What is a plant chemotype anyway?

Caroline Müller[1,2*], **Thomas Dussarrat**[1], **Nicole M. van Dam**[3,4]

[1]Chemical Ecology, Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany

[2]Joint Institute for Individualisation in a Changing Environment (JICE), University of Münster and Bielefeld University, Bielefeld, Germany

[3]Leibniz Institute of Vegetable and Ornamental Crops, Theodor-Echtermeyer-Weg 1, 14979 Großbeeren, Germany

[4]Institute of Biodiversity, Ecology and Evolution (IBEE), Dornburgerstraße 159, 07743 Jena, Germany

* Correspondence:

caroline.mueller@uni-bielefeld.de (C. Müller)

https://www.uni-bielefeld.de/fakultaeten/biologie/forschung/arbeitsgruppen/chem_eco/

Orcid-IDs:

Caroline Müller - https://orcid.org/0000-0002-8447-534X

Thomas Dussarrat - https://orcid.org/0000-0001-6245-3652

Nicole M. van Dam - https://orcid.org/0000-0003-2622-5446

KEYWORDS

Chemical family, chemodiversity, chemotype, intraspecific variation, organ-specificity, metabolomics

ABSTRACT

Many plant species show chemical polymorphisms regarding the composition of specialized metabolites belonging to certain chemical families. This led to the classification of

chemotypes, that is, groups of plants that can be distinguished by their chemical profiles of metabolites within one chemical family. We present existing definitions and approaches for classifying chemotypes, and describe factors determining them. We argue that it should always be made explicit on which organ the chemotype specification is based, because chemical profiles can differ among organs. Moreover, the chemical family needs to be explicitly stated, as plants may be grouped differently when other metabolites are taken into account. We argue that gaining more knowledge on chemotypes is of high relevance for basic and applied science.

MAIN TEXT

Chemotypes and their terminology

Within various species, different chemotypes can be distinguished. These chemotypes are based on distinct profiles expressed within certain biochemical pathways, or chemical families. The term “chemotype” was first used for a *Drosophila* mutant lacking xanthine dehydrogenase activity, leading to a maroon eye color [1]. The term has also been used for bacteria, such as *Salmonella typhi* strains, that either can or cannot attack D(+)-xylose [2]. However, these early examples refer to the presence or absence of an enzyme, rather than profiles of certain chemical families. In plants, the German term “*Chemische Rassen*” (chemical races) was introduced to distinguish individuals of medicinal plants or crops with different profiles of essential oils, for example, in *Tanacetum vulgare* [3] or *Daucus carota* [4], or in phenylpropanoid derivatives, found in *Petroselinum* [5]. The German term “chemische Sippen” (chemical clans) was applied to differentiate between chemotypes of *Solanum dulcamara* differing in steroidal alkaloid and sapogenin profiles during fruit development [6]. Other famous examples of chemotypes are the distinct occurrence of steroidal lactones in *Withania somnifera* [7], of pyrrolizidine alkaloids in *Senecio* species [8], of glucosinolates in different Brassicaceae [9-11], and of essential oils in various spices, such as *Thymus* [12], in which dominant monoterpenoids cause a distinct smell and taste. In other plant species, the complex interplay between both substrates and enzymes determines the presence of toxic

metabolites, such as in *Trifolium* species that can be cyanogenic or not, giving rise to the term “cyanotypes” [13]. Similar terms such as “chemovarieties” [14] or “chemical phenotypes” [15] have been used as well. Chemical differences can also coincide with morphological differences, which led to using the term “morphochemotypes”, such as in *Annona emarginata* [16]. The term “metabotype” was coined to describe distinct metabolic phenotypes that refer to numerous metabolites of different families, explored in untargeted metabolic fingerprinting approaches [17]. The diversity of terms and meanings calls for a more unified terminology.

Assignment of chemotypes

Various approaches are used to assign chemotypes. Often, the percentage of a major metabolite in relation to all metabolites of that chemical family is considered. For example, in *T. vulgare*, “mono-chemotypes”, in which the main terpenoid accounts for 41-99 % of the profile, are distinguished from “mixed chemotypes”, in which one to three additional (satellite) terpenoids contribute to this amount [18]. Alternatively, different statistical methods taking all metabolites within a chemical family into account can be applied. For example, using hierarchical cluster analysis, 21 chemotypes of *T. vulgare* were detected across Europe and North America [19]. Using a similar approach, 121 cultivars of *Piper methysticum* were assigned to six chemotypes [20]. Using principal component analysis (PCA), steroidal glycoside chemotypes were determined for *S. dulcamara* [21], cannabinoid chemotypes for commercial *Cannabis* samples [15], and essential oil chemotypes in *Crithmum maritimum* [22]. With non-metric multidimensional scaling (NMDS), glucosinolate chemotypes in *Arabidopsis halleri* and *Bunias orientalis* were discriminated [10, 11]. Finally, heatmaps can be used to depict differences among chemical families in chemotypes or metabotypes [17, 23]. The use of different classification methods may result in chemotypes being assigned differently. Polatoğlu [14] proposed a nomenclature where the frequency of chemotype occurrence in a given location is also considered. This requires the full sampling of all individuals within this given population and limits comparisons with other populations. Moreoevere, the chemotype composition can also vary substantially among populations [24,

25]. In order to stringently assign chemotypes, ideally multiple populations should be screened completely. Considering that this is not possible, splitting the available dataset into training, testing and validation sets, can help to identify robust chemotypes [26]. This procedure would also prevent overfitting by the models that are used. While all methods are legitimate, it should always be clearly stated how chemotypes were determined.

