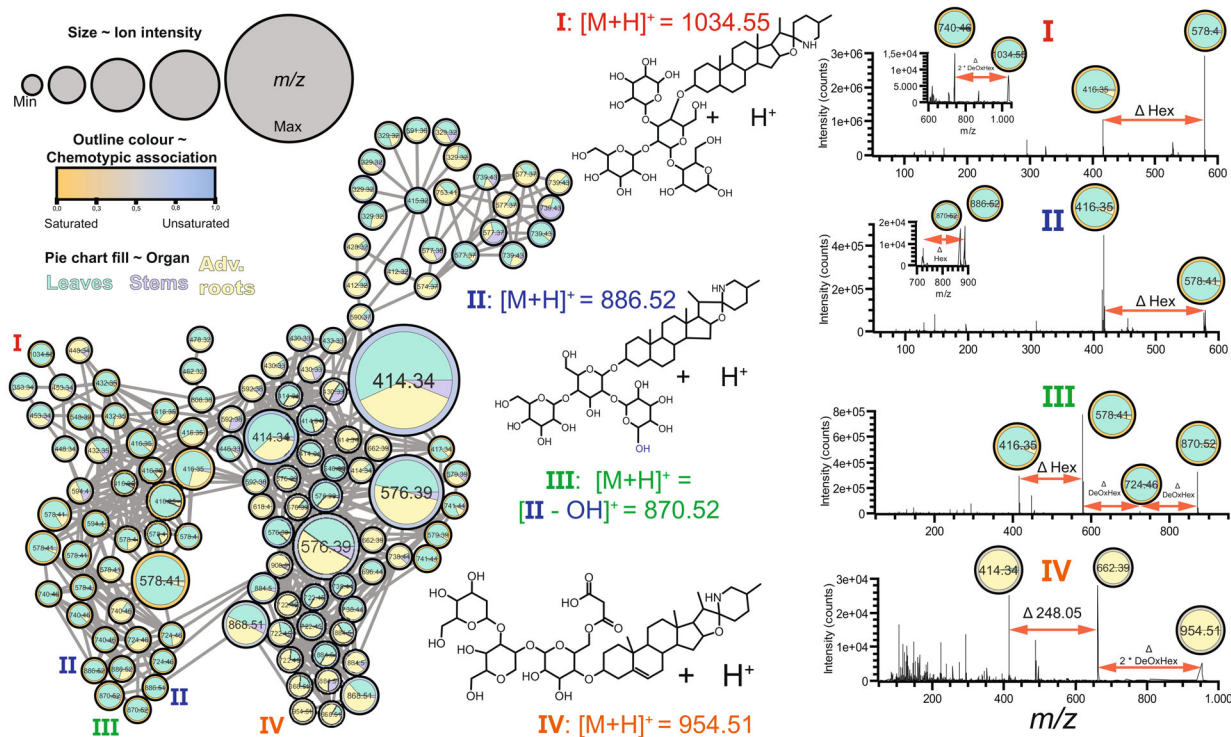
Fig. 3. *Ab initio* inferred mass difference networks (MDN) of 118 LC–MS features (nodes) associated with steroidal glycosides, including steroidal glycoalkaloids (SGAs) and steroidal sapogenin glycosides (SSGs). Number in nodes depicts mass-over-charge (*m/z*) signal resulting from in-source fragmentation, while connecting edges represent neutral losses (e.g. biotransformations such as glycosylation, hydroxylation, malonylation, etc.). Node size is proportional to ion intensity in LC–MS. Nodes are coloured by chemotypic association of the node (yellow: saturated SGs, blue: unsaturated SGs; grey: neutral). Relative intensity for every node is visualized as a pie chart coloured by organ (yellow: adventitious roots, green: leaves, purple: stems). Note that isobaric ions in the MDN elute at different retention times. Neutral losses are visualized within mass spectra using red double-headed arrows. Hex, hexose; DeOxHex, deoxyhexose; Δ 248.05, malonylhexoside. Mass spectra of hypothesized compounds I–VI in this figure are shown in Figure S2, while for compounds V and VI, mass spectra are also shown in Fig. 4.

886.52), which are putatively annotated as 'soladulcine B' and 'soladulcine A', respectively, were detected in the *S* but not in the *U*-chemotype. Interestingly, the saturated SGAs I and II associated with *S*-chemotype plants (Figs 2b and 3) were most prominently detected in leaf samples, predominantly in those of vegetative plants (Fig. 4). Furthermore, one saturated SGA triose III (Fig. 3, Figure S2; *m/z* values: 416.35, 454.75, 578.40, 724.46 and 870.52) with putative molecular ion *m/z* 870.52 and multiple constitutional isomers were detected, as the molecular ion is found at least twice in the MDN (Fig. 3). Compared to compound II, isomers of III were associated with the substitution of a hexose for a deoxyhexose in the glycoside moiety of III (Fig. 3).

Malonylglucoside SGA exclusively detected in roots of vegetative *S. dulcamara*

In the MDN coloured by organ-type, most of the nodes had two or three colours, meaning that they represent features found in leaves, roots and stems. The nodes representing SGs **I**, **II**, **III**, however, were predominantly found in leaves. Interestingly, we found a few nodes that coloured completely yellow, meaning that these features were exclusively detected in roots (Fig. 3). These *m/z* features were putatively assigned to a malonylated SGA with molecular ion *m/z* 954.51 and base peak *m/z* 662.39 (**IV**; Fig. 3). We also found a completely yellow node for *m/z*

414.34, which is indicative of a root-specific unsaturated nitrogenated steroidal aglycone (Fig. 3, Figure S2). The fragmentation pattern of **IV** shows two sequential neutral losses of deoxyhexose ($\Delta_{m/z}$ 146.05) moieties, followed by the loss of $\Delta_{m/z}$ 248.05 (Fig. 3), the latter of which indicates the loss of a malonylglucoside moiety. This fragmentation pattern, together with the base peak *m/z* 662.39, indicates that malonylglucoside is part of a larger trisaccharide moiety that is conjugated to the 3-hydroxyl position of an unsaturated steroidal aglycone (Fig. 3).

Steroidal saponin glycosides vary across organs and ontogeny in *S. dulcamara*

A cluster of nodes associated with unsaturated SSGs (**V**) formed a subnetwork on top of the larger MND (Figs 3 and 4). We assigned the *m/z* features in these nodes to putative SSG pentosides with molecular ion *m/z* 1195.5662 and base peak *m/z* 739.4221. Additionally, we detected a node with *m/z* 415.31, which is indicative of an unsaturated oxygenated steroidal aglycone (**V**, Fig. 4). The fragmentation pattern of this putative saponin pentoside (Fig. 4, Figure S2) showed sequential neutral losses, indicative of the cleavage of a hexose and a pentose, followed by the loss of three hexoses.

Interestingly, a cluster of five nodes, three of which with *m/z* 739.43, were found at the top of the MDN (**V**), suggesting that there are (sub)structural isomers of **V** in *S. dulcamara* (Fig. 4).
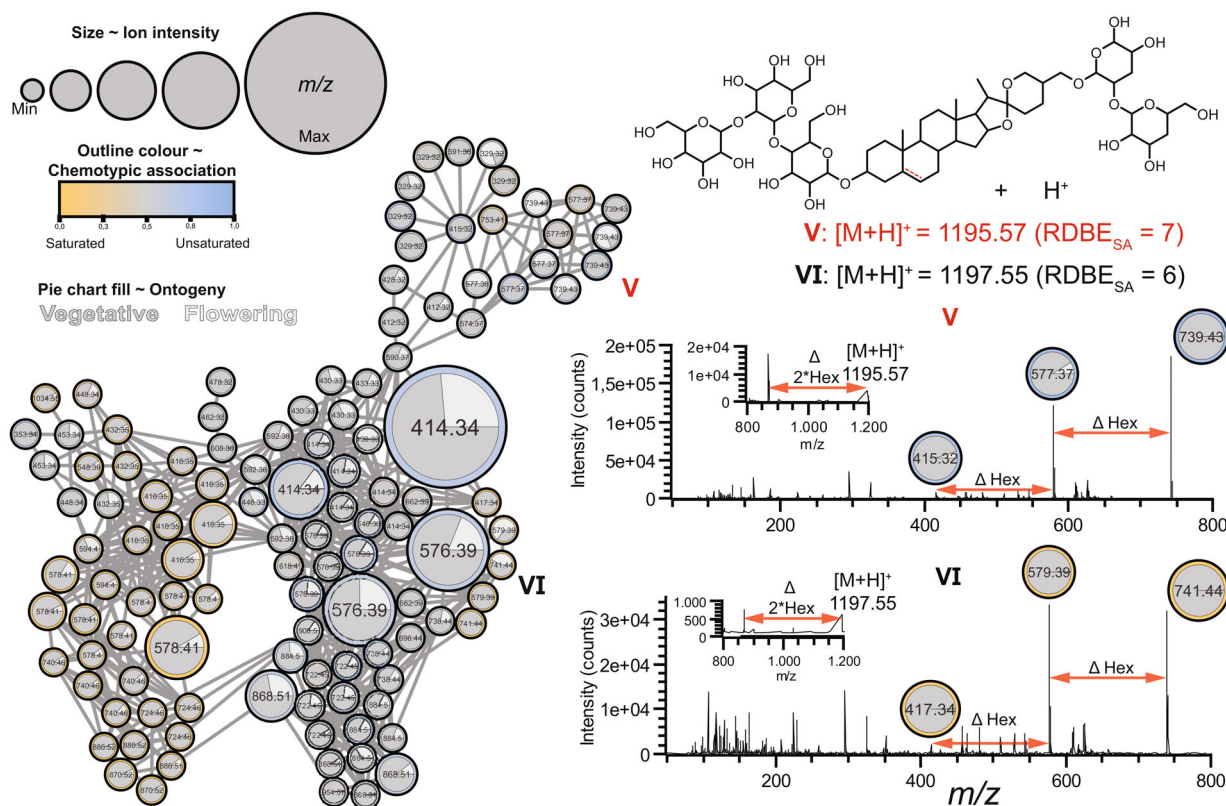
Fig. 4. *Ab initio* inferred mass difference networks (MDN) of 118 LC–MS features (nodes) associated with steroidal glycosides, including steroidal glycoalkaloids (SGAs) and steroidal sapogenin glycosides (SSGs). Number in nodes depicts mass-over-charge (*m/z*) signals resulting from in-source fragmentation, while connecting edges represent neutral losses (e.g. biotransformations, such as glycosylation, hydroxylation, malonylation, etc.). Node size is proportional to ion intensity in LC–MS. Node outline and legend network are coloured by chemotypic association of the node (yellow: saturated SGs, blue: unsaturated SGs; grey: neutral). Relative intensity for every node is visualized as a pie chart coloured by ontogeny (dark grey: vegetative, light grey: flowering). Note isobaric ions in MDN elute at different retention times. Neutral losses are visualized within mass spectra using red double-headed arrows. Hex, hexose; DeOxHex, deoxyhexose. Mass spectra of hypothesized compounds V and VI in this figure are shown in Figure S2, while compounds I-IV mass spectra are also in Fig. 3.