Determinants of an individual's chemotype

The chemotype of each individual is determined by several internal and external factors (Fig. 1), with the (epi)genome being central. Only if the gene coding for a specific transcription factor or an enzyme involved in the biosynthesis of a particular metabolite is present, transcribed, and functional, the metabolite can be produced [27, 28]. Whereas there is evidence for epigenetic regulation in specialized metabolism, for example, for terpenoid biosynthesis in *Arabidopsis thaliana* and monoterpene indole alkaloid synthesis in *Cantharanthus roseus* [28], little is known about how epigenetic regulation may contribute to chemotype formation. The second internal layer determining chemotypes is the plant's physiology. Genes involved in metabolite synthesis may show chemotype- [29] as well as organ-specific expression patterns [30]. For example, in *Senecio vulgaris* pyrrolizidine alkaloids are synthesized in the roots and transported to the shoots [31]. For glucosinolates, the typical defenses of Brassicaceae, there are specific transporter proteins which are responsible for the allocation of different classes of glucosinolates across the plant [32]. Other metabolites are emitted into the air or into the rhizosphere, which is likewise regulated by (specific) transporter proteins [33, 34]. Next to transport, storage is important. Many metabolites are stored in specific cells or structures, such as terpenoids in trichomes [35], or glucosinolates or alkaloids in the vacuoles [31, 36].

These internal physiological processes are further modulated by external abiotic and biotic factors. For example, temperature, water stress, and ultraviolet (UV) light are all known to elicit the production of specific metabolites that should reduce damage to the plant, such as proline after drought [37], and phenolics in response to UV exposure [38]. Similarly, attacks

by herbivores or pathogens can trigger induced defense responses that affect the plant's metabolome [39]. Such environmental factors, alone or in combination, may promote the biosynthesis of particular metabolites within a chemical family. For instance, drought and herbivory induce indole glucosinolates in *Arabidopsis thaliana* [40]. When unnoticed, differences in the intensity of such challenges may lead to chemotype assignments that are not visible or robust under all environmental conditions.

Finally, plant metabolomes may be affected by temporal processes, such as time of day, ontogeny, and the season. In particular, plant volatile emissions vary over the day, due to the availability of sunlight, or due to the availability of mutualists, such as pollinators [41]. Across ontogeny, the expression of genes and the resulting glycoalkaloid chemotype of *S. dulcamara* plants differed between vegetative and flowering stages [42]. Shifts in chemotypes have also been found between juvenile and mature leaves of *Musa* spp. [23] and across the season in *Conyza bonariensis* [43]. Despite these various influences on plant chemical profiles, individual chemotypes commonly remain distinguishable [3, 44, 45].

Differences within chemical families among organs

Chemotypes are usually determined based on the metabolite composition of one organ, most often the leaves. However, more in-depth studies revealed that the metabolite profiles within a chemical family can differ among organs. For example, *Barbarea vulgaris* has two chemotypes based on leaf composition, one dominated by 2-phenylethylglucosinolate (NAS type), the other by the hydroxylated form, (*S*)-2-hydroxy-2-phenylethylglucosinolate (BAR-type) [9]. Within each chemotype, this respective glucosinolate also dominates in the seeds and flowers. In contrast, both glucosinolates occur in comparable amounts in the roots of BAR-type plants, meaning that chemotypes can no longer be distinguished there. In *T. vulgare,* several chemotypes are distinguishable by their leaf monoterpenoid profiles, which are mostly also reflected in the flower heads [46], while the profiles are very distinct in the roots. In roots, fewer terpenoids, mostly sesquiterpenoids, and no separation into distinct clusters are found [47, 48] (Fig. 2A). This may be due to the distinct localization of the

respective biosynthetic pathways: while monoterpenoid biosynthesis is mostly taking place in the plastids, sesquiterpenoids are formed in the cytosol [49]. In addition to the biosynthetic machinery, the eco-physiological function of different terpenoids may determine this allocation pattern; monoterpenoids are more volatile than sesquiterpenoids, and therefore better suited to mediate interactions with other organisms in the air than in the soil, and *vice versa*. Differences in terpenoid profiles between organs were also found in other species, such as *Smyrnium olusatrum* [50] and *Limoniastrum guyonianum* [51]. In these species, the term chemotype was even used to distinguish between the profiles of these organs within individuals. Different steroidal glycoside chemotypes could also be determined in *S. dulcamara*, with striking differences among leaves [45], but less in roots [52] (Fig. 2B). Even within an organ, the metabolite composition can differ, as revealed for terpenoid profiles across different root sections in a metabolic atlas for *T. vulgare* [53]. Similarly, root parts of *Brassica* species differed in their glucosinolate profiles, with 2-phenylethylglucosinolate dominating the profile of tap roots, whereas indole glucosinolates were more prominent in the fine roots [54]. These findings underscore the need to explicitly state on which organ or plant part the chemotype assignment is based.

Differences within organs between chemical families

A largely neglected aspect is that even within organs potentially distinct chemotypes can be found, depending on which chemical family is considered. Using an existing metabolomics dataset from leaves of five terpenoid chemotypes of *T. vulgare* [55], we show that these chemotypes could also be predicted from alkaloids and fatty acids with over 70% accuracy (Fig. 3A). Moreover, fatty acids predicted alkaloid and flavonoid clusters, and *vice versa*, revealing strong cross-family co-variation. Despite these high levels of co-variation, the same chemical families can define additional and (partially) independent chemotypes in the same organ. For example, cluster analyses revealed different numbers of chemotypes if alkaloids (three), fatty acids (four) or flavonoids (four) were considered, compared to the five terpenoid chemotype clusters (Fig. 3B-D).

Overall, an organ's chemistry reflects the integrated interplay of multiple chemotypes whose chemical building blocks co-vary to some extent. This may be explained by different gene expression patterns and transcription factors regulating distinct biosynthetic pathways in parallel [56]. Genes coding for different chemical families may also be located on different chromosomes. In tomato, the GlycoAlkaloid MEtabolism (GAME) genes, involved in the synthesis of steroidal glycoalkaloids, are clustered on two chromosomes [57], while the genes coding for terpenoid synthases are clustered on five other chromosomes [58]. This means that chemotypes in these two chemical families may evolve independently of each other, and that alkaloid chemotypes may not predict terpenoid chemotypes very well. At the same time, some chemical families are closely linked via shared biosynthetic pathways, for example terpene indole-alkaloids in *Catharantus roseus* [59]. Under such conditions, it may be more likely that terpenoid chemotypes can be predicted by alkaloid chemotypes. However, little is known on the underlying mechanisms of co-variation or coupling versus decoupling of chemotype formation, highlighting the need for more research in this direction.

The how and why of differences in chemotypes

A central premise of evolutionary theory is that selection acts on the phenotype. Considering that plant metabolites are important for interactions with the environment, these interactions likely contribute to the emergence and maintenance of chemodiversity [60, 61] and diverse chemotypes within plant species. Both abiotic and biotic factors vary over time and space, which may also explain the above-mentioned differences among organs and life stages. Roots grow in the dark and dense soil, where they are confronted with a large diversity of beneficial and harmful micro- and macro-organisms in the rhizosphere [54]. The physicochemical properties of soils may impact the types of metabolites that perform best. For example, 2-phenylethylglucosinolate, which is particularly prominent in *Brassica* tap root profiles [54], yields breakdown products that are less volatile and more toxic in solid medium than those of other glucosinolates that are more prominent in leaves [62]. Thus, the distinct root and shoot chemical profiles may well result from differential selection pressures exerted

by different organisms. Interactions also vary over ontogeny, in particular when plants start to flower and pollinators must be attracted, again modulating plant chemistry [48].

Selection pressures also vary over larger spatial and temporal scales. This variation may contribute to the maintenance of different chemotypes within a plant species. In particular wind- or bird-dispersed seeds may germinate far away from their mother plant, where environmental conditions may be completely different. If all plants were of the same chemotype, the species may fail in establishing itself. For example, different herbivore spectra were found on different chemotypes of *T. vulgare* and *S. dulcamara* within populations, with some being more, some less resistant to certain herbivore species [21, 63]. Because the frequency of these species can vary in time and space, having multiple chemotypes enhances the chances of survival of the species. Also, associational resistance may reduce herbivore pressure if plants of different chemotypes grow in close proximity [63].

Relevance of chemotypes in applied fields

Distinguishing among chemotypes within a species depicts an important part of chemodiversity. Next to its ecological consequences, chemotypic variation is also of high relevance from an applied perspective. The exact chemical composition and chemotype of a plant are, for example, important for their medicinal value. The different chemotypes for cannabinoids and terpenoids of *Cannabis* determine their psycho-activity [15] while chemotypes of *W. somnifera* differing in their composition of whitanolides and other metabolites differ in their pharmacological activity [30]. Likewise, chemotypes need to be considered for plants potentially used as pesticides. For example, in *Tephrosia vogelii* only the chemotype containing rotenoids is bioactive against insects [64].

The chemical profiles of crop plants determine their value as human food or animal feed as well as their level of resistance to herbivores and pathogens. Sometimes these aspects may conflict. Metabolites acting as defenses against pests can also render the crop unpalatable and even toxic. Therefore, crop breeders have selected certain chemotypes with low levels of these metabolites. An example are the double zero canola (*Brassica napus*) varieties,

which were bred to have low erucic acid and glucosinolate levels, turning the oil safe for human consumption and the meal for animal feed [65]. At the same time, these varieties are more susceptible to slug herbivory than their wild relatives with higher levels of glucosinolates [66]. Chemodiversity has also dropped inadvertently, leading to higher susceptibility to antagonists. For example, most American maize varieties lost the ability to produce *E*-β-caryophyllene in their roots, which reduced their ability to attract entomopathogenic nematodes to herbivore-damaged roots [67]. In view of growing concerns regarding pesticide toxicity and resistance development in pest organisms [68], older chemotypes or landraces should be revisited to breed pest-resistant plants. This not only requires knowledge of the efficacy of different chemical families [69], but also knowledge on the genetic and regulatory mechanisms determining chemotypes, including temporal and spatial allocation patterns of metabolites within crops.

Besides, exploring chemodiversity through chemotypes represents a step forward in efforts to preserve chemical functions in ecosystems. In an ecosystem, the distribution of chemical families, and therefore chemotypes, is linked to the identity and coverage of plant species and to the environment [70]. Understanding and, ultimately, predicting loss or gain of chemotypes under climate change is another dimension and objective for preserving ecosystem services and protect biodiversity.

Concluding remarks

Overall, the chemotype concept is useful, as it helps us to analyze how chemotypes emerge and are maintained in natural plant populations, what roles they play in wild and cultivated plant species and how chemotype variation may enhance resilience of plant populations. Explicit mentioning of the organ, the chemical family, and the (statistical) method on which the chemotype is defined is needed. Otherwise, one is left with the question what a chemotype is anyway.

Acknowledgements: We thank Dominik Ziaja, Valeria Mendoza, and Paula Bueno for drawing parts of Fig. 2. We thank the Deutsche Forschungsgemeinschaft (DFG) for funding the research unit FOR 3000, with funds for MU1829/28-2, MU 1829/29-2, and DA1201/10-2.