The colours of the SSG-associated nodes in Fig. 3 indicate that most features were found in roots and leaves, wherein two nodes (*m/z* 577.37 and 739.43) were more frequently found in stems and roots (Fig. 3; left upper cluster). Interestingly, four out of five nodes with *m/z* 739.43 are circled in blue and thus associated with *U*-chemotype plants, while one node is grey and thus occurs more-or-less equally in both chemotypes (Fig. 4). Of the four nodes associated with *U*-chemotype plants, three were predominantly associated with leaf samples, and one with stem samples (Fig. 3). Nodes in the SSG cluster that were encircled in grey (Fig. 3) were predominantly associated with roots (Fig. 3). Three of the five nodes with *m/z* 739.43 are predominantly associated with flowering plants, whereas the other two nodes are mainly detected in vegetative plants (Fig. 4). Additionally, two nodes with *m/z* 741.44 (Fig. 4, VI) were found in *S*-chemotype plants. These *m/z* features were annotated as SGG pentosides, which are the saturated analogues of **V** with putative molecular ion *m/z* 1197.55. (Fig. 4).

Variation partitioning of SG counts using GLM ANOVA simultaneous component analysis (GLM-ASCA) reveals dynamics of SG variability in *S. dulcamara*

Based on inspection of the EICs of the 12 annotated SA species (Fig. 2c), we counted a total of 3149 scans associated with SGs.

To study the variation in SG counts per SA species among chemotypes, organs and ontogenetic stages in the two *S. dulcamara* chemotypes, variation partitioning using a GLM-based ASCA was performed. The first GLM-ASCA model describes the interaction between 'Organ' and 'Chemotype' (Figure S3). Interestingly, SGs with SA species *m/z* 416.35 are associated with latent variables (LV) 1 and 2 for the interaction (Figure S3a), and LV1 for the term 'Chemotype'. This is in accordance with the chemotype selection, which was done based on *m/z* 416.35 (Figure S3b, c). Furthermore, the LV1 for the term 'Organ' (Figure S3a) projects leaf SG profiles in the negative direction (Figure S3b), which is associated with the higher number of SSGs (SA species with *m/z* 415.32, 417.33 and 433.33; Figure S3c) in leaves than in roots or stems. Lastly, LV2 of the term 'Organ' (Figure S3a) projects SG profiles of root samples in the positive direction (Figure S3b), which is associated with a higher number of SSGs with SA species *m/z* 433.33 (Figure S3c).

The second GLM-ASCA model describes the interaction between 'Sample type' (different organs per chemotype) and 'Ontogeny' (Figure S4). Interestingly, SGs with SA species *m/z* 416.35 are associated with LV1 and LV2 for the interaction and LV1 for the term 'Chemotype' (Figure S4a). This is in accordance with the chemotype selection, which was done based on the same *m/z* signal (Figure S4b, c). The LV1 of the interaction

term (Figure S4a) separates SG profiles of roots of vegetative and flowering plants (Figure S4b), which is associated with a higher number of SGs with SA species *m/z* 412.32 and *m/z* 433.33 in roots of flowering plants (Figure S4c). In contrast, SSGs with SA species *m/z* 417.33 (Figure S4c) were more present in roots of vegetative plants (Figure S4b). The LV1 for the term 'Ontogeny' (Figure S4a), indicated that chemical profiles of samples taken from vegetative and flowering were always clearly separated (Figure S4b). This was related to the higher number of SGs with SA species *m/z* 417.33 and 446.33 (Figure S4c) in vegetative organs, independent of chemotype. Lastly, the LV1 for the term 'Sample' (Figure S4a) showed that stems differentiate from the other samples (Figure S4b), which was associated with higher numbers of SGs with SA species *m/z* 416.35, 417.33 and 446.33 (Figure S4c). Furthermore, leaf SG profiles were separated from other organs in the negative direction (Figure S4b), which was associated with higher numbers of SGs with SA species *m/z* 412.32, 415.3 and 433.33 (Figure S4c) in leaves.

Chemical diversity indices provide insight in SG metabolism across organs and ontogeny in *S. dulcamara*

To compare the levels of SG chemodiversity among organs and across ontogeny in the two chemotypes, we counted the number of glycosylated signals of putative SGs for every SA species annotated by MDN (Fig. 2c). With these data, we calculated the Margalef's richness (D_{mg}) and Pielou's evenness (J) for every sample. Furthermore, we used the total ion current TIC_{SG} as a proxy for total SG quantity and studied its relationship with SG chemodiversity indices D_{mg} and J in different organs across ontogeny.

In general, the median SG richness was higher in samples of flowering ($D_{EMM} > 3.5$) than in those of vegetative ($D_{EMM} < 3.5$; Fig. 5a) plants. In vegetative plants, SG richness was higher in leaves (*S*-chemotype: $D_{EMM} = 3.23 \pm 0.0455$; 95% CI: [3.14–3.32]; *U*-chemotype: $D_{EMM} = 3.31 \pm 0.0455$; 95% CI: [3.22–3.40]) than in roots (*S*-chemotype: $D_{EMM} = 2.91 \pm 0.0499$; 95% CI: [2.81–3.01]; *U*-chemotype: $D_{EMM} = 2.92 \pm 0.0482$; 95% CI: [2.82–3.02]; Fig. 5a). In flowering plants of the *U*-chemotype, SG richness values were lower for leaves ($D_{EMM} = 3.72 \pm 0.0705$; 95% CI: [3.58–3.86]) than for stems ($D_{EMM} = 4.02 \pm 0.0665$; 95-% CI: [3.88–4.15]; Fig. 5a). Furthermore, stems of flowering plants of the *U*-chemotype have higher predicted SG richness than those of the *S*-chemotype ($D_{EMM} = 3.70 \pm 0.0665$; 95% CI: [3.56–3.83]; Fig. 5a).

Next to SG richness, we also analysed SG evenness. The measures of SG evenness varied more across flowering (EMMs $J_{min-max} = 0.832$–0.911; Fig. 5b, right) than vegetative (EMMs $J_{min-max} = 0.853$–0.885; Fig. 5b, left) organs, with a maximum observed EMM of $J_{max} = 0.911$ in leaves of flowering plants (Fig. 5b, right). In vegetative plants, SG evenness tended to be higher in *S*-chemotypes than in *U*-chemotype for both leaves (*S*-chemotype $J_{EMM} = 0.878 \pm 0.00641$; 95% CI: [0.865–0.890]; *U*-chemotype $J_{EMM} = 0.853 \pm 0.00641$; 95% CI: [0.841–0.866]) and roots (*S*-chemotype $J_{EMM} = 0.885 \pm 0.00701$; 95% CI: [0.871–0.899]; *U*-chemotype $J_{EMM} = 0.880 \pm 0.00679$; 95% CI: [0.867–0.894]), but the differences were not significant (Fig. 5b). In flowering plants, predicted leaf SG evenness is higher (*S*-chemotype $J_{EMM} = 0.911 \pm 0.00899$; 95% CI: [0.894–

0.929]; *U*-chemotype $J_{EMM} = 0.909 \pm 0.00899$; 95% CI: [0.890–0.928]) than that of stems (*S*-chemotype $J_{EMM} = 0.861 \pm 0.00964$; 95% CI: [0.842–0.880]; *U*-chemotype $J_{EMM} = 0.832 \pm 0.00964$; 95% CI: [0.812–0.851]), but not than that of roots (*S*-chemotype $J_{EMM} = 0.892 \pm 0.01007$; 95% CI: [0.872–0.912]; *U*-chemotype $J_{EMM} = 0.909 \pm 0.00949$; 95% CI: [0.890–0.928]; Fig. 5b). Taken together, this means that both organ and ontogeny interactively impact variation in the measures of chemodiversity.