REFERENCES

1. Hubby, J.L. and Forrest, H.S. (1960) Studies on the mutant maroon-like in *Drosophila melanogaster*. Genetics 45, 211-224.
2. Nicolle, P. et al. (1961) Recherches sur le comportement fermentatif des bacilles typhiques a legard du xylose naturel D(+) et de son inverse optique L(-). Annales de l'Institut Pasteur 101, 211-+.
3. Stahl, E. and Schmitt, G. (1964) Chemische Rassen bei Arzneipflanzen. 2. Mitt.: über die verschiedenartig zusammengesetzten ätherischen Öle des Rainfarns. Archiv der Pharmazie und Berichte der Deutschen Pharmazeutischen Gesellschaft 297, 385-391.
4. Stahl, E. (1964) Chemische Rassen bei Arzneipflanzen. 3. Mitt.: Die unterschiedliche Zusammensetzung des ätherischen Öls der Früchte von Kultur- und Wildmöhren (*Daucus carota* L. s. l.). Archiv der Pharmazie und Berichte Der Deutschen Pharmazeutischen Gesellschaft 297, 500-511.
5. Stahl, E. and Jork, H. (1964) Chemische Rassen bei Arzneipflanzen. 1 Mitt.: Untersuchung der Kulturvarietäten europäischer Petersilienherkünfte. Archiv der Pharmazie und Berichte der Deutschen Pharmazeutischen Gesselschaft 297, 273-281.
6. Willuhn, G. (1967) Untersuchungen zur chemische Differenzierung bei *Solanum dulcamara* L. 2. Der Steroidgehalt in Früchten verschiedener Entwicklungsstadien der Tomatidenol- und Soladulcidin-Sippe. Planta Medica 15, 58-+.
7. Abraham, A. et al. (1968) A chemotaxonomic study of Withania somnifera (L) Dun. Phytochemistry 7, 957-&.
8. Witte, L. et al. (1992) Chemotypes of 2 pyrrolizidine alkaloid-containing *Senecio* species. Phytochemistry 31, 559-565.
9. van Leur, H. et al. (2006) A heritable glucosinolate polymorphism within natural populations of *Barbarea vulgaris*. Phytochemistry 67, 1214-1223.
10. Tewes, L.J. et al. (2018) Intracontinental plant invader shows matching genetic and chemical profiles and might benefit from high defence variation within populations. Journal of Ecology 106, 714–726.
11. Kazemi-Dinan, A. et al. (2015) Is there a trade-off between glucosinolate-based organic and inorganic defences in a metal-hyperaccumulator in the field? Oecologia 178, 369-378.
12. Linhart, Y.B. and Thompson, J.D. (1999) Thyme is of the essence: Biochemical polymorphism and multi-species deterrence. Evolutionary Ecology Research 1, 151-171.

13. Tillbottraud, I. et al. (1988) Variable phenotypes and stable distribution of the cyanotypes of *Trifolium repens* L in Southern France. Acta Oecologica-Oecologia Plantarum 9, 393-404.
14. Polatoğlu, K. (2013) “Chemotypes”– a fact that should not be ignored in natural product studies. The Natural Products Journal 3, 10-14.
15. Smith, C.J. et al. (2022) The phytochemical diversity of commercial *Cannabis* in the United States. Plos One 17, 33.
16. Mimi, C.O. et al. (2021) Chemophenetics as a tool for distinguishing morphotypes of *Annona emarginata* (Schltdl.) H. Rainer. Chemistry & Biodiversity 18, e202100544.
17. Clancy, M.V. et al. (2018) Metabotype variation in a field population of tansy plants influences aphid host selection. Plant Cell and Environment 41, 2791-2805.
18. Holopainen, M. et al. (1987) A study on tansy chemotypes. Planta Medica 53, 284-287.
19. Wolf, V.C. et al. (2011) High chemical diversity of a plant species is accompanied by increased chemical defence in invasive populations. Biological Invasions 13, 2091–2102.
20. Lebot, V. and Levesque, J. (1996) Genetic control of kavalactone chemotypes in *Piper methysticum* cultivars. Phytochemistry 43, 397-403.
21. Calf, O.W. et al. (2019) Gastropods and insects prefer different *Solanum dulcamara* chemotypes. Journal of Chemical Ecology 45, 146-161.
22. Jallali, I. et al. (2023) *Crithmum maritimum* L. Volatile compound's diversity through Tunisian populations: use of a plant organ-based statistical approach for chemotype identification. Chemistry & Biodiversity 20, 10.
23. Drapal, M. et al. (2019) Metabolite profiling characterises chemotypes of *Musa* diploids and triploids at juvenile and pre-flowering growth stages. Scientific Reports 9, 4657.
24. Wolf, V.C. et al. (2012) Genetic and chemical variation of *Tanacetum vulgare* in plants of native and invasive origin. Biological Control 61, 240–245.
25. Pormetter, L. et al. (2025) Glucosinolate diversity in seven field-collected Brassicaceae species. PloS ONE 20 (11), e0336172.
26. Dussarrat, T. et al. (2022) Predictive metabolomics of multiple Atacama plant species unveils a core set of generic metabolites for extreme climate resilience. New Phytologist 234 (5), 1614-1628.
27. Beekwilder, J. et al. (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. Plos One 3, e2068.
28. Méteignier, L.V. et al. (2023) Emerging mechanistic insights into the regulation of specialized metabolism in plants. Nature Plants 9, 22-30.
29. Padovan, A. et al. (2013) Differences in gene expression within a striking phenotypic mosaic *Eucalyptus* tree that varies in susceptibility to herbivory. BMC Plant Biology 13, 29.

30. Gupta, P. et al. (2011) Differential expression of farnesyl diphosphate synthase gene from *Withania somnifera* in different chemotypes and in response to elicitors. Plant Growth Regulation 65, 93-100.
31. Hartmann, T. and Dierich, B. (1998) Chemical diversity and variation of pyrrolizidine alkaloids of the senecionine type: biological need or coincidence? Planta 206, 443-451.
32. Nour-Eldin, H.H. et al. (2012) NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. Nature 488, 531–534.
33. Liao, P. et al. (2023) Emission of floral volatiles is facilitated by cell-wall non-specific lipid transfer proteins. Nature Communications 14, 330.
34. Ziegler, J. et al. (2017) Arabidopsis transporter ABCG37/PDR9 contributes primarily highly oxygenated coumarins to root exudation. Scientific Reports 7, 3704.
35. Jakobs, R. and Müller, C. (2019) Volatile, stored and phloem exudate-located compounds represent different appearance levels affecting aphid niche choice. Phytochemistry 159, 1-10.
36. Kopriva, S. and Gigolashvili, T. (2016) Chapter Five - Glucosinolate synthesis in the context of plant metabolism. Advances in Botanical Research, 80, 99-124.
37. Ahuja, I. et al. (2010) Plant molecular stress responses face climate change. Trends in Plant Science 15, 664-674.
38. Kuhlmann, F. and Müller, C. (2011) Impact of ultraviolet radiation on interactions between plants and herbivorous insects: a chemo-ecological perspective. Progress in Botany 72, 305-347.
39. Karban, R. and Myers, J.H. (1989) Induced plant responses to herbivory. Annual Review of Ecology and Systematics 20, 331-348.
40. Pineda, A. et al. (2016) Negative impact of drought stress on a generalist leaf chewer and a phloem feeder is associated with, but not explained by an increase in herbivore-induced indole glucosinolates Environmental and Experimental Botany 123, 88-97.
41. Schuman, M.C. et al. (2016) Temporal dynamics of plant volatiles: mechanistic bases and functional consequences. In Deciphering Chemical Language of Plant Communication (Blande, J.D. and Glinwood, R. eds), pp. 3-34, Springer International Publishing Ag.
42. Anaia, R.A. et al. (2025) Ontogeny and organ-specific steroidal glycoside diversity is associated with differential expression of steroidal glycoside pathway genes in two *Solanum dulcamara* leaf chemotypes. Plant Biology 27, 651-668.
43. Mabrouk, S. et al. (2011) Chemical composition of essential oils from leaves, stems, flower heads and roots of *Conyza bonariensis* L. from Tunisia. Natural Product Research 25, 77-84.

44. Kleine, S. and Müller, C. (2011) Intraspecific plant chemical diversity and its relation to herbivory. Oecologia 166, 175-186.
45. Calf, O.W. et al. (2020) Slug feeding triggers dynamic metabolomic and transcriptomic responses leading to induced resistance in *Solanum dulcamara*. Frontiers in Plant Science 11, 803.
46. Sasidharan, R. et al. (2024) Intraspecific plant chemodiversity at the individual and plot levels influences flower visitor groups with consequences for germination success. Functional Ecology 38, 2665-2678.
47. Kleine, S. and Müller, C. (2013) Differences in shoot and root terpenoid profiles and plant responses to fertilisation in *Tanacetum vulgare*. Phytochemistry 96, 123-131.
48. Ziaja, D. and Müller, C. (2025) Intraspecific and intra-individual chemodiversity and phenotypic integration of terpenes across plant parts and developmental stages in an aromatic plant species. Plant Biology 27, 637-650.
49. Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Current Opinion in Plant Biology 9, 297-304.
50. Maggi, F. et al. (2015) Essential oil chemotypification and secretory structures of the neglected vegetable *Smyrnium olusatrum* L. (Apiaceae) growing in central Italy. Flavour and Fragrance Journal 30, 139-159.
51. Hammami, S. et al. (2011) Chemical analysis and antimicrobial effects of essential oil from *Limoniastrum guyonianum* growing in Tunisia. Journal of Medicinal Plants Research 5, 2540-2545.
52. Chiocchio, I. et al. (2023) Steroidal glycoside profile differences among primary roots system and adventitious roots in *Solanum dulcamara*. Plants 257, 37.
53. Rahimova, H. et al. (2025) Exogenous stimulation of *Tanacetum vulgare* roots with pipecolic acid leads to tissue-specific responses in terpenoid composition. Plant Biology 27, 891-902.
54. Tsunoda, T. et al. (2017) Root and shoot glucosinolate allocation patterns follow optimal defence allocation theory. Journal of Ecology 105, 1256-1266.
55. Dussarrat, T. et al. (2023) Influences of chemotype and parental genotype on metabolic fingerprints of tansy plants uncovered by predictive metabolomics. Scientific Reports 13, 11645.
56. Shi, M. et al. (2024) Molecular regulation of the key specialized metabolism pathways in medicinal plants. Journal of Integrative Plant Biology 66, 510-531.
57. Cárdenas, P.D. et al. (2015) The bitter side of the nightshades: Genomics drives discovery in Solanaceae steroidal alkaloid metabolism. Phytochemistry 113, 24-32.
58. Falara, V. et al. (2011) The tomato terpene synthase gene family. Plant Physiology 157, 770-789.

59. O'Connor, S.E. and Maresh, J.J. (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. Natural Product Reports 23, 532-547.
60. Thon, F.M. et al. (2024) Evolution of chemodiversity – From verbal to quantitative models. Ecology Letters 27, e14365.
61. Wittmann, M.J. and Bräutigam, A. (2024) How does plant chemodiversity evolve? Testing five hypotheses in one population genetic model. New Phytologist.
62. Kirkegaard, J.A. and Sarwar, M. (1998) Biofumigation potential of brassicas - I. Variation in glucosinolate profiles of diverse field-grown brassicas. Plant and Soil 201, 71-89.
63. Ziaja, D. and Müller, C. (2023) Intraspecific chemodiversity provides plant individual- and neighbourhood-mediated associational resistance towards aphids. Frontiers in Plant Science 14, 1145918.
64. Mkindi, A.G. et al. (2019) Phytochemical analysis of *Tephrosia vogelii* across East Africa reveals three chemotypes that influence its use as a pesticidal plant. Plants 8, 597.
65. Azhar, M. et al. (2025) A brief history of canola genetic gains: from classical breeding to genome editing. Physiologia Plantarum 177, e70644.
66. Baaij, B.M. et al. (2018) Slug herbivory on hybrids of the crop *Brassica napus* and its wild relative *B. rapa*. Basic and Applied Ecology 31, 52-60.
67. Köllner, T.G. et al. (2008) A maize (*E*)-β-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. Plant Cell 20, 482-494.
68. Gensch, L. et al. (2024) Pesticide risk assessment in European agriculture: Distribution patterns, ban-substitution effects and regulatory implications. Environmental Pollution 348, 123836.
69. Whitehead, S.R. and Poveda, K. (2019) Resource allocation trade-offs and the loss of chemical defences during apple domestication. Annals of Botany 123, 1029-1041.
70. Defossez, E. et al. (2021) Spatial and evolutionary predictability of phytochemical diversity. Proceedings of the National Academy of Sciences of the United States of America 118, e2013344118.