Additionally, we analysed the relationship between SG diversity indices D_{mg} and J with TIC_{SG}. Overall D_{mg} correlated negatively with TIC_{SG} ($R = -0.73$, $P = 0.017$; Figure S7a). Specifically, in flowering *S*-chemotype plants, D_{mg} correlated negatively with the TIC_{SG} in leaves ($R_L = -0.82$, $P = 0.012$) and stems ($R_S = -0.77$, $P = 0.024$; Figure S7b). Overall J did not correlate with TIC_{SG} ($R = 0.12$, $P = 0.74$; Figure S7c). Furthermore, J correlated positively with the TIC_{SG} in leaves of *S*-chemotype ($R_L = 0.8$, $P = 0.017$) and stems of *U*-chemotype ($R_S = 0.76$, $P = 0.028$) plants in the flowering stage (Figure S7d).

Expression patterns of candidate *GAME* genes explain SG diversification among organs across ontogeny

To investigate the potential chemotype- and organ-specific expression for selected genes, RT-qPCR analyses were performed with total RNA extracted from different organs of vegetative and flowering *S. dulcamara* individuals. Primers were designed using a homology-based approach with previously characterized *GAME* genes from closely related cultivated *Solanum* spp. Transcript abundances (counts) were calculated based on C_q values given the primer efficiencies of candidate and reference genes. First, PCA and PERMANOVA were performed on a naïve model (Fig. 6a). The expression levels of *U*-chemotype leaf samples separated from all other samples on PCo1, which explained 38.4% of the observed variation (Fig. 6a). Analogously, the separation of *S*-chemotype leaf samples from stem and root samples was associated with PCo2, which explained 8.8% of the observed variation (Fig. 6a). Furthermore, leaf samples were organized in subclusters based on ontogeny (symbols; Fig. 6a).

Second, we analysed transcript abundances for every candidate gene based on an informed model which considers *EXP* and *SAND* as reference genes (Fig. 6b, c). *GAME9* codes for a transcription factor associated with steroidal glycoalkaloid biosynthesis (Cárdenas *et al.* 2016; Nakayasu *et al.* 2018). *SdGAME9* transcript abundance did not differ across organ or chemotype, neither in vegetative plants (Fig. 6b, *SdGAME9*) nor in flowering plants (Fig. 6c, *SdGAME9*). It is noteworthy that the expression pattern of *SdGAME9* in different organs, though not significantly different, closely resembles that of *SdGAME4* and *SdS5αR2* in flowering (Fig. 6b), but not in vegetative (Fig. 6a), plants.

GAME4 codes for a cytochrome P450 enzyme that is involved in early oxidation of steroidal precursors towards SGA biosynthesis (Itkin *et al.* 2013; Paudel *et al.* 2017). In vegetative plants, *SdGAME4* transcript abundances were higher in leaves than in roots. Additionally, *S*-chemotype roots had 1.37-fold higher transcript abundance than *U*-chemotype roots ($P = 0.0062$, Fig. 6b). When flowering, *SdGAME4* transcript abundances in leaves and roots of *S*-chemotype plants were
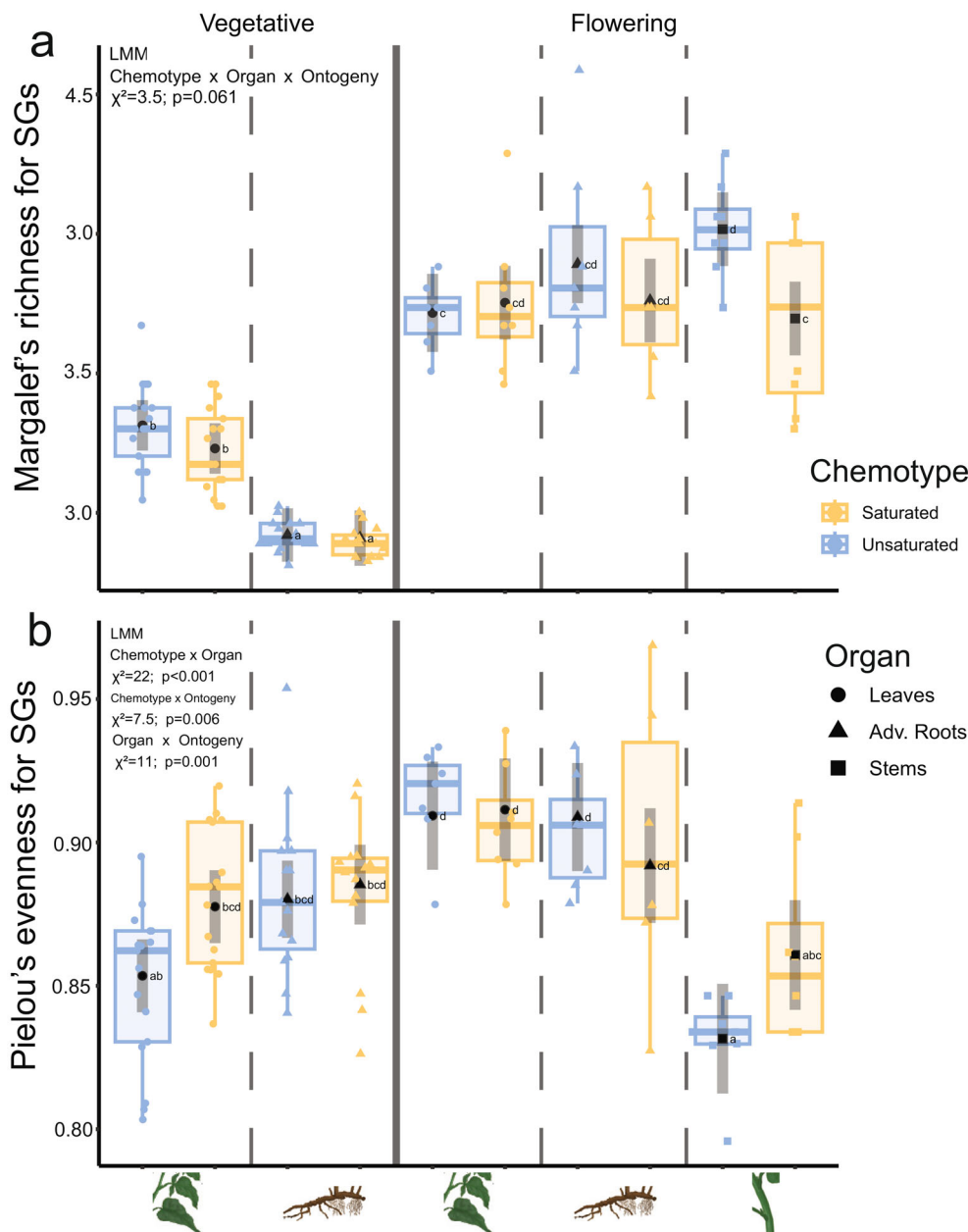
Fig. 5. Diversity indices based on a total of 3149 counted steroidal glycosides (SG) of 12 steroidal aglycone (SA) species. (a) Margalef's richness and (b) Pielou's evenness for SGs based on SG counts for 12 annotated SA species selected from mass-difference networking. Estimated marginal means (black symbols) and 95% confidence intervals (dark grey lines). Symbol shape represents organ (triangle: adventitious roots, dot: leaves, square: stems). Raw data are summarized as boxplots and individual data points are plotted for vegetative (left from thick line) and flowering (right) plants. Linear mixed models (LMMs) were built with the three-way interaction 'chemotype', 'organ', and 'ontogeny' as fixed effects, and plant individuals nested within genotype as random effect. For SG evenness, the three-way interaction term was removed from the model because it was not significant. Letters show outcome of Tukey's post-hoc test on estimated marginal means, and groups sharing a letter were not significantly different. Wald's χ^2-test (Table S2) was used to test for interactions between explanatory variables. $P < 0.1$; *$P < 0.05$; ***$P < 0.001$.

1.24-fold ($P = 0.0039$) and 2.9-fold ($P = 0.0094$) higher, respectively, than in leaves and roots of *U*-chemotype plants (Fig. 6c).

GAME25 and S5αR2 are two genes associated with double bond reduction in other *Solanum* spp. They code for a short-chain dehydrogenase/reductase (Lee *et al.* 2019; Sonawane *et al.* 2018) and a steroid Ç5α-reductase (Akiyama *et al.* 2019), respectively. GAME25 catalyses the first dedicated step towards saturated SGs. In leaves of vegetative *U*-chemotype plants, *SdGAME25* abundance is significantly lower, in fact close to the detection limit, compared to leaves of *S*-chemotype plants (Fig. 6b, *SdGAME25*). In contrast, *SdGAME25* transcript abundances are not significantly different among leaves of the two chemotypes when plants are flowering (Fig. 6c, *SdGAME25*). S5αR2 is the second enzyme involved in the reduction of the double bond in the B-ring of steroidal glycosides in *Solanum*
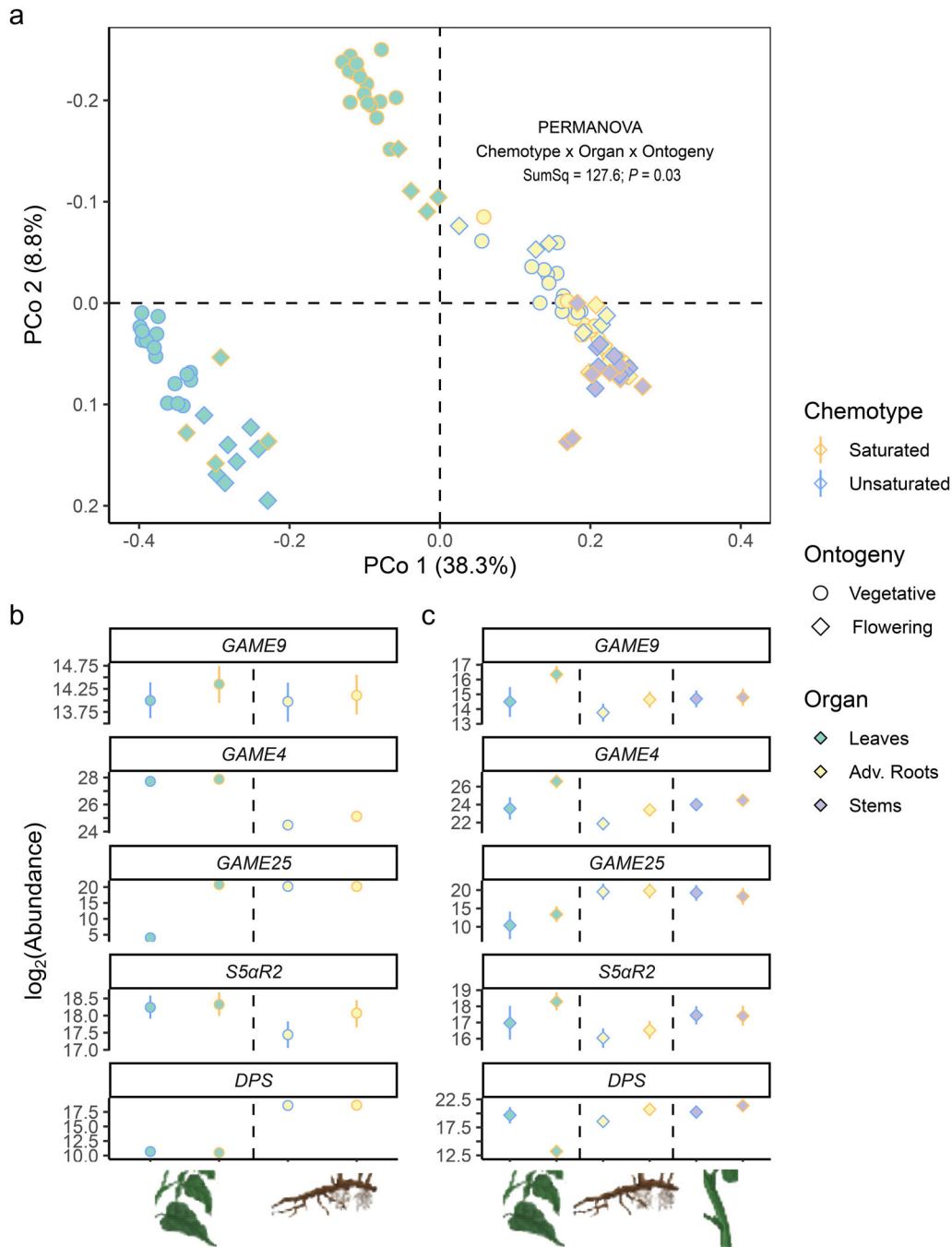
Fig. 6. The log$_2$-transformed transcript abundances from reverse transcription–quantitative polymerase chain reaction (RT-qPCR) analyses for (a) naïve and (b, c) soft-normalization models based on five candidate steroidal glycoside (SG) metabolism genes and two reference genes. (a) Principal coordinates analysis of log$_2$-transformed transcript abundances from the naïve model. Symbol shape represent ontogeny (diamonds: vegetative plants; asterisks: flowering plants); fill colour and line type represent organ-of-origin (light yellow triangles: adventitious roots, green dots: leaves, purple squares: stems). (b, c) Estimated marginal means (black symbols) and 95% credible intervals (coloured lines). Colours represent leaf chemotypes (blue: unsaturated steroidal glycosides (SGs); yellow: saturated SGs. Symbol shape represents ontogeny (triangle: adventitious roots, dot: leaves, square: stems) while line colour represents leaf chemotype (yellow: saturated SGs, blue: unsaturated SGs). Different organs are separated by a dashed line. A soft-normalization model was built using *EXP* and SAND as reference genes (priors) providing informed posteriors for target genes *GAME9*, *GAME4*, *GAME25*, *S5αR2* and *DPS*. (b) vegetative plants (c) flowering plants.

spp. Neither in leaves nor in roots of vegetative plants did we find statistically significant chemotypic differences in *SdS5αR2* abundance, although the expression levels in roots of *S*-chemotypes were visibly higher than those in *U*-chemotypes (Fig. 6b, *SdS5αR2*). In flowering plants, leaves of *S*-chemotype plants had higher *SdS5αR2* transcript abundances than leaves of *U*-chemotype plants, but the difference was not significant (Fig. 6c).

DPS is a 2-oxoglutarate-dependent dioxygenase that catalyses spirostanes into solanidanes in *S. tuberosum*. DPS catalyses the first dedicated step towards solanidanes by downstream C16α-hydroxylation. In vegetative plants, *SdDPS* abundance is higher in roots ($>2^{17}$) than leaves ($<2^{13}$; Fig. 6b). In flowering plants, *SdDPS* abundance was the lowest in *S*-chemotype leaves ($<2^{15}$), much lower than in all other organs ($>2^{17.5}$; Fig. 6c). In contrast, the absolute expression levels were 4.47-fold higher in roots of *S*-chemotypes than in *U*-chemotype plants ($P < 0.01$; Fig. 6c, *SdDPS*).

DISCUSSION

To date, SG variation in *S. dulcamara* leaves (Calf *et al.* 2018) and roots (Chiocchio *et al.* 2023) has been described using metabolomic approaches, but the extent of other levels of intra-individual variation in SG chemodiversity in *S. dulcamara* chemotypes remained unknown. Here, we show that there are additional levels of organ- and ontogeny-specific variation in SG chemodiversity in two studied *S. dulcamara* leaf chemotypes. Our untargeted metabolomic approached yielded 2906 picked LC–MS features. Combining PCA with mass-difference networking resulted in 118 SG-associated features. This allowed us to investigate intraspecific SG metabolism and to postulate the presence of at least 12 SA species in the *S. dulcamara* extracts. These analyses revealed that leaves of vegetative individuals of the two selected chemotypes have very distinct SG profiles, whereas roots on vegetative and flowering plants do not. On flowering plants, the SG profiles of leaves and stems were less distinct, showing that chemodiversity is affected by ontogeny. Finally, we have explored SG chemodiversity through manual investigation of EICs of the 12 annotated SA species. We counted 3149 mass spectra associated with SGs. These SG counts were used for variation partitioning using GLM-ASCA and to calculate the chemical diversity indices D_{mg} and J, which were used for univariate analyses using L(M)Ms. These analyses showed that overall chemodiversity varies among organs and over ontogeny, but not as much among chemotypes. Lastly, we used a homology-based approach to select five candidate and two reference genes of which the expression was measured using RT-qPCRs. These analyses showed that gene expression patterns of enzymes involved in SG biosynthesis were in line with the observed SG profiles. Our experiments show that organ- and ontogeny-specific variation in SG chemodiversity relates to the expression of candidate genes in SG metabolism.