FIGURES

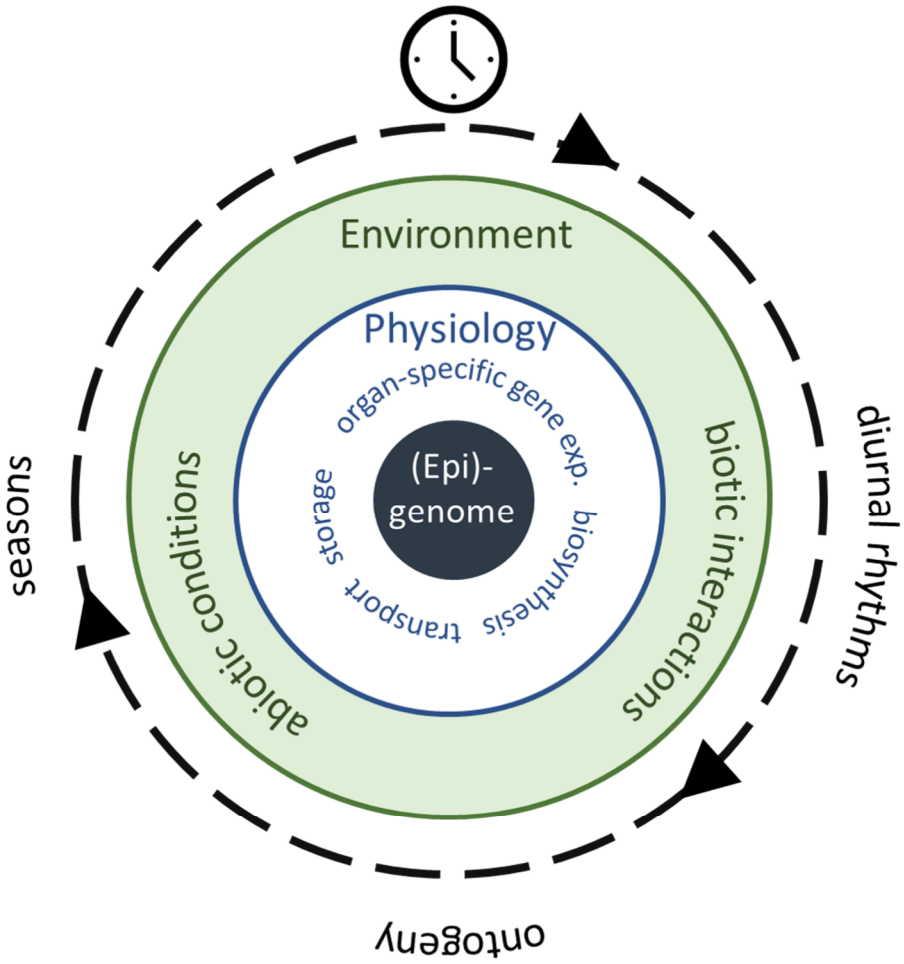

Figure 1: Intrinsic and external factors together determine the observed chemotype of an individual plant. From the inside out: 1) the (epi)genome determines whether a functional gene or transcription factor is coded for and can be transcribed; 2) the plant's internal physiological program determines organ-specific metabolite biosynthesis, transport, and storage; 3) the abiotic and biotic environment determine whether specific genes and metabolites are upregulated in response to stressors, and 4) time causes diurnal, ontogenetic, and seasonal variation in the metabolic profiles of plants. This means that chemotypes must be chosen such that they can be consistently identified accounting for additional levels of variation. Exp.: expression.