Organ and ontogenetic variation in SG chemodiversity

Chemical profiling of the organs of SG chemotypes across ontogeny followed by MDN revealed that SG diversity in *S. dulcamara* is broader than previously described (Calf *et al.* 2018; Chiocchio *et al.* 2023). As expected, the putatively annotated unsaturated soladulcine B (**I**), and soladulcine A (**II**), are both predominantly detected in leaves of the *S*-chemotype (Lee *et al.* 1994). Unsaturated analogues of these SGAs were found in leaves of the *U*-chemotype, but also in roots of both chemotypes. In this study, unsaturated SSGs such as **V** were detected at high levels in the leaves and the roots of *U*-chemotype plants. Interestingly, these compounds were also detected at high levels in roots of *S*-chemotype plants, whereas they were very low in their leaves.

This implies that chemotypic differentiation is leaf-specific and not associated with the absence or presence of genes coding for SG biosynthesis. Furthermore, soladulcine B is a stereoisomer of α-tomatine, which suggests that soladulcine B may have similar herbivore-deterrent properties as these described for α-tomatine (Bailly 2021; You & van Kan 2021). In terms of intra-individual organ-specific variation, we detected a malonylglucoside SGA (**IV**) that was exclusively found in roots. Modifications like glycosylation and malonylation may increase the polarity of SGs and thereby their transportability, storability, and biological activity (Wolters *et al.* 2023). Malonylation is a process that is described for diterpene glycosides, a class of defence compounds that are known in other Solanaceae such as *Nicotiana attenuata* and *Capsicum* spp. (Heiling *et al.* 2010; Macel *et al.* 2019). Furthermore, specific decorations of 17-hydroxy-geranyl linalool by malonyl- and glycosylation are shown to solve the autotoxicity problem of diterpene-based defences in *N. attenuata* (Heiling *et al.* 2021). In addition, malonylated compounds may be more suitable for exudation into the rhizosphere, as root exudates are mostly polar, water-soluble compounds (Van Dam & Bouwmeester 2016). Lastly, one of the saturated SSGs (**VI**), was detected in flowering, but not in vegetative aboveground organs of *S*-chemotype plants, which suggests that this compound (class) may have a specific function in the interactions with pollinators. Further experiments are needed to infer the ecological functions of the various SGs in *S. dulcamara*.

We counted SGs for every annotated SA species. These SG counts allowed us to perform variation partitioning, thereby revealing the dynamics of SG chemodiversity among chemotypes and organs, and across ontogeny. These GLM-ASCA analyses corroborated the conclusions of the molecular network analyses. First, they confirmed that *m/z* 416.35, which was used to chemotype plants before chemical profiling, was clearly associated with chemotypic differences. Furthermore, we found that the numbers of SSGs with SA species *m/z* 415.32, 417.33 and 433.33 were higher in leaves than other organs, which we did not anticipate from results of previous studies (Calf *et al.* 2018; Chiocchio *et al.* 2023). In these previous studies, plants were sampled during the vegetative stage, which might explain why SSGs such as **V** and **VI** were previously not detected. In addition, our analyses showed that SG chemodiversity can also significantly vary across ontogeny. In particular, ontogenetic differentiation may also occur in roots, as evidenced by LV1 of the interaction term in the GLM-ASCA model that partitioned the variation between sample type and plant ontogeny. In addition to variation partitioning, the SG counts were also used to calculate SG richness and evenness, which served to compare SG chemodiversity in different chemotype–organ combinations. Interestingly, SG richness was higher in flowering than vegetative plants. This is in line with the common view that SGs are constitutive chemical defences to protect flowering and fruiting plants from leaf, and potential fitness, loss (Paudel *et al.* 2017; Panda *et al.* 2022). In vegetative plants, SG richness is significantly higher in leaves than in roots, while the opposite trend is observed in flowering plants. Considering that *S. dulcamara* is a perennial with overwintering roots, allocating more metabolites to the root may reflect patterns of optimal defence allocation (De Jong & Van Der 2000; van Dam & van der Meijden 2018). As genes related to the biosynthesis of SGs are expressed in both roots and shoots in both flowering and vegetative plants, differences in

root and shoot chemodiversity may be related to differences in transport dynamics. For glucosinolates, the expression of specific transporters in shoots and roots of plants are important drivers of differences in root and shoot glucosinolate profiles (Nour-Eldin *et al.* 2017). Whether similar mechanisms are regulating SG allocation in *S. dulcamara* has yet to be determined.

Chemodiveristy of SG in relation to expression of candidate *GAME* genes in *S. dulcamara*

We found that leaf SG chemotypes are not expressed similarly in roots of vegetative plants at the level of both SGs and transcripts. Surprisingly, when flowering, half of the *S*-chemotype plants showed SG profiles characteristic of *U*-chemotype plants in aboveground organs. We found that the *S*-chemotype plants which clustered with *U*-chemotype plants in the multivariate analyses also had lower expression levels of *SdGAME25* in their leaves compared to plants that retained a SG profile characteristic of *S*-chemotype plants. GAME25 catalyses the conversion of unsaturated into saturated SGs (Lee *et al.* 2019; Sonawane *et al.* 2018) and hence it was suggested that allelic variation in the *S*-chemotype would be the cause of the chemotypic differences between *S. dulcamara* leaves (Calf 2019). However, since the roots of *S*-chemotype plants also express *SdGAME25*, there must be another level at which this leaf chemical polymorphism is maintained. Furthermore, the relatively high expression levels of *SdGAME25* in leaves in the flowering *S*-chemotype correlates with the presence of saturated analogues of unsaturated SSGs, such as **V**. Sonawane *et al.* (2018) showed that overexpression of *GAME25* in eggplant not only increases saturated SGA production, but also results in the increased production of saturated SSGs. The concurrent increase of saturated SGAs and SSGs in leaves of the *S*-chemotype thus strongly suggests that *SdGAME25* plays an important role in maintaining SG chemodiversity at different times during the ontogeny of *S. dulcamara*.

The high levels of (unsaturated) SGs in leaves of the *U*-chemotype can be explained by differential expression of *SdGAME4* across chemotypes. When flowering, *SdGAME4* expression is lower in leaves of *U*- than *S*-chemotype plants. The enzyme coded for by *GAME4* catalyses the first dedicated step in SGA production. Therefore, a higher expression level of *SdGAME4* likely drives SG biosynthesis towards SGA biosynthesis, thereby downregulating SSG accumulation (Paudel *et al.* 2017). Indeed, RNA interference-mediated silencing of *SdGAME4* decreased SGA, and increased SSG accumulation in *S. lycopersicum* (Itkin *et al.* 2013), and *S. tuberosum* (Paudel *et al.* 2017). This is in line with our observation that roots of *U*-chemotype plants show lower expression levels of *SdGAME4* and higher abundance of SSG. The low *SdGAME4* may have redirected steroidal precursors into the SSG branch of the SG biosynthetic pathway. Furthermore, the relatively low *SdGAME4* expression in roots compared to leaves, explains why we previously found many SSGs in roots (Chiocchio *et al.* 2023), but not as many in leaves (Calf *et al.* 2018).

The enzyme encoded by *DPS* converts spirostanes into solanidanes in *S. tuberosum* through an initial C16α hydroxylation. Hydroxylation reactions of the steroidal aglycone lead to additional hydroxyl groups (Sonawane *et al.* 2022). These can potentially be glycosylated, thereby generating bidesmodic steroids, such as the putative compounds **V** and **VI** in *S.*

dulcamara. In the case where the product of *SdDPS* has similar catalytic activity as that of *StDPS* (Akiyama *et al.* 2021), then the differential expression of *SdDPS* in roots of *U*- and *S*-chemotypes may cause an additional level of SG diversity. In leaves of flowering plants, *SdDPS* expression is lower in *S*-chemotype than *U*-chemotype plants, while the opposite is observed for *SdGAME4* and *SdGAME9*. Specifically, *SdDPS* abundance is increased in leaves of flowering *U*-chemotype plants compare to vegetative plants. Taken together, this suggests that there might be a trade-off between expression of *SdGAME4* and *SdGAME9* on the one hand, and *SdDPS* on the other hand. Recently, other 2-oxoglutarate-dependent dioxygenases, such as GAME33 and GAME34, were associated with expansion of steroidal alkaloid structural diversity in *S. lycopersicum* and *S. habrochaites*, respectively (Sonawane *et al.* 2022). It is likely that 2-oxoglutarate-dependent dioxygenases have also driven expansion of SG chemodiversity in *S. dulcamara*. Lastly, the congruence among the expression patterns of different genes, in particular *SdGAME25*, *SdGAME9*, *SdGAME4* and *SdS5αR2*, suggests that these genes are co-regulated. Indeed, in tomato and potato genes related to SSG and SGA biosynthesis are organized into metabolic gene clusters (Cárdenas *et al.* 2015), which may be co-regulated by common transcription factors. Further studies, for example using gene-edited plants lacking one or more of these genes, are needed to fully understand the regulatory mechanisms of these different layers of intra-individual chemodiversity.