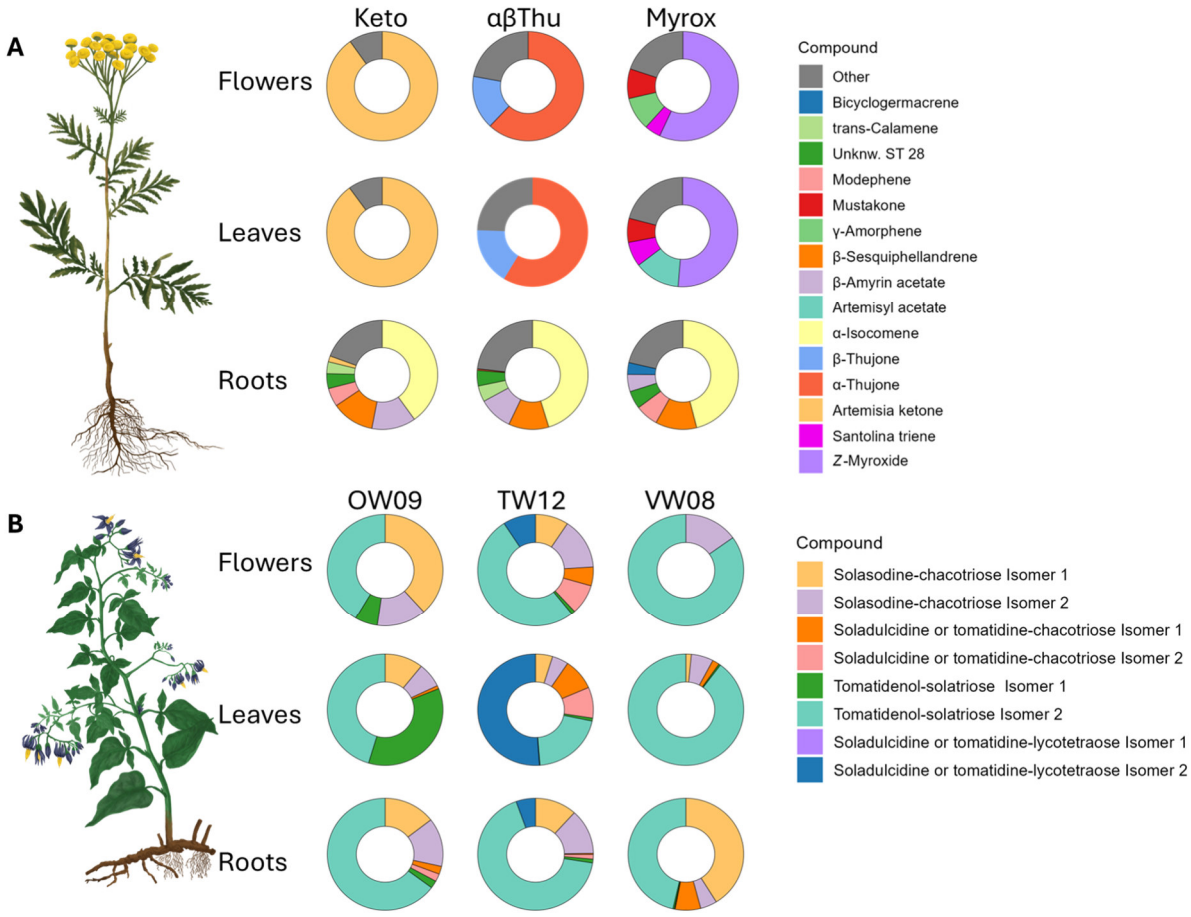

Figure 2: Organ-specific relative composition of **A.** stored terpenoids (extracted with *n*-heptane and analyzed with GC-MS) in three leaf-terpenoid-chemotypes of *Tanacetum vulgare* dominated either by artemisia ketone (Keto), α- and β-thujone (αβThu), or a mixture of *Z*-myroxide, artemisyl acetate, and santolina triene (Myrox) (data redrawn from [48], average of 10-11 replicates per chemotype), and **B.** of steroidal glycosides (extracted with water:methanol 3:1 and analyzed with LC-QToF-MS) in three accessions of *Solanum dulcamara* (samples of clones from accessions mentioned in [45], average of 3 replicates). Unknw. ST – unknown sesquiterpene.