Implications for chemodiversity research

In conclusion, our analyses provide new insights into the extent and regulation of intraspecific and intra-individual SG chemodiversity in two *S. dulcamara* chemotypes by combining metabolic analyses with expression analyses of genes involved in the biosynthesis of SGs. Gene expression analyses associated transcripts abundances of candidate genes in SG metabolism with SG chemodiversity. The expression patterns of *SdGAME4*, *SdGAME25* and *SdDPS* were linked to chemotype-, organ- and ontogeny-specific intra-individual variation in SG chemodiversity. We used a homology-based approach for the gene expression analyses, assuming that these gene structure and functions are conserved among related *Solanum* spp. However, functional genetic analyses are needed to show that candidate genes are indeed the casual agents of SG chemodiversity among organs, ontogeny and chemotypes of *S. dulcamara*. Although the combination of MDN, chemical diversity and gene expression analyses provides new insights into the regulation of intraindividual SG chemodiversity in two *S. dulcamara* chemotypes, we realize that the current approach of quantifying SG chemodiversity in terms of richness and evenness may underestimate the total extent of structural SG diversity in *S. dulcamara*. Additional MS-based studies investigating structural SG chemodiversity in *S. dulcamara* would benefit from a tandem MS approach, allowing for spectral–database (Wang 2017) and compound–database (Dührkop *et al.* 2019) based dereplication and subsequent propagation of annotations (Ernst *et al.* 2019; Quinlan *et al.* 2022). Alternatively, NMR-based metabolomics approaches may provide complementary insights into structural SG diversity, especially for chemical evenness, as peak intensities in NMR are directly proportional to concentration of the metabolite. Since our work focused on chemical

diversity, we controlled for the effect of genetic diversity by studying clonally propagated plants from selected genotypes of known leaf chemotypes. To study whether differential regulation of *SdGAME25* across ontogeny indeed regulates SG chemodiversity, an experiment with plants grown from seed, rather than cuttings, would need to be performed. Combined with segregation pattern analysis of the chemotype, such an approach allows testing whether differential regulation of *SdGAME25* is controlled by a single locus across ontogeny in *S. dulcamara*. Considering that SG chemotype is heritable (Calf 2019) and related to differences in leaf herbivore and pathogen resistance (Calf *et al.* 2018, 2019; Sonawane *et al.* 2018; Wolters *et al.* 2023), our findings suggest that the existing intraspecific diversity in *S. dulcamara* may have resulted from differential selection pressures exerted by biotic interactors. In addition, the observed intra-individual chemodiversity suggests that aboveground and belowground chemodiversity may be regulated and selected for independently. We also found that over the course of ontogeny, different types of SG become more prominent, which may be an indication that other interactions, e.g. with pollinators, may be prioritized when plants are flowering. Our work highlights that phytochemical variation among organs and across ontogeny are important dimensions of chemodiversity that need to be considered in chemodiversity experiments. We hypothesize that such processes increase phytochemical dissimilarity in *S. dulcamara* populations, which, in turn, may increase individual plant fitness under field conditions.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

RAA and NMvD conceived and designed the study. NMvD prepared the F1 plants. RAA, IC, and BS performed greenhouse experiments. RAA and IC conducted phytochemical analyses using the method of FV. BS and RS performed gene expression. RAA and IC analysed LC–MS data. IC extracted, validated and counted steroidal glycosides. RAA applied mass-difference networking, built models and prepared figures. RAA wrote a draft manuscript with feedback from NMvD and FV. All authors provided feedback and agreed the final manuscript.

FUNDING INFORMATION

CONFLICT OF INTEREST

The authors declare no competing interests.

ETHICS STATEMENT

There were no studies with human and/or animal participants.

DATA AVAILABILITY STATEMENT

Data (including metadata) presented here are publicly available through Zenodo (10.5281/zenodo.11080314). Code will be made available through GitHub (https://github.com/redouanadam/CH1).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Supporting Information.

Figure S1. (a) Scree and (b, c) loadings plots of principal components analysis (PCA) of features (dots) from liquid-coupled to mass spectrometry (LC–MS).

Figure S2. Mass spectra of steroidal glycosides that vary across leaf chemotypes (I, II, III), organ type (IV) and ontogeny (steroidal saponin glycosides; V and VI).

Figure S3. Generalized linear model (GLM) based ANOVA–simultaneous component analysis (ASCA) of 3149 counted steroidal glycoside (SG) based on 12 steroidal aglycone (SA) species selected from mass-difference networking. GLM was modelled with the interaction between chemotype and tissue as fixed effect. (a) Scree plot showing latent variables (LV) and their explained variances. (b) Latent variables for interaction terms (LV1 and LV2), chemotype (LV1) and tissue (LV1 and LV2). Symbol shape represents organ-of-origin (triangle: adventitious roots, dot: leaves, square: stems) while symbol colour represents leaf chemotype (yellow: saturated SGs, blue: unsaturated SGs). (c) Loading bar plots showing relative importance of SA species in separation of treatments in the latent variables plot. Bar colours represent nature of the SA species (steroidal glycoalkaloids (SGA): orange; steroidal saponin glycosides (SSG): purple), bar colour shade is based on ring double bond equivalent (RDBE) of SA species.

Figure S4. Generalized linear model (GLM) based ANOVA–simultaneous component analysis (ASCA) of 3149 counted steroidal glycoside (SG) based on 12 steroidal aglycone (SA) species selected from mass-difference networking. GLM modelled with interaction between sample type and ontogeny as fixed effect. (a) Scree plot showing latent variables (LV) and their explained variances. (b) Latent variables for interactions term (LV1 and LV2), chemotype (LV1) and tissue (LV1 and LV2). Symbol shape represents ontogeny (diamonds: vegetative plants; asterisks: flowering plants). (c) Loading bar plots showing the relative importance of SA species in the separation of treatments in the latent variables plot. Bar colours represent nature of the SA species (steroidal glycoalkaloids (SGA): orange; steroidal saponin glycosides (SSG): purple), while bar shade is based on the ring double bond equivalent (RDBE) of the SA species.

Figure S5. Validation of GLM–ASCA model with interaction between chemotype and organ as fixed effect, showing (a) significance of model terms; (b) univariate model fit (pseudo-R^2) for every model variable; and (c) variable importance (Vector Norm) for every model variable. Only responsive variables with

higher pseudo-R^2 in the specified (blue) than in the null model (red) were retained for decomposition in the presented GLM–ASCA model (Figure S3).

Figure S6. Validation of GLM–ASCA model with interaction between sample type (organ-chemotype combination) and ontogeny as fixed effect, showing (a) significance of model terms; (b) univariate model fit (pseudo-R^2) for every model variable; and (c) variable importance (Vector Norm) for every model variable. Only responsive variables with higher pseudo-R^2 in the specified (blue) than the null model (red) were retained for decomposition in the presented GLM–ASCA model (Figure S4).

Figure S7. Total ion current for SG-associated features (TIC$_{SG}$; x-axes) in relation to predictions from linear mixed models (LMMs; see Fig. 5) for overall (a) and group-specific (b) Margalef richness, and overall (c) and group-specific (d) Pielou evenness. Black lines represent ±SE of mean. Symbol shape represents ontogeny (diamond: vegetative plants; asterisk: flowering plants); fill colour and line type represent organ-of-origin (light yellow triangle: adventitious roots, green dot: leaves, purple square: stems) and symbol outline colour represents leaf chemotype (yellow: saturated steroidal glycosides (SGs), blue: unsaturated SGs). Horizontal and vertical lines represent ±SE of mean for TIC$_{SG}$ and predicted SG indices, respectively. Significant Pearson correlation coefficients are shown with R$_L$ and R$_S$ for leaves and stems, respectively.

Table S1. Estimates, confidence intervals (CI), test statistic (χ^2), and P-values (P) of linear mixed models for Margalef's richness and Pielou's evenness for steroidal glycosides (SGs) found in extracts of different *S. dulcamara* organs. For Margalef's richness, fixed effects were modelled as the three-way interaction between 'leaf chemotype', 'organ', and 'ontogeny'.

Table S2. Analysis of deviance table (Type-III Wald χ^2-tests) for models presented in Figure 4 and Table S1. Test statistic (χ^2), and P-values (P).

REFERENCES

Akiyama R., Lee H.J., Nakayasu M., Osakabe K., Osakabe Y., Umemoto N., Saito K., Muranaka T., Sugimoto Y., Mizutani M. (2019) Characterization of steroid 5α-reductase involved in α-tomatine biosynthesis in tomatoes. *Plant Biotechnology*, **36**, 253–263. https://doi.org/10.5511/plantbiotechnology.19.1030a

Akiyama R., Watanabe B., Nakayasu M., Lee H.J., Kato J., Umemoto N., Muranaka T., Saito K., Sugimoto Y., Mizutani M. (2021) The biosynthetic pathway of potato solanidanes diverged from that of spirosolanes due to evolution of a dioxygenase. *Nature Communications*, **12**, 1300. https://doi.org/10.1038/s41467-021-21546-0

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bailly C. (2021) The steroidal alkaloids α-tomatine and tomatidine: panorama of their mode of action and pharmacological properties. *Steroids*, **176**, 108933. https://doi.org/10.1016/j.steroids.2021.108933

Bates D., Mächler M., Bolker B., Walker S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Brooks, M., E., Kristensen, K., Benthem, K., J.,van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H. J., Mächler, M., Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, **9**, 378. https://doi.org/10.32614/RJ-2017-066

Calf O.W. (2019) B'sweet or B'sour—An Ecological, Metabolomic and molecular analysis of slug resistance in *Solanum dulcamara* [PhD thesis, Radboud University] https://repository.ubn.ru.nl/bitstream/handle/2066/207517/207517.pdf?sequence=1

Calf O.W., Huber H., Peters J.L., Weinhold A., Poeschl Y., van Dam N.M. (2019) Gastropods and insects prefer different *Solanum dulcamara* chemotypes. *Journal of Chemical Ecology*, **45**, 146–161. https://doi.org/10.1007/s10886-018-0979-4

Calf O.W., Huber H., Peters J.L., Weinhold A., van Dam N.M. (2018) Glycoalkaloid composition explains variation in slug resistance in *Solanum dulcamara*. *Oecologia*, **187**, 495–506. https://doi.org/10.1007/s00442-018-4064-z

Calf O.W., Lortzing T., Weinhold A., Poeschl Y., Peters J.L., Huber H., Steppuhn A., van Dam N.M. (2020) Slug feeding triggers dynamic metabolomic and transcriptomic responses leading to induced resistance in *Solanum dulcamara*. *Frontiers in Plant Science*, **11**, 803. https://doi.org/10.3389/fpls.2020.00803

Cárdenas P.D., Sonawane P.D., Heinig U., Bocobza S.E., Burdman S., Aharoni A. (2015) The bitter side of the nightshades: genomics drives discovery in Solanaceae steroidal alkaloid metabolism. *Phytochemistry*, **113**, 24–32. https://doi.org/10.1016/j.phytochem.2014.12.010

Cárdenas P.D., Sonawane P.D., Pollier J., Vanden Bossche R., Dewangan V., Weithorn E., Tal L., Meir S., Rogachev I., Malitsky S., Giri A.P., Goossens A., Burdman S., Aharoni A. (2016) GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nature Communications*, **7**, 10654. https://doi.org/10.1038/ncomms10654

Cheng Q., Zeng L., Wen H., Brown S.E., Wu H., Li X., Lin C., Liu Z., Mao Z. (2023) Steroidal saponin profiles and their key genes for synthesis and regulation in *Asparagus officinalis* L. by joint analysis of metabolomics and transcriptomics. *BMC Plant Biology*, **23**, 207. https://doi.org/10.1186/s12870-023-04222-x

Chiocchio I., Andrés N.P., Anaia R.A., van Dam N.M., Vergara F. (2023) Steroidal glycoside profile differences among primary roots system and adventitious roots in *Solanum dulcamara*. *Planta*, **257**, 37. https://doi.org/10.1007/s00425-023-04072-9

Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. https://igraph.org

D'Agostino N., Golas T., van de Geest H., Bombarely A., Dawood T., Zethof J., Driedonks N., Wijnker E., Bargsten J., Nap J.-P., Mariani C., Rieu I. (2013) Genomic analysis of the native European *Solanum* species, *S. dulcamara*. *BMC Genomics*, **14**, 356. https://doi.org/10.1186/1471-2164-14-356

van Dam N.M., van der Meijden E. (2018) A role for metabolomics in plant ecology, *Annual plant reviews online*. Wiley, Chichester, UK, pp 87–107. https://doi.org/10.1002/9781119312994.apr0464

De Jong T.J., Van Der Meijden E. (2000) On the correlation between allocation to defence and regrowth in plants. *Oikos*, **88**, 503–508. https://doi.org/10.1034/j.1600-0706.2000.880305.x

Dührkop K., Fleischauer M., Ludwig M., Aksenov A.A., Melnik A.V., Meusel M., Dorrestein P.C., Rousu J., Böcker S. (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, **16**, 299–302. https://doi.org/10.1038/s41592-019-0344-8

Eich E. (2008) *Solanaceae and convolvulaceae - secondary metabolites: biosynthesis, chemotaxonomy, biological and economic significance: a handbook*. Springer, Berlin, Germany.

Ernst M., Kang K.B., Caraballo-Rodríguez A.M., Nothias L.-F., Wandy J., Chen C., Wang M., Rogers S., Medema M.H., Dorrestein P.C., van der Hooft J.J.J. (2019) MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites*, **9**, 144. https://doi.org/10.3390/metabo9070144

Franceschi, P. (2022). *gASCA: glm based ANOVA – Simultaneous component analysis (ASCA)*.

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31**, 3784–3788. https://doi.org/10.1093/nar/gkg563

Hartig F. (2022) DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models http://florianhartig.github.io/DHARMa/

Heiling S., Llorca L.C., Li J., Gase K., Schmidt A., Schäfer M., Schneider B., Halitschke R., Gaquerel E., Baldwin I.T. (2021) Specific decorations of 17-hydroxygeranyllinalool diterpene glycosides solve the autotoxicity problem of chemical defense in *Nicotiana attenuata*. *The Plant Cell*, **33**, 1748–1770. https://doi.org/10.1093/plcell/koab048

Heiling S., Schuman M.C., Schoettner M., Mukerjee P., Berger B., Schneider B., Jassbi A.R., Baldwin I.T. (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *The Plant Cell*, **22**, 273–292. https://doi.org/10.1105/tpc.109.071449

Heinig U., Aharoni A. (2014) Analysis of steroidal alkaloids and saponins in Solanaceae plant extracts using UPLC-qTOF mass spectrometry. In: Rodríguez-Concepción M. (Ed), *Plant isoprenoids*, Vol. **1153**.

Springer, New York, USA, pp 171–185. https://doi.org/10.1007/978-1-4939-0606-2_12

Hilker M. (2014) New synthesis: parallels between biodiversity and chemodiversity. *Journal of Chemical Ecology*, **40**, 225–226. https://doi.org/10.1007/s10886-014-0402-8

Itkin M., Heinig U., Tzfadia O., Bhide A.J., Shinde B., Cardenas P.D., Bocobza S.E., Unger T., Malitsky S., Finkers R., Tikunov Y., Bovy A., Chikate Y., Singh P., Rogachev I., Beekwilder J., Giri A.P., Aharoni A. (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**(6142), 175–179. https://www.science.org/doi/abs, https://doi.org/10.1126/science.1240230

Kessler A., Kalske A. (2018) Plant secondary metabolite diversity and species interactions. *Annual Review of Ecology, Evolution, and Systematics*, **49**, 115–138. https://doi.org/10.1146/annurev-ecolsys-110617-062406

Lee H.J., Nakayasu M., Akiyama R., Kobayashi M., Miyachi H., Sugimoto Y., Umemoto N., Saito K., Muranaka T., Mizutani M. (2019) Identification of a 3β-hydroxysteroid dehydrogenase/ 3-ketosteroid reductase involved in α-tomatine biosynthesis in tomato. *Plant and Cell Physiology*, **60**, 1304–1315. https://doi.org/10.1093/pcp/pcz049

Lee Y.-Y., Hashimoto F., Yahara S., Nohara T., Yoshida N. (1994) Steroidal Glycosides from *Solanum dulcamara*. *Chemical & Pharmaceutical Bulletin*, **42**, 707–709. https://doi.org/10.1248/cpb.42.707

Lenth R.V. (2024) Emmeans: estimated marginal means, aka least-squares means [Computer software] https://CRAN.R-project.org/package=emmeans

van Leur H., Raaijmakers C.E., van Dam N.M. (2006) A heritable glucosinolate polymorphism within natural populations of *Barbarea vulgaris*. *Phytochemistry*, **67**, 1214–1223. https://doi.org/10.1016/j.phytochem.2006.04.021

van Leur H., Vet L.E.M., van der Putten W.H., van Dam N.M. (2008) *Barbarea vulgaris* glucosinolate phenotypes differentially affect performance and preference of two different species of lepidopteran herbivores. *Journal of Chemical Ecology*, **34**, 121–131. https://doi.org/10.1007/s10886-007-9424-9

Macel M., Visschers I.G.S., Peters J.L., Kappers I.F., de Vos R.C.H., van Dam N.M. (2019) Metabolomics of thrips resistance in pepper (Capsicum spp.) reveals monomer and dimer acyclic diterpene glycosides as potential chemical defenses. *Journal of Chemical Ecology*, **45**, 490–501. https://doi.org/10.1007/s10886-019-01074-4

Margalef R. (1958) Information theory in ecology. *General Systems, Yearbook of the Society for General Systems Research*, **3**, 36–71.

Marion Z.H., Fordyce J.A., Fitzpatrick B.M. (2015) Extending the concept of diversity partitioning to characterize phenotypic complexity. *The American Naturalist*, **186**, 348–361. https://doi.org/10.1086/682369

Matz M.V., Wright R.M., Scott J.G. (2013) No control genes required: Bayesian analysis of qRT-PCR data. *PLoS One*, **8**, e71448. https://doi.org/10.1371/journal.pone.0071448

Moss G.P. (1998) Nomenclature of fused and bridged fused ring systems (IUPAC recommendations 1998). *Pure and Applied Chemistry*, **70**, 143–216. https://doi.org/10.1351/pac199870010143

Naake T., Fernie A.R. (2019) MetNet: metabolite network prediction from high-resolution mass spectrometry data in R aiding metabolite annotation.

Analytical Chemistry, **91**, 1768–1772. https://doi.org/10.1021/acs.analchem.8b04096

Nakayasu M., Shioya N., Shikata M., Thagun C., Abdelkareem A., Okabe Y., Ariizumi T., Arimura G., Mizutani M., Ezura H., Hashimoto T., Shoji T. (2018) JRE4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. *The Plant Journal*, **94**, 975–990. https://doi.org/10.1111/tpj.13911

Nour-Eldin H.H., Madsen S.R., Engelen S., Jørgensen M.E., Olsen C.E., Andersen J.S., Seynnaeve D., Verhoye T., Fulawka R., Denolf P., Halkier B.A. (2017) Reduction of antinutritional glucosinolates in Brassica oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, **35**, 377–382. https://doi.org/10.1038/nbt.3823

Oñate-Sánchez L., Vicente-Carbajosa J. (2008) DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Research Notes*, **1**, 93. https://doi.org/10.1186/1756-0500-1-93

Panda S., Jozwiak A., Sonawane P.D., Szymanski J., Kazachkova Y., Vainer A., Vasuki Kilambi H., Almekias-Siegl E., Dikaya V., Bocobza S., Shohat H., Meir S., Wizler G., Giri A.P., Schuurink R., Weiss D., Yasuor H., Kamble A., Aharoni A. (2022) Steroidal alkaloids defence metabolism and plant growth are modulated by the joint action of gibberellin and jasmonate signalling. *New Phytologist*, **233**, 1220–1237. https://doi.org/10.1111/nph.17845

Pang Z., Chong J., Zhou G., de Lima Morais D.A., Chang L., Barrette M., Gauthier C., Jacques P.-É., Li S., Xia J. (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*, **49**(W1), W388–W396. https://doi.org/10.1093/nar/gkab382

Paudel J.R., Davidson C., Song J., Maxim I., Aharoni A., Tai T.H.H. (2017) Pathogen and pest responses are altered due to RNAi-mediated knockdown of GLYCOALKALOID METABOLISM 4 in *Solanum tuberosum*. *Molecular Plant-Microbe Interactions*, **30**, 876–885. https://doi.org/10.1094/MPMI-02-17-0033-R

Petrén H., Anaia R.A., Aragam K.S., Bräutigam A., Eckert S., Heinen R., Jakobs R., Ojeda-Prieto L., Popp M., Sasidharan R., Schnitzler J.-P., Steppuhn A., Thon F., Tschikin S., Unsicker S.B., van Dam N.M., Weisser W.W., Wittmann M.J., Yepes S., Ziaja D., Müller C., Junker R.R. (2023a) Understanding the phytochemical diversity of plants: quantification, variation and ecological function. *bioRxiv*. https://doi.org/10.1101/2023.03.23.533415

Petrén H., Köllner T.G., Junker R.R. (2023b) Quantifying chemodiversity considering biochemical and structural properties of compounds with the R package chemodiv. *New Phytologist*, **237**, 2478–2492. https://doi.org/10.1111/nph.18685

Pielou E.C. (1966) The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, **13**, 131–144. https://doi.org/10.1016/0022-5193(66)90013-0

Quinlan Z.A., Koester I., Aron A.T., Petras D., Aluwihare L.I., Dorrestein P.C., Nelson C.E., Wegley Kelly L. (2022) ConCISE: consensus annotation propagation of ion features in untargeted tandem mass spectrometry combining molecular networking and in silico metabolite structure prediction. *Metabolites*, **12**, 1275. https://doi.org/10.3390/metabo12121275

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schreiber K., Rönsch H. (1965) Die Steroidalkaloide und -sapogenine chemisch unterschiedlicher Sippen von Solanum dulcamara L. LIII. Mitt.: solanumalkaloide. *Archiv der Pharmazie*, **298**, 285–293. https://doi.org/10.1002/ardp.19652980504

Shannon C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504. https://doi.org/10.1101/gr.1239303

Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D., Higgins D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539. https://doi.org/10.1038/msb.2011.75

Sonawane P.D., Heinig U., Panda S., Gilboa N.S., Yona M., Pradeep Kumar S., Alkan N., Unger T., Bocobza S., Pliner M., Malitsky S., Tkachev M., Meir S., Rogachev I., Aharoni A. (2018) Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus Solanum. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E5419–E5428. https://doi.org/10.1073/pnas.1804835115

Sonawane P.D., Jozwiak A., Barbole R., Panda S., Abebie B., Kazachkova Y., Gharat S.A., Ramot O., Unger T., Wizler G., Meir S., Rogachev I., Doron-Faigenboim A., Petreikov M., Schaffer A., Giri A.P., Scherf T., Aharoni A. (2022) 2-oxoglutarate-dependent dioxygenases drive expansion of steroidal alkaloid structural diversity in the genus Solanum. *New Phytologist*, **234**, 1394–1410. https://doi.org/10.1111/nph.18064

Sonawane P.D., Jozwiak A., Panda S., Aharoni A. (2020) 'Hijacking' core metabolism: a new panache for the evolution of steroidal glycoalkaloids structural diversity. *Current Opinion in Plant Biology*, **55**, 118–128. https://doi.org/10.1016/j.pbi.2020.03.008

Sonawane P.D., Pollier J., Panda S., Szymanski J., Massalha H., Yona M., Unger T., Malitsky S., Arendt P., Pauwels L., Almekias-Siegl E., Rogachev I., Meir S., Cárdenas P.D., Masri A., Petrikov M., Schaller H., Schaffer A.A., Kamble A., Giri A.P., Goossens A., Aharoni A. (2017) Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nature Plants*, **3**, 16205. https://doi.org/10.1038/nplants.2016.205

Thon F.M., Müller C., Wittmann M.J. (2024) The evolution of chemodiversity in plants—From verbal to quantitative models. *Ecology Letters*, **27**, e14365. https://doi.org/10.1111/ele.14365

Tsunoda T., Krosse S., van Dam N.M. (2017) Root and shoot glucosinolate allocation patterns follow optimal defence allocation theory. *Journal of Ecology*, **105**, 1256–1266. https://doi.org/10.1111/1365-2745.12793

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. (2012) Primer3—New capabilities and interfaces. *Nucleic Acids Research*, **40**, e115. https://doi.org/10.1093/nar/gks596

Van Dam N.M., Bouwmeester H.J. (2016) Metabolomics in the rhizosphere: tapping into belowground

chemical communication. *Trends in Plant Science*, **21**, 256–265. https://doi.org/10.1016/j.tplants.2016.01.008

Viswanathan D.V., Thaler J.S. (2004) Plant vascular architecture and within-plant spatial patterns in resource quality following herbivory. *Journal of Chemical Ecology*, **30**, 531–543. https://doi.org/10.1023/B:JOEC.0000018627.26420.e0

Wang M. (2017) Spectral Library Construction and Matching of MS/MS Spectra at Repository Scales [University of California, San Diego] https://escholarship.org/uc/item/08g414s7

Wang S., Alseekh S., Fernie A.R., Luo J. (2019) The structure and function of major plant metabolite modifications. *Molecular Plant*, **12**, 899–919. https://doi.org/10.1016/j.molp.2019.06.001

Wetzel W.C., Whitehead S.R. (2020) The many dimensions of phytochemical diversity: linking theory to practice. *Ecology Letters*, **23**, 16–32. https://doi.org/10.1111/ele.13422

Wickham H. (2016) *ggplot2: elegant graphics for data analysis*. Springer, New York, USA. https://ggplot2.tidyverse.org

Willuhn G. (1966) UNTERSUCHUNGEN ZUR CHEMISCHEN DIFFERENZIERUNG BEI SOLANUM DULCAMARA L. – I. Genetische Fixierung der unterschiedlichen Steroidalkaloidführung 1. *Planta Medica*, **14**, 408–420. https://doi.org/10.1055/s-0028-1100068

Wolters P.J., Wouters D., Tikunov Y.M., Ayilalath S., Kodde L.P., Strijker M.F., Caarls L., Visser R.G., Vleeshouwers V.G. (2023) Tetraose steroidal glycoalkaloids from potato provide resistance against *Alternaria solani* and Colorado potato beetle. *eLife*, **12**, RP87135. https://doi.org/10.7554/eLife.87135

You Y., van Kan J.A.L. (2021) Bitter and sweet make tomato hard to (b)eat. *New Phytologist*, **230**, 90–100. https://doi.org/10.1111/nph.17104

Zhao D.-K., Zhao Y., Chen S.-Y., Kennelly E.J. (2021) *Solanum* steroidal glycoalkaloids: structural diversity, biological activities, and biosynthesis. *Natural Product Reports*, **38**, 1423–1444. https://doi.org/10.1039/D1NP00001B