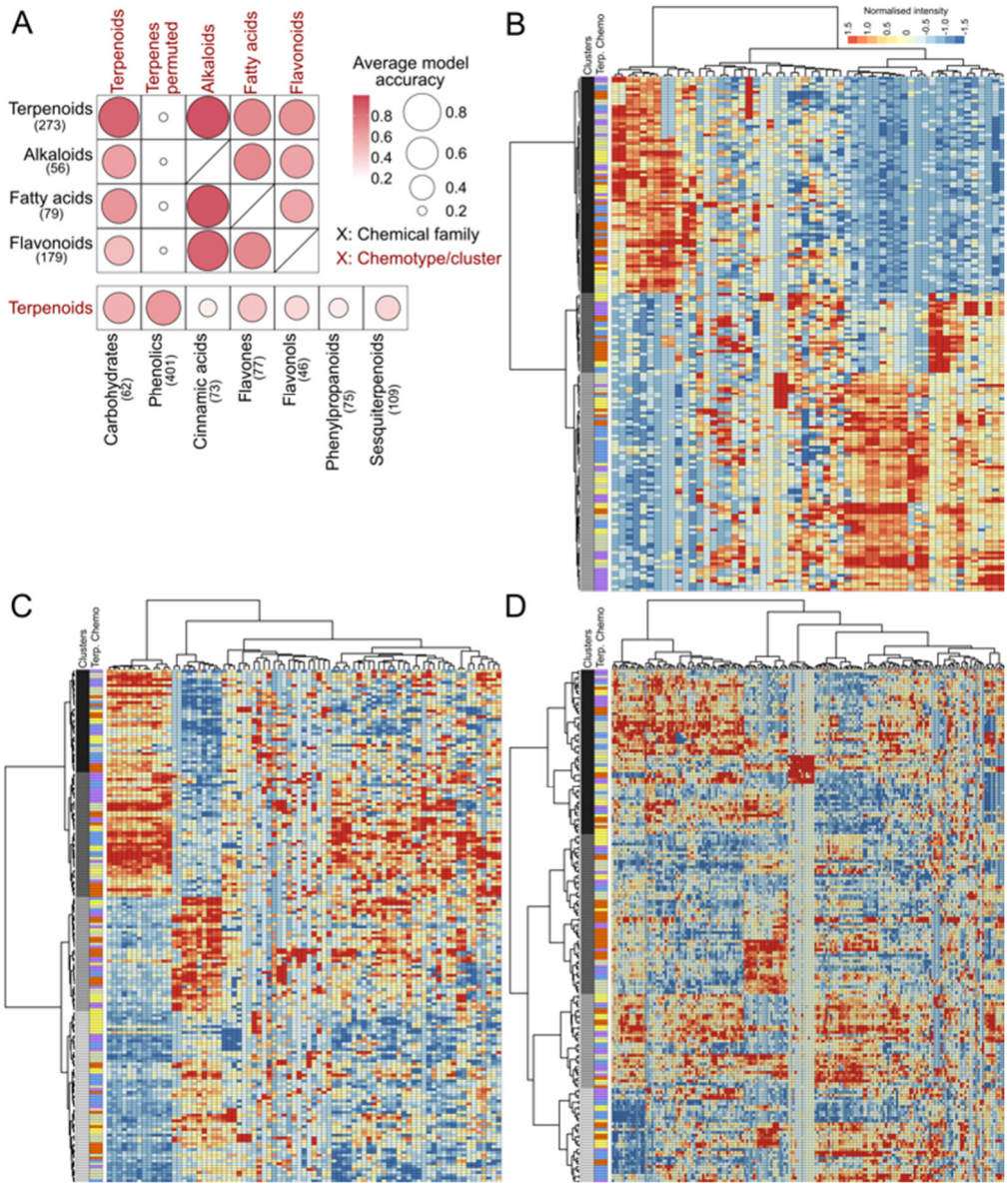

Figure 3: Predictability of different leaf chemotypes using various chemical families in *Tanacetum vulgare*. **A.** Average capacity of different chemical families (in black) to predict distinct chemotypes (in red). For instance, the first column of the matrix represents the average accuracy of terpenoid chemotypes prediction using alkaloids, fatty acids and flavonoids. The bottom line represents additional predictions of terpenoid chemotypes using other chemical families. Average model accuracy was defined on 500 generalized linear models to predict terpenoid chemotype and 100 models for the other chemical families, as previously described [26]. To test the likelihood of spurious predictions, 500 permuted datasets, where chemotypes were randomly swapped among samples, were created. Numbers between parentheses represent the number of chemical features in the corresponding chemical family. Terpenoid chemotypes, defined using GC-MS data, were predicted from LC-MS data from [26]. **B-D**. Clusters defined by different chemical families (B: alkaloids, C: fatty acids, D: flavonoids) and link with terpene chemotypes (second column in

each heatmap, five leaf-terpene chemotypes). Data from 181 plants, leaves collected in the field, extracted in 90% methanol (v:v) and analyzed by UHPLC-QToF-MS/MS (data from [55] reanalyzed). For details see Supplement 1.

SUPPLEMENT 1

To investigate covariations among chemical families and to study the existence of other chemotypes in *Tanacetum vulgare* (see main manuscript, Fig. 3), we used an existing LC-MS dataset comprising leaf analysis from 181 plants belonging to five distinct terpenoid chemotypes [1]. Raw data were re-processed on a newer version of the R-ReX 3D algorithm of Metaboscape (v. 2021b, Bruker Daltonics) with the same parameters (intensity threshold 1000, minimum peak length 11, maxsum method). Raw intensities were normalized by the area of the internal standard hydrocortisone and sample weight, and similar data cleaning was performed (average quality control intensity higher than five times the blank average, features in a minimum of two samples). The pre-processed LC-MS dataset was then normalized using median normalization, cube root transformation, and Pareto scaling using MetaboAnalyst (v. 6) [2], as previously described [1]. Metabolite structure and chemical class predictions were obtained using CSI:FingerID and CANOPUS with the Natural Products Classifier (NPC) ontology [3]. As previously recommended [4], classifications were excluded if the classification approximate score was lower than 0.8. Chemical families including at least 5% of MSMS chemical features (*i.e.*, at least 46 features) were subjected to modelling and clustering analyses. Generalized linear models (GLM) were performed to assess the capacity of different chemical families (*e.g.*, flavonoids) to predict distinct chemotypes (*e.g.*, terpenoid or fatty acids chemotypes). Models were developed using the *glmnet* package [5, 6] as previously described [1, 7]. Briefly, the dataset was divided using stratified sampling into a training set (70%) and a validation set (20%), while real predictions were performed on the testing set (10%). For each modelling condition predicting terpenoid chemotypes (*e.g.*, using fatty acids), 500 models were created, and the average accuracy (real chemotype *versus* predicted chemotype) was calculated. To limit the ecological impact of such models, 100 models were run to measure the predictive accuracy for other clusters (*e.g.*, predicting fatty acid cluster). In addition, 500 permuted datasets, in which chemotypes were randomly swapped between samples, were used to test the likelihood of spurious predictions. To explore whether leaves contain additional clusters based on chemical families other than terpenoids, we used the *factoextra* package on R (v. 4.5.1) [8, 9]. The optimal number of clusters was defined using the 'gap_stat' method as the first cluster preceding a stabilisation of the gap statistic (k). We proposed a visualisation of the tree chemical families (fatty acids, alkaloids and flavonoids) that showed the clearest clustering through heatmaps designed using the *pheatmap* package [10]. Figures were designed with *ggplot2* [11].

References

1. Dussarrat, T. et al. (2023) Influences of chemotype and parental genotype on metabolic fingerprints of tansy plants uncovered by predictive metabolomics. *Sci. Rep.* 13, 11645.

2. Pang, Z.Q. et al. (2024) MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucl. Acids Res.* 52, W398-W406.
3. Kim, H.W. et al. (2021) NPClassifier: A deep neural network-based structural classification tool for natural products. *J. Nat. Prod.* 84, 2795-2807.
4. Hoffmann, M.A. et al. (2022) High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotech*. 40, 411-+.
5. Friedman, J. et al. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stats. Softw.* 33, 1-22.
6. Tay, J.K. et al. (2023) Elastic net regularization paths for all generalized linear models. *J. Stats. Softw.* 106 (1).
7. Dussarrat, T. et al. (2022) Predictive metabolomics of multiple Atacama plant species unveils a core set of generic metabolites for extreme climate resilience. *New Phytol.* 234, 1614-1628.
8. Kassambara, A. and Mundt, F. (2020) factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
9. R Developmental Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vielnna, austria. URL Available at: https://www.R-project.org/.
10. Kolde, R. (2025) pheatmap: Pretty Heatmaps.
11. Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag.