Posted on 10 Jul 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.168900665.52892543/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

The evolution of chemodiversity - From verbal to quantitative models

Frans Thon[1], Caroline Müller[1], and Meike Wittmann[1]

[1]Bielefeld University

July 10, 2023

Abstract

Plants harbour a great chemodiversity, i.e., diversity of specialized metabolites (SMs), at different scales. For instance, individuals can produce a large number of SMs and populations can differ in their metabolite composition. Given the ecological and economic importance of plant chemodiversity, it is important to understand how it arises and is maintained over evolutionary time. For other dimensions of biodiversity, i.e., species diversity and genetic diversity, quantitative models play an important role in addressing such questions. Here we provide a synthesis of existing hypotheses and quantitative models, i.e. mathematical models and computer simulations, for the evolution of plant chemodiversity. We describe each model's ingredients, i.e., the biological processes that shape chemodiversity, the scales it considers, and whether it has been formalized as a quantitative model. Although we identify several quantitative models, not all are dynamic and many influential models have remained verbal. To fill these gaps, we identify quantitative models used for genetic variation that may be adapted for chemodiversity. We end by outlining our vision for the future of chemodiversity modeling, presenting a flexible framework for the creation of individual-based models that address different scales of chemodiversity and combine different ingredients that bring this chemodiversity about.

The evolution of chemodiversity – From verbal to quantitative models

July 7, 2023

Frans M. Thon
Bielefeld University, Faculty of Biology, Theoretical Biology
frans.thon@uni-bielefeld.de
Caroline Müller
Bielefeld University, Faculty of Biology, Chemical Ecology
caroline.mueller@uni-bielefeld.de
Meike J. Wittmann
Bielefeld University, Faculty of Biology, Theoretical Biology
meike.wittmann@uni-bielefeld.de

Statement of Authorship:
Frans Thon and Meike Wittmann wrote the first draft of the manuscript, with input from Caroline Müller. All authors contributed substantially to rewrites and revisions.
Data Accessibility: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.
Running title: Chemodiversity: Verbal to quantitative models
Keywords: chemodiversity, phytochemical diversity, secondary metabolites, mathematical model, simulation
Article type: Synthesis
Abstract: 197 Main text: 6782 Box 1: 749 Box 2: 111 Box 3: 381 Box 4: 404
Nr. References: 94 Nr. Figures: 4 Nr. Text boxes: 4
Correspondence:

Frans Thon

UHG W4-107 Universitätsstraße 25, 33615 Bielefeld, Germany

Tel. +49 521 106-5564

e-mail: frans.thon@uni-bielefeld.de

Abstract

Plants harbour a great *chemodiversity*, i.e., diversity of specialized metabolites (SMs), at different scales. For instance, individuals can produce a large number of SMs and populations can differ in their metabolite composition. Given the ecological and economic importance of plant chemodiversity, it is important to understand how it arises and is maintained over evolutionary time. For other dimensions of biodiversity, i.e., species diversity and genetic diversity, quantitative models play an important role in addressing such questions. Here we provide a synthesis of existing hypotheses and quantitative models, i.e. mathematical models and computer simulations, for the evolution of plant chemodiversity. We describe each model's ingredients, i.e., the biological processes that shape chemodiversity, the scales it considers, and whether it has been formalized as a quantitative model. Although we identify several quantitative models, not all are dynamic and many influential models have remained verbal. To fill these gaps, we identify quantitative models used for genetic variation that may be adapted for chemodiversity. We end by outlining our vision for the future of chemodiversity modeling, presenting a flexible framework for the creation of individual-based models that address different scales of chemodiversity and combine different ingredients that bring this chemodiversity about.

1 Introduction

Chemodiversity is the diversity of specialized metabolites (SMs) at different levels of organisation – from single tissues to entire communities (see Wetzel & Whitehead, 2020, for a review). SMs are compounds produced by an organism that are not directly involved in basic physiological mechanisms such as photosynthesis and reproduction, but may be important for the interaction with herbivores, pollinators, and conspecifics, or protect against abiotic stresses (Wetzel & Whitehead, 2020). For many SMs, the function is currently unknown. SMs are also often called 'secondary metabolites' (Dyer *et al.*, 2018; Moore *et al.*, 2014; Rokas *et al.*, 2020; Stone & Williams, 1992), 'phytochemicals' (in plants, Defossez *et al.*, 2021; Richards *et al.*, 2016), or 'natural products' (Firn & Jones, 2003). SMs belong to metabolite classes including alkaloids, terpenoids, or flavonoids. If individuals within a species differ in the metabolite composition within such metabolite classes, they are often divided into distinct chemotypes. For example, in *Tanacetum vulgare* (Lokki *et al.*, 1973) or *Thymus vulgaris* (Thompson *et al.*, 2003), different terpenoid chemotypes differing in terpenoid composition can be found.

Chemodiversity can be quantified in various ways (Petrén *et al.*, 2023; Wetzel & Whitehead, 2020): it can focus on a group of metabolites belonging to one or several metabolite classes or large parts of the metabolome, and it can be measured within as well as between units of scale. For example, both the number of SMs per individual and the differences in SMs between individuals in a population are aspects of chemodiversity.

The existence of widespread chemodiversity is puzzling from an evolutionary perspective. SMs are often synthesized in complex metabolic pathways that involve multiple enzymes modifying a precursor metabolite into the SM over several steps (Firn & Jones, 2003, 2009; Jones & Firn, 1991). Given the inherent costs of these pathways, one would expect evolution towards a small number of the most beneficial metabolites (Jones & Firn, 1991). Despite this, a high diversity of SMs has been found within and between plant populations and to some extent also in fungal and bacterial populations (Calf *et al.*, 2018; Defossez *et al.*, 2021; Li *et al.*, 2020; Rhoades *et al.*, 1976; Rokas *et al.*, 2020). Evidently, there are mechanisms for the maintenance of this chemodiversity. These mechanisms are what models around chemodiversity attempt to elucidate.

For other dimensions of biodiversity such as species diversity and genetic diversity, quantitative models – both mathematical and simulation models – have long been an important part of scientific inquiry (see e.g. Hubbell, 2001; Kimura, 1983; Wright, 1937). They are for instance used as proof-of-concept models to test the validity of verbal models (Servedio *et al.*, 2014), to generate predictions and hypotheses that can be tested empirically (Servedio *et al.*, 2014), and to estimate parameters from data. Running *in silico* experiments

with quantitative models can give ideas about those aspects of a system that are the most relevant for an outcome or quantity of interest, and thus about which measurements should be taken in an experiment. In that way they can make empirical studies more efficient, and sometimes provide a statistical model that can be fit to data. Additionally, a good quantitative model can unify several studies that differ in methodology so they form a coherent narrative (Otto & Rosales, 2019). Because of these various contributions of quantitative models to scientific inquiry, we argue that chemodiversity should likewise be investigated in this way.

A better understanding of the evolution of chemodiversity that might be conferred by such models also has an applied relevance. For instance, models that predict local adaptation and geographic structure of chemodiversity (Calf *et al.*, 2018; Defossez *et al.*, 2021), could be taken into account in conservation management, as implemented already for genetic diversity (see e.g. Frankham *et al.*, 2002, Ch. 16). Moreover, models for the evolution of chemodiversity might help predict why some introductions of plants or herbivores to new places succeed while others do not, which will be discussed further in Box 3.

In this article, we synthesize the work that has been done so far on developing verbal and, in particular, quantitative models for chemodiversity. For each model, we discuss which biological processes it focuses on and at which scale it addresses chemodiversity. We are not the first to review possible explanations for chemodiversity. There are for instance the excellent reviews by Dyer *et al.* (2018), Kessler & Kalske (2018), Moore *et al.* (2014), Stamp (2003), and Wetzel & Whitehead (2020); therefore we do not want to dwell on reviewing the empirical support for the different explanations for chemodiversity. Instead, we focus on how the verbal models described in those reviews have been tested through quantitative models, and discuss how quantitative models for other types of diversity could be adapted to modeling chemodiversity. We also discuss some of the empirical work that has been done on them to elucidate how quantitative models can connect verbal models and empirical studies. We end by outlining an individual-based modeling framework for chemodiversity and important avenues for future research.

2 Chemodiversity models

There is currently no single unifying theory that can explain chemodiversity at all scales and within as well as between units at each scale. However, there are numerous hypotheses and verbal models (Fig. 1), each of which focuses on a different set of "ingredients", i.e., biological players and processes that can contribute to shaping chemodiversity (Fig. 3, Box 1). Moreover, the different hypotheses and models cover different scales of organization (Fig. 2, Box 2), and can be modeled in different ways (Fig. 4, Box 4). Although there

are connections and overlaps between many hypotheses and they are certainly not all mutually exclusive, to provide some structure for this review, we have placed all models into a hierarchy and grouped similar models together (Fig. 1). Note that sometimes different terms are used in different publications to describe the same or similar models (for instance in the reviews of Dyer *et al.*, 2018; Stamp, 2003; Wetzel & Whitehead, 2020).

Although these various hypotheses on the development and maintenance of chemodiversity are commonly used to generate hypotheses for empirical research (Li *et al.*, 2020; Whitehead *et al.*, 2021), there are fewer quantitative models of these hypotheses. Here, we will review extant quantitative models of chemodiversity, as well as research that could serve as a starting point for the creation of models to fill the numerous gaps in the state of the art.

3 Model descriptions

3.1 Models based on resource constraints

The SMs an organism produces are limited by the resources the organism has access to. Furthermore, the organism may optimize SM production subject to constraints based on available resources or other abiotic or biotic factors (Fig. 3). These ideas are the basis for a group of models that aim to explain variation in quantitative defense level depending on abiotic and biotic conditions. For instance, according to the **carbon nutrient balance hypothesis**, the relative availability of carbon and nitrogen determines which types of compounds are preferentially produced (Bryant *et al.*, 1983). For instance, in low-carbon (i.e. shady) habitats, likely more nitrogen-based toxins are produced. While the hypothesis at times correctly predicts the chemicals which are found in empirical studies, it has come under criticism for frequently failing to do so (Hamilton *et al.*, 2008).

Many of the hypotheses in this group are based on the idea that a plant's investment in defense competes with investment into growth, reproduction, and storage. According to the **resource availability hypothesis** (Coley *et al.*, 1985; Hahn & Maron, 2016), plants in environments with low resource availability are selected to grow slowly and be well defended, whereas plants grown at high resource availability are selected to grow fast and invest less in defense. The **growth-differentiation hypothesis** (Herms & Mattson, 1992) and the **coordinated resource allocation hypothesis** (Monson *et al.*, 2021) expand on the resource availability hypothesis by modeling the maximum possible growth rate in more detail and by adding reserve safety margins, i.e., storage pools that allow plants to plastically respond to biotic and abiotic challenges

(Fig. 3). Coley *et al.* (1985) also propose an equation (which is then extended by Herms & Mattson, 1992; Monson *et al.*, 2021) modeling how plant growth rate depends on investment in defense. In these models (at least with the selected parameter values), plant growth rate is optimized at intermediate defense levels and the optimal strategy depends on the maximum possible growth rate in the absence of herbivores, i.e. resource availability. Thus, if the microenvironments of individuals in a population differ in resource availability, these hypotheses thus offer a quantitative prediction for differences in defense levels between individuals within a population (Fig. 2). Analogously, differences between different populations of the same species are predicted if the habitats of the populations differ in resource availability. Similarly, in a verbal model, Vannette & Hunter (2011) assume that the expression of plant defenses has a nonlinear relationship with the density of arbuscular mycorrhizal fungi (AMF), such that variation in plant defense between populations in different habitats could be explained by variation in AMF density.

Differences in defense levels between individuals or populations in response to environmental conditions are often based on phenotypic plasticity. Many SMs are not produced continuously, but induced only when needed, being then beneficial. For example, exposure to higher ultraviolet-B radiation induces the production of flavonoids, which are important SMs in photoprotection (Agati & Tattini, 2010). Using a simple mathematical framework called **error-management theory**, Orrock *et al.* (2015) computed the optimal herbivore cue strength at which a plant should express its induced defense. This threshold represents a balance between the costs of a false alarm and the costs of being attacked unprepared. Since these costs but also the exposure to herbivores might differ between different plant individuals and populations (Fig. 3), this model could explain chemodiversity in expressed defense between individuals and between populations (Fig. 2).

Clearly, plasticity cannot explain all differences in defense expression and many differences are also genetically determined (see e.g. Garrido *et al.*, 2012). The SMs of populations and species as a whole can be adapted to the typical resources in their location, consequently affecting their chemodiversity (Defossez *et al.*, 2021; Hahn & Maron, 2016; Stamp, 2003). However, at the intraspecific level patterns appear to be more complex, in part because locations with high resource availability also tend to be locations with high herbivore pressure, such that it is still unclear how much of intraspecific defense variation can be explained by the resource-availability hypothesis and other models based on constraints (Hahn & Maron, 2016).

3.2 Apparency

Feeny (1976) developed the **plant apparency hypothesis** as a verbal model where plants in a community

produce different SMs depending on how easy it is for herbivores to find them relative to other plants in the community (Fig. 3): highly apparent plants (e.g. oak trees) produce certain SMs in high concentrations, so called quantitative defenses, that affect most herbivores by acting as digestibility-reducers; less apparent species (e.g. small annual plants) produce toxic SMs in low concentrations, so called qualitative defenses, which some herbivores are immune to. The apparency hypothesis addresses chemodiversity at the community level as well as between clades (see Fig. 2).

Partial support that the verbal apparency model can work in principle comes from a quantitative model by Yamamura & Tsuji (1995). They used an optimal control theory approach to find the optimal investment in growth vs. defense over the life time of the plant. With apparency understood as herbivory pressure, their model results supported the apparency hypothesis. They found that more apparent plants produced quantitative defenses whereas less apparent ones did not, and that for those plants who produced quantitative defenses, the investment in defense increased with apparency. With apparency understood as length of the growth period, only the first prediction was true and only under some parameter settings.

The plant apparency hypothesis has been criticized for being difficult to test in practice. SMs are not easily divided into quantitative and qualitative defenses, but often have properties of both. This makes it difficult to correlate apparency to the presence of quantitative and qualitative defenses (Stamp, 2003). However, which SMs are beneficial to particular plants in a population still might differ depending on their apparency.

3.3 Mechanistic explanations of chemodiversity

Hypotheses in this group focus on why having multiple defense SMs is beneficial to plants, but they do not explicitly consider evolution on the herbivore side. Those hypotheses will be covered in the next section.

The **synergy hypothesis** proposes that the effect of a mixture of SMs is more than the sum of the effects of each SM in isolation. These nonlinear effects of mixtures of SMs then explain why organisms produce multiple SMs. On the molecular level, synergistic effects are explained by molecular interactions between SMs, for example by one SM facilitating the movement of another SM across cell membranes (Richards *et al.*, 2016). The synergy hypothesis addresses chemodiversity within tissues and within individuals (see Fig. 2, and Wetzel & Whitehead, 2020). In empirical studies, it was found that mixtures of SMs affect generalist herbivores more than specialists (Richards *et al.*, 2016). Synergistic effects are difficult to study in the lab in part because it is difficult to isolate large enough amounts of the SMs involved (Dyer *et al.*, 2018). In this and similar situations, quantitative modeling can be a useful tool to formulate hypotheses and guide

experiments. So far, to our knowledge, such models do not exist.

The **moving target hypothesis** is a term proposed by Adler & Karban (1994) and used by Li *et al.* (2020) to refer to an inducible defense where a plant in a population randomly changes its phenotype in response to an attack by herbivores (Fig. 3). This is closely related to the 'novel weapons hypothesis' that is proposed in biological invasions (Box 3). Herbivores then cannot adjust effectively to this defense as it is unpredictable. Wetzel & Whitehead (2020) use the term more broadly to cover hypotheses that describe changes in SMs within individuals over time, functioning like a moving target that herbivores have a hard time adjusting to, even if it does not follow the precise mechanisms described in Adler & Karban (1994). Relatedly, variation between plant individuals in a population can suppress herbivores because they need to adjust their detoxification mechanism when moving between plants with different defense metabolites, which costs energy and is not instantaneous (Pearse *et al.*, 2018). However, the conditions under which this phenomenon would provide an advantage to individuals with rare metabolites and promote the evolution of chemodiversity between individuals are not well understood yet and would probably depend on the details of herbivore behavior.

Adler & Karban (1994) analysed a quantitative, stochastic, plant-herbivore model. In this model, plants could have three defense strategies: a strategy that does not adjust to herbivore damage (constitutive defense), one where plants adjust to herbivore damage by randomly changing their defense (moving target), or one where all plants switch to whichever defense results in the highest growth rate in the current situation (optimal inducible defense). They found that in an environment with a fluctuating number of herbivores, the optimal inducible defense is superior when there is just one herbivore species. When there are at least two herbivores with different resistances to defense strategies, the moving target strategy is superior as long as the cost of defense is not too high. When the cost of defense is high, low levels of constitutive defense are superior.

Eagle-eyed readers will have noticed that this model only included chemodiversity on the within population and between-individual level, not on the within-individual level, although this is part of the verbal model (Fig. 2). It is arguable that different defense strategies can be analogous to chemotypes with different defensive properties. To make this an explicit chemodiversity model capable of investigating chemodiversity on the individual level, a model would need to be added that simulates the actual chemotypes.

The **interaction diversity hypothesis** (Kessler & Kalske, 2018; Whitehead *et al.*, 2021) suggests that plants possess a variety of SMs because they interact with a large number of other organisms (both antagonists and mutualists, as well as interspecific or intraspecific competitors, see Fig. 3) and different metabolites are

active in interaction with the different organisms. Independent selection on each metabolite, potentially in coevolution with the respective interacting species, then leads to the emergence of chemodiversity. Although the name of the hypothesis is rather new, the idea is relatively old and has also been called the "common-sense scenario" (Berenbaum & Zangerl, 1996). Although the interaction diversity hypothesis mostly seeks to explain the diversity of SMs within an individual (Fig. 2, with empirical support by Whitehead *et al.*, 2021), the reasoning can be extended to explain chemodiversity at other scales. For example, if two populations or species differ in the set of interacting animals, we would expect differences in their metabolite composition. Although the ideas of multiple interaction partners driving chemodiversity is very prominent and well-supported, there is currently no quantitative model based on this idea.

3.4 Evolutionary games

Models and hypotheses explaining plant chemodiversity or variation in defenses via evolutionary games focus on the competition of plants "playing" different defense strategies. That is, they have interactions with other plants as important ingredient (Fig. 3). One way in which variation in defense traits can be maintained between individuals in a population and within populations (see Fig. 2) is **negative frequency-dependence** where the respective rare phenotype or genotype has an advantage. Sato *et al.* (2017) and Augner *et al.* (1991) created quantitative models of evolutionary change in plant defense in populations over time. Sato *et al.* (2017) found a pattern of negative frequency-dependent selection that could lead to stable coexistence of undefended plants and defended plants which paid a cost for their defense. The underlying mechanism is that when defended plants become too common, optimally foraging herbivores start to forage also on defended plants despite lower profitability because it is too costly to search for the few undefended plants. This matched up well with their data on a field site of *Arabidopsis halleri* that comprised hairy and glabrous morphs. Similarly, Augner *et al.* (1991), showed in a game-theoretical model of a grazed population with two morphs, that both morphs could coexist if the fitness benefit a defended player has if an undefended opponent is grazed is higher than the profit a non-defended player makes if another non-defended opponent is grazed. These models show that diversity, including chemodiversity, can exist when there is a benefit to being the rare morph in a mixed population.

A coexistence mechanism closely related to negative frequency-dependence is a rock-paper-scissors game. This appears to explain the coexistence of *Brassica nigra* genotypes with high levels of the defense metabolite sinigrin, low-sinigrin *B. nigra* genotypes, and other competing species (Lankau & Strauss, 2007). Here the high-sinigrin type has an advantage in the competition with other plant species and thus has an advantage

when *B. nigra* is rare. As *B. nigra* is then becoming more common and intraspecific competition becomes stronger, it no longer pays to produce so much sinigrin which is not effective against conspecifics, and the low-sinigrin type has an advantage. These dynamics were captured in an individual-based simulation model by Lankau (2009).

Associational effects are the effects that plants have on each other without direct interaction with each other but through interaction with other species, in particular herbivores (Hambäck *et al.*, 2014) (Fig. 3). These exist under many names. One kind of associational effect is social heterosis, which can explain chemodiversity at the within-population scale. Social heterosis refers to a scenario wherein individuals have different traits that are beneficial both to themselves and others in their neighbourhood but cannot all exist in one organism at the same time. In this case, individuals in a diverse group have higher fitness than those in a monoculture, as demonstrated in a quantitative model by Nonacs & Kapheim (2007). In chemodiversity, this could take the form of a population of plants where different plants have different chemical profiles which repel different insect herbivores, both from the individual plant itself and its neighbours. If the production of SMs is costly, it may be detrimental to an individual to produce too many SMs, but beneficial to be in an environment where other individuals produce SMs different from oneself. In this way, social heterosis generates negative frequency-dependent selection on a single or community scale (Fig. 2) as rare types are more likely to find themselves in mixed neighbourhoods with other types. This type of associational effect is also called associational resistance (increased resistance through associational effects).

Another example of associational effects of chemodiversity is the model by Hambäck *et al.* (2014), which modeled associational effects in a community of plants with different traits involved in both visual and olfactory detection by herbivores. Volatile SMs form odor plumes that can be detected by herbivores. The model included both repellent and attractant traits, different rates of detection of these plumes from outside the patch by herbivores, the relative attraction to different types of plants, and the within- and without-patch migration rates. Their model found scenarios where mixed-trait plots had associational resistance to herbivory, scenarios where mixed-trait plots had associational susceptibility, and scenarios where one trait displayed associational susceptibility and the other associational resistance, depending on the nature of the trait of the plant and whether the compared patches comprised the same number of plants. Hambäck *et al.* (2014) did not model evolutionary changes, nor were SMs modeled beyond 'plant type'- which could be taken to correspond to chemotypes with different effects. However, their model could be modified in the future to include these aspects, and in a multi-chemotype system these kinds of associational effects should certainly be taken into account. Similarly, (Till-Bottraud & Gouyon, 1992) used a quantitative model to study the

evolution of the proportion of leaves containing cyanogenic compounds per plant (chemodiversity within individuals) assuming that herbivores leave a patch after tasting the first leaf with cyanogenic compounds. In this way, there were associational effects between leaves both between different plants in the population and between leaves on the same plant.

3.5 Coevolution

Coevolution is a phenomenon where the evolution of two or more interacting species is influenced by the evolution of the respective other species in the interaction (Ehrlich & Raven, 1964). In the case of the evolution of chemodiversity, chemodiversity can be conceptualized in different ways in the co-evolutionary framework. In some studies, the richness or evenness of SMs is used as a trait of individuals (e.g. Calf *et al.*, 2018), in others, the individual SMs are treated as separate traits (Speed *et al.*, 2015). Most coevolutionary models and hypotheses for chemodiversity focus on coevolution with herbivores and envision plants and herbivores to be in an "arms race". There is evidence from field, lab and phylogenetic studies for the coevolutionary arms race hypothesis of chemodiversity on a population, species and lineage level (Becerra *et al.*, 2009; Jander, 2014; Richards *et al.*, 2016; Thorsteinson, 1953).

Many quantitative models of **coevolution with herbivores** can be found in the literature (e.g. Ashby & Boots, 2017; Sandoval-Castellanos & Núñez-Farfán, 2023; Gilman *et al.*, 2012; Speed *et al.*, 2015). However, the one that most explicitly addresses chemodiversity is by Speed *et al.* (2015). This model deals with chemodiversity within individual plants and within populations (see Fig. 2). In this individual-based model, the traits that make plants produce toxins have direct resistance counterparts in insect herbivores, so that an insect that feeds on a plant with trait A has to have the corresponding resistance A to the same degree as the plant to be able to reproduce. We have dubbed this style of model a **toxin-resistance matching model**. When plants and insects were allowed to evolve toxins and resistances through random mutation and selection (Fig. 3), an arms race developed. Selection favoured plants which produced whatever toxin the insects were least resistant to at a time, while other toxins became less prevalent until the insect resistances evolved in turn and a different toxin would be the toxin they were least resistant to. This caused cycles in which multiple toxins were present in the plants at any time in different concentrations, giving one possible explanation of chemodiversity within individuals, within populations, and in populations through time (Fig. 2). Which chemicals dominate in the population at any given time is independent of the chemicals which dominate in separate populations which are going through their own coevolutionary processes, thus causing chemodiversity between populations and within the species.

Other quantitative coevolutionary models do not directly address chemodiversity, but contain elements which could be used in chemodiversity models. Gilman *et al.* (2012) developed a model similar to the one by Speed *et al.* (2015) and also show that coevolving with multiple traits can give the 'victim' species an edge, although their model is not explicitly about chemodiversity. Ashby & Boots (2017) developed a host-parasite model which uses a toxin-resistance matching model to model host defenses and parasite resistance traits. This model displayed dynamics of both stable equilibrium where either all traits or none were present, and infinite cyclic evolution. In the case of infinite cycles, there were two types of cycles occurring in the same simulation: fast-cycling of the possible resistances at the same number of resistance loci in a plant and slow cycling of number of resistances that existed in a plant. This was possible through a dynamic where immediate benefits or detriments of the exact resistances an individual possessed caused rapid changes in the resistances present, while subtle fitness costs of maintaining resistances created a slow dwindling of the number of resistances until a threshold was reached where having many was beneficial again. Both types of cycles displayed negative frequency-dependent-selection and are thus linked to the game theory models discussed above. These dual cycles are a feature not found in the model of Speed *et al.* (2015) or in the model by Gilman *et al.* (2012) which is very similar to the model by Ashby & Boots (2017) but only found stable equilibria. While the models of Ashby & Boots (2017) and Gilman *et al.* (2012) are not specific to chemodiversity, these and other matched-trait models of coevolution could easily be translated to a chemodiversity context by making the resistance traits correspond to metabolites.

Some authors have proposed that chemodiversity can come about when individual plants produce multiple toxins to make it more difficult for herbivores to adapt to counter every single toxin of the set a plant produces (Speed *et al.*, 2012; Wetzel & Whitehead, 2020). This **slowed adaptation hypothesis**, as coined by Wetzel & Whitehead (2020) is the outcome of an evolutionary process for which toxin-resistance matching is the mathematical description. The slowed adaptation hypothesis may play a role in biological invasions (Box 3). However, this is not the only possible verbal explanation for a mechanism of toxin-resistance matching. In the model by Speed *et al.* (2015), the observed chemodiversity is the result of constant, moving target-like innovation and the remains of innovations past, rather than of slowed adaptation of herbivores to a standing set of toxins. Additionally, while all models described in this review that have an explicit or implicit description of individual metabolites use a toxin-resistance model, this is not the only possible way to model interactions between herbivores with metabolites and combinations of metabolites. Quantitative models which explore synergy between metabolites, for example, would require the effects of metabolites to differ depending on which other metabolites are present.

Another coevolutionary hypothesis which is based on spatial structure is the **geographic mosaic of coevolution hypothesis**. Differences in SM composition between populations may derive from opposite selection pressures of specialists, which use certain SMs for host plant finding, versus generalists that are repelled or deterred by the same SMs (Enge *et al.*, 2012; van der Meijden, 1996). Geographic differences in herbivore abundance may thus result in a selection pressure mosaic (Thompson, 1999; Zangerl & Berenbaum, 2003). In a general mathematical model for the geographic mosaic of coevolution that could also be applied to toxin-resistance matching, Gomulkiewicz *et al.* (2000) showed that a spatial setup with coevolutionary hot spots (mutual selection) and cold spots (only one species exerts selection on the other, but not vice versa) could under some conditions allow the maintenance of polymorphism in both species and differences in allele frequency between populations. For a plant toxin coevolving with a herbivore, this would mean that the model could explain coexistence of toxic and nontoxic plant individuals in the same patch as well as differences in the frequency of toxic plants between patches (see Fig. 2).

A recent coevolutionary model that does not fit clearly into any of the previously discussed categories investigates the circumstances under which non-linearity of costs and benefits of herbivore resistance can lead to a mix of herbivory-resistant and herbivory-tolerating plants in a population (Sandoval-Castellanos & Núñez-Farfán, 2023). The model is an individual-based model that compares additive and multiplicative versions of a fitness function that includes the fitness costs and benefits of resistance and tolerance as well as the fitness cost of inbreeding depression. It concludes that the nonlinear fitness function needs to be concave, the allocation of costs and benefits multiplicative, selfing non-heritable, and tolerance costly to promote a mix of strategies. The 'resistance' in this model does not necessarily refer to SMs, but the model could be easily modified to include different SMs.

To really make the models specific to chemodiversity, introducing ingredients like the branching pathways through which chemicals are created would be useful. Additionally, while toxin-resistance matching has been used in most quantitative coevolutionary models, empirical studies frequently test linear additive or nonlinear synergistic effects of SMs (Dyer *et al.*, 2018; Richards *et al.*, 2016). Therefore, developing models that use these additive or synergistic effects could bring the empirical and quantitative modeling research in this field together.

Coevolution with seed dispersers is another way in which coevolution can potentially promote chemodiversity. For example, in the genus *Piper*, fruit toxicity affects dispersal by coevolved frugivores. This was statistically modeled by Baldwin *et al.* (2020) based on empirical data collected in a system where *Carollia* fruit bats consume *Piper* fruits. They found that higher levels of defensive amides correlate with

shorter gut retention times and lower dispersal distances, while lower concentrations of amides correlate with higher dispersal distances as well as optimal ripeness of the fruit and maximum attractiveness to the fruit bats. In this way, the ripe *Piper* seeds are dispersed as far as possible, while unripe fruits are less likely to be eaten in the first place. Statistical models like this can be used to make quantitative predictions for future empirical studies. In its current form, such a model could explain differences in amide concentration between fruits within an individual or between individuals when fruits differ in ripeness (Fig. 2). For purposes of studying chemodiversity, it would be interesting to not only include total amide concentrations as a predicting variable, but also diversity of amides. It may be possible to shine a light on how coevolution of seed dispersal influences chemodiversity on the within-plant and between-population scale. Moreover, gut retention time in frugivores is not the only way in which chemodiversity may play a role in the coevolution of fruit-bearing plants and their seed dispersers. For example, SMs also may act as olfactory cues to frugivores, inhibit seed germination, and act as toxins (Cipollini & Levey, 1997). So far, we did not find any quantitative models addressing the potentially important consequences of coevolution with seed dispersers for the evolution of plant chemodiversity.

3.6 Screening hypothesis

The models reviewed in the previous sections generally assume that every metabolite has a specific biological activity in the interaction with one or more species. Metabolites without a specific advantageous activity would be expected to be selected away because of their inherent costs. The screening hypothesis first proposed by Jones & Firn (1991) (see also Firn & Jones, 2003; Firn, 2009) challenges this view. Based on the observation that many metabolites produced by plants and fungi have no known function, they argue that pronounced specific biological activity is in fact a rare phenomenon. Thus, in order to defend themselves against a variety of herbivores, plants need to "screen" a large number of "candidate metabolites". They keep a diversity of metabolites, even many without a function, because this diversity and the underlying multitude of enzymes allows them via mutation to rapidly generate new metabolites and thus increase their chances to find at least some that have strong activity. Since maintaining a diversity of enzymes and metabolites still has costs, the screening hypothesis predicts that the metabolic pathways would be selected such that they can produce a high number of metabolites with as little enzymatic machinery as possible, mostly via promiscuous enzymes that can take multiple substrates, leading to grid-like metabolic pathways. Keeping a diversity of metabolites around can then also allow plants to rapidly respond to new herbivores or herbivores that have evolved a counter-defense (Jones & Firn, 1991).

In summary, the screening hypothesis is based on the ingredients of random mutation and natural selection, metabolic pathways, and interactions with other organisms (Fig. 3). Though the focus in the original formulation is on herbivores, by the same logic, a diversity of metabolites can help plants to develop metabolites that are beneficial in the interaction with other plants or with mutualists and soil communities. The screening hypothesis has been formulated as a verbal model and to our knowledge, it has not been formulated as a quantitative model. Nor are the ideas of grid-like metabolic pathways, promiscuous enzymes, and rare metabolic activity incorporated into other quantitative models.

The screening hypothesis mostly addresses chemodiversity within units of scale, i.e. it attempts to explain why a plant individual, or a plant population, species or clade produces so many different metabolites (Fig. 2). Although this has been less discussed by the authors of the screening hypothesis, if a complex metabolic network allows plants to rapidly generate new metabolites, this could also help explain differences between individuals, populations, and species. This appears to be supported by the observation that many enzymes involved in generating chemodiversity of SMs are less conserved than those that are involved in more primary metabolism (Weng *et al.*, 2012).

The screening hypothesis has been criticized because evolutionary foresight appears to be required for evolution to create a complex network of pathways just based on the chance that some of its products might eventually prove useful (Pichersky *et al.*, 2006). Proponents of the hypothesis counter that no evolutionary foresight is required if the rapid adaptation made possible through the network is sufficiently beneficial to be selected for (Firn & Jones, 2006). A formal mathematical description or computer model could demonstrate to what extent these criticisms are warranted. Such models have, to our knowledge, not yet been developed, but could therefore play an important role in the debate around the merits of the screening hypothesis.

3.7 Population genetics models

Chemodiversity is, at least in part, underpinned by genetic variation. Thus, it would seem natural to extend classical population genetic models to study chemodiversity.

In biodiversity research (Hubbell, 2001) and in population genetics (Kimura, 1983), **neutral models** play an important role as null models for more interesting ecological and evolutionary processes though there is also much debate over how useful they are (see e.g. Clark, 2009; Kern & Hahn, 2018). While neutral biodiversity models assume that species within a guild are equivalent in terms of their birth, death, and migration rates, neutral models in population genetics assume that allele frequencies change just through mutation and drift without selection. We argue that similar neutral models should also be developed for chemodiversity. That

is, diversity in such models would come about and be maintained through random mutation and genetic drift without selection (Box 1). A neutral model for chemodiversity should additionally include pathways and enzymes that are encoded by the modeled genes and that thus also evolve neutrally (Fig. 3). A neutral model could in principle be formulated at any scale of organization, though a neutral model derived from neutral models in population genetics would most straightforwardly address chemodiversity within populations, within and between individuals (Fig. 2). This model would serve as a null model for the origin and maintenance of chemodiversity. Any explanations of the origin and maintenance of chemodiversity which propose that there is a selective advantage to producing a diversity of SMs can then be compared to it. We have not found any sources where a model of chemodiversity is compared to such a population genetic null model, but we think that such an approach would be extremely useful. However, patterns of chemical traits between species have been compared to null models for trait evolution on a phylogeny. For example, Volf *et al.* (2018) used Brownian motion and white noise models. So far, however, these models cannot address intraspecific chemodiversity.

An explanation for chemodiversity that is slightly more complex than neutrality is **mutation-selection-drift balance** (see e.g. Wright, 1937). Mutation-selection-drift models in population genetics generally assume that mutations are either unconditionally beneficial or deleterious. There are no complex selection patterns like balancing selection or negative frequency-dependent selection. Thus variation under this model is observed due to mutations that are transiently segregating in the mutation before they are getting either lost or fixed. In a chemodiversity extension of such models, new SMs would come about through random mutation in existing metabolic pathways. Here it is assumed that at least some SMs improve the organisms' fitness and are under positive selection, while others may be detrimental or are not worth the cost of their production and are thus under negative selection. The probability that a mutation is fixed or lost and the expected time to do so then depends on the strength of selection relative to genetic drift. During their time to loss or fixation, mutations and their effects on the chemical phenotype would then transiently contribute to chemodiversity. Like neutral models, mutation-selection-drift models could explain chemodiversity within populations, and within and between individuals (Fig. 2), but potentially also at all other scales. For example if different SMs randomly become fixed in different populations and clades this could explain differences between these units. Models based on mutation-selection-drift may thus be highly applicable to chemodiversity.

4 Vision

It would be easy to fall into the trap of assuming that the different frameworks and hypotheses described here must be in competition, and that through modeling and empirical work, we can find one correct and all others incorrect. Instead, we agree with Dyer *et al.* (2018) that the hypotheses do not have to contradict each other. In fact, many of the hypotheses and selection mechanisms described here may be at play at once within the same plant species. While the members of a particular set of SMs cannot all have both synergistic and toxin-resistance matching fitness effects at the same time, this set of SMs might have synergistic effects on herbivory at the plant level, and also be under negative frequency-dependent selection on the population level. This set can then simultaneously be in a co-evolutionary arms race with herbivores on the lineage level, where new SMs come about within a pathway through the process well-described by the screening hypothesis. Another set of SMs in the same plant species may only exist transiently due to random mutations in a pathway under mutation-selection-drift balance. For this reason, it is useful to think of the different hypotheses and selection methods as different possible factors in the evolution and maintenance of chemodiversity, and not as mutually exclusive. At the same time, some of these hypotheses may well be invalid, or some might contribute less to explaining chemodiversity than others. Quantitative models can serve to investigate the validity and explanatory power of these hypotheses. Combining multiple hypotheses in a single quantitative model would enable us to compare the relative explanatory power of different models (Box 4).

When we make a model to test a verbal hypothesis, we need to incorporate enough detail to describe what that hypothesis is about. That means that a test of the coevolutionary arms race needs to incorporate a focal species and an interacting co-evolving species. For the screening hypothesis it is necessary that a model of the metabolic pathways is included that works as described by Firn & Jones (2003); Jones & Firn (1991), but it is not fundamentally necessary that any other species is involved. When modeling a very specific system, it is useful to ensure that all aspects of that system are carefully considered, and all relevant details are included, and all assumptions are justified (Servedio *et al.*, 2014). For example, if one were to model a biological invasion as described in Box 3, it would make sense to include elements of the slowed adaptation and novel weapons/moving target hypotheses (but see Box 4).

We have shown here that there is a great body of work on the description of current chemodiversity and its evolutionary history, and there are meaningful hypotheses for the evolution of chemodiversity and maintenance of chemodiversity. For many of the hypotheses, we also found a small number of quantitative modeling studies. Quantitative models that explicitly address chemodiversity where individuals produce a

diverse set of metabolites are few and far between. We found such a model for a coevolutionary toxin-resistance matching model where negative frequency-dependence was found Speed *et al.* (2015). The other models did not explicitely consider metabolites but rather a more abstract numerical. These are nevertheless rendered in red in Fig. 1 and Fig. 2, because they can be extended to explicitly include this chemodiversity. So when one looks at Fig. 2, there is work to be done both to extend these models, particularly for those hypotheses which could possibly explain chemodiversity on levels that are not between-individual or within-population. Such models will often be sufficiently complicated that it might not be possible to capture them in a system of equations that can then be solved analytically. Thus, numerical solutions and individual-based models as in Speed *et al.* (2015) appear to be the most productive ways to model the evolution of chemodiversity in these cases. In Box 4 we outline a modular approach to constructing such an individual-based model.

5 Conclusion

In conclusion, while there has been much verbal theory-crafting on chemodiversity and there is a small number of interesting quantitative models for many of the verbal hypotheses, there is much space for the development of mathematical and simulation models. This can be done by expanding existing models in topics related to chemodiversity, adapting existing models of diversity, or by combining different submodels to create coherent models of chemodiversity. In this way, the existing theory can be tested for validity, the explanatory power of different hypotheses can be explored, and the results can be used to interpret empirical data to better understand the dynamics underlying extant chemodiversity.

6 Acknowledgements

We are grateful to Andrea Bräutigam, Redouan Adam Anaia, Hampus Petrén and the Theoretical Biology group at Bielefeld University for comments on earlier versions of the manuscript, and Theresa Rohrberg for creating some of the artwork used in the figures. This project was funded by the German Research Foundation (DFG), projects WI 4544/2 and MU 1829/28-1, as part of the Research Unit (RU) FOR 3000 Ecology and Evolution of Intraspecific Chemodiversity of Plants.

References

Adler, F. & Karban, R. (1994) Defended fortresses or moving targets? Another model of inducible defenses inspired by military metaphors. *The American Naturalist* **144**, 813–832.

Agati, G. & Tattini, M. (2010) Multiple functional roles of flavonoids in photoprotection. *New Phytologist* **186**, 786–93.

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C. & Tawfik, D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nature Genetics* **37**, 73–76.

Akiyama, K., Matsuzaki, K.i. & Hayashi, H. (2005) Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**, 824–827.

Allstadt, A., Caraco, T., Molnár, F. & Korniss, G. (2012) Interference competition and invasion: Spatial structure, novel weapons and resistance zones. *Journal of Theoretical Biology* **306**, 46–60.

Ashby, B. & Boots, M. (2017) Multi-mode fluctuating selection in host-parasite coevolution. *Ecology Letters* **20**, 357–365.

Augner, M., Fagerstrom, T. & Tuomi, J. (1991) Competition, defense and games between plants. *Behavioral Ecology and Sociobiology* **29**, 231–234.

Baldwin, J.W., Dechmann, D.K.N., Thies, W. & Whitehead, S.R. (2020) Defensive fruit metabolites obstruct seed dispersal by altering bat behavior and physiology at multiple temporal scales. *Ecology* **101**, e02937.

Becerra, J., Noge, K. & Venable, D. (2009) Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18062–18066.

Berenbaum, M.R. & Zangerl, A.R. (1996) Phytochemical diversity: adaptation or random variation? *Phytochemical diversity and redundancy in ecological interactions* (eds. J.T. Romeo, J.A. Saundders & P. Barbosa), pp. 1–24, Springer.

Borghi, M., Junker, R., Lucas-Barbosa, D. & Zych, M. (2021) Editorial: Flower metabolism and pollinators. *Frontiers in Plant Science* **12**, 776978.

Bryant, J., Chapin III, F.S. & Klein, D. (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos* **40**, 357–368.

Calf, O.W., Huber, H., Peters, J., Weinhold, A. & van Dam, N. (2018) Glycoalkaloid composition explains variation in slug resistance in *Solanum dulcamara*. *Oecologia* **187**, 495–506.

Callaway, R. & Ridenour, W. (2004) Novel weapons: Invasive success and the evolution of increased competitive ability. *Frontiers in Ecology and the Environment* **2**, 436–443.

Cheng, A.X., Lou, Y.G., Mao, Y.B., Lu, S., Wang, L.J. & Chen, X.Y. (2007) Plant terpenoids: Biosynthesis and ecological functions. *Journal of Integrative Plant Biology* **49**, 179 – 186.

Cipollini, M. & Levey, D. (1997) Secondary metabolites of fleshy vertebrate-dispersed fruits: Adaptive hypotheses and implications for seed dispersal. *The American Naturalist* **150**, 346–72.

Clark, J.S. (2009) Beyond neutral science. *Trends in Ecology & Evolution* **24**, 8–15.

Coley, P., Bryant, J. & Chapin III, F.S. (1985) Resource availability and plant anti-herbivore defense. *Science* **230**, 895–899.

Cooper, J. (2004) Multiple responses of rhizobia to flavonoids during legume root infection. *Advances in Botanical Research* **41**, 1–62.

Cox, G.W. (2004) *Alien species and evolution - The evolutionary ecology of exotic plants, animals, microbes, and interacting native species*. Island Press, Washington, D.C.

De la Peña, C. & Loyola-Vargas, V. (2014) Biotic interactions in the rhizosphere: A diverse cooperative enterprise for plant productivity. *Plant Physiology* **166**, 701–719.

Defossez, E., Pitteloud, C., Descombes, P., Glauser, G., Allard, P.M., Walker, T. & et al. (2021) Spatial and evolutionary predictability of phytochemical diversity. *Proceedings of the National Academy of Sciences* **118**, e2013344118.

Dyer, L., Philbin, C., Ochsenrider, K., Richards, L., Massad, T., Smilanich, A. & et al. (2018) Modern approaches to study plant–insect interactions in chemical ecology. *Nature Reviews Chemistry* **2**, 50–64.

Ehrlich, P.R. & Raven, P.H. (1964) Butterflies and plants, a study in coevolution. *Evolution* **8**, 586–608.

Enge, S., Nylund, G., Harder, T. & Pavia, H. (2012) An exotic chemical weapon explains low herbivore damage in an invasive alga. *Ecology* **93**, 2736–2745.

Feeny, P. (1976) Plant apparency and chemical defense. *Biochemical Interaction Between Plants and Insects*, pp. 1–40, Springer US, Boston, MA.

577 Firn, R. (2009) *Nature's chemicals: the natural products that shaped our world*. Oxford University Press.

578 Firn, R.D. & Jones, C.G. (2003) Natural products – a simple model to explain chemical diversity. *Natural*
579 *Product Reports* **20**, 382–391.

580 Firn, R.D. & Jones, C.G. (2006) Response to Pichersky et al.: Correcting a misconception about the screening
581 hypothesis. *Trends in Plant Science* **11**, 422–422.

582 Firn, R.D. & Jones, C.G. (2009) A Darwinian view of metabolism: molecular properties determine fitness.
583 *Journal of Experimental Botany* **60**, 719–726.

584 Frankham, R., Ballou, J.D. & Briscoe, D. (2002) *Introduction to Conservation Genetics*. Cambridge Univer-
585 sity Press.

586 Garrido, E., Andraca-Gómez, G. & Fornoni, J. (2012) Local adaptation: simultaneously considering herbi-
587 vores and their host plants. *New Phytologist* **193**, 445–453.

588 Gilman, R.T., Nuismer, S.L. & Jhwueng, D.C. (2012) Coevolution in multidimensional trait space favours
589 escape from parasites and pathogens. *Nature* **483**, 328–330.

590 Gomulkiewicz, R., Thompson, J.N., Holt, R.D., Nuismer, S.L. & Hochberg, M.E. (2000) Hot spots, cold
591 spots, and the geographic mosaic theory of coevolution. *The American Naturalist* **156**, 156–174.

592 Hahn, P. & Maron, J. (2016) A framework for predicting intraspecific variation in plant defense. *Trends in*
593 *Ecology & Evolution* **31**, 646–656.

594 Hamberger, B. & Bak, S. (2013) Plant P450s as versatile drivers for evolution of species-specific chemical
595 diversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **368**,
596 20120426.

597 Hambäck, P., Inouye, B., Andersson, P. & Underwood, N. (2014) Effects of plant neighborhoods on plant-
598 herbivore interactions: Resource dilution and associational effects. *Ecology* **95**, 1370–1383.

599 Hamilton, J., Zangerl, A., DeLucia, E. & Berenbaum, M. (2008) The carbon-nutrient balance hypothesis:
600 Its rise and fall. *Ecology Letters* **4**, 86 – 95.

601 Herms, D. & Mattson, W. (1992) The dilemma of plants: To grow or defend. *The Quarterly Review of*
602 *Biology* **67**, 283–335.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton and Oxford.

Jander, G. (2014) *Revisiting Plant-Herbivore Co-Evolution in the Molecular Biology Era*, vol. 47, chap. 11, pp. 361–384. Wiley-Blackwell.

Jones, C.G. & Firn, R.D. (1991) On the evolution of plant secondary chemical diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **333**, 273–280.

Kern, A.D. & Hahn, M.W. (2018) The neutral theory in light of natural selection. *Molecular Biology and Evolution* **35**, 1366–1371.

Kessler, A. & Kalske, A. (2018) Plant secondary metabolite diversity and species interactions. *Annual Review of Ecology Evolution and Systematics* **49**, 115–138.

Kimura, M. (1983) The neutral theory of molecular evolution. *Scientific American* **241**, 98–100.

Lankau, R.A. (2009) Genetic variation promotes long-term coexistence of *Brassica nigra* and its competitors. *The American Naturalist* **174**, E40–E53.

Lankau, R.A. & Strauss, S. (2007) Mutual feedbacks maintain both genetic and species diversity in a plant community. *Science* **317**, 1561–1563.

Lankau, R.A. & Strauss, S. (2008) Community complexity drives patterns of natural selection on a chemical defense of *Brassica nigra*. *The American Naturalist* **171**, 150–61.

Li, D., Halitschke, R., Baldwin, I. & Gaquerel, E. (2020) Information theory tests critical predictions of plant defense theory for specialized metabolism. *Science Advances* **6**, eaaz0381.

Lokki, J., Sorsa, M., Forsén, K. & Schantz, M.V. (1973) Genetics of monoterpenes in chrysanthemum vulgare. *Hereditas* **74**, 225–232.

Moghe, G.D. & Last, R.L. (2015) Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology* **169**, 1512–1523.

Monson, R., Trowbridge, A., Lindroth, R. & Lerdau, M. (2021) Coordinated resource allocation to plant growth-defense trade-offs. *New Phytologist* **233**, 1051–1066.

Moore, B.D., Andrew, R.L., Kulheim, C. & Foley, W.J. (2014) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist* **201**, 733–750.

Mueller, L., Borstein, S., Tague, E., Dearth, S., Castro, H., Campagna, S. & et al. (2020) Populations of *Populus angustifolia* have evolved distinct metabolic profiles that influence their surrounding soil. *Plant and Soil* **448**, 399–411.

Nevo, O., Razafimandimby, D., Jeffrey, J., Schulz, S. & Ayasse, M. (2018) Fruit scent as an evolved signal to primate seed dispersal. *Science Advances* **4**, eaat4871.

Nonacs, P. & Kapheim, K.M. (2007) Social heterosis and the maintenance of genetic diversity. *Journal of Evolutionary Biology* **20**, 2253–2265.

Orrock, J.L., Sih, A., Ferrari, M.C., Karban, R., Preisser, E.L., Sheriff, M.J. & et al (2015) Error management in plant allocation to herbivore defense. *Trends in Ecology & Evolution* **30**, 441–445.

Otto, S. & Rosales, A. (2019) Theory in service of narratives in evolution and ecology. *The American Naturalist* **195**, 290–299.

Pankoke, H., Tewes, L., Matties, S., Hensen, I., Schädler, M., Ebeling, S. & et al. (2019) Pre-adaptations and shifted chemical defences provide *Buddleja davidii* populations with high resistance against antagonists in the invasive range. *Biological Invasions* **21**, 333–347.

Pearse, I.S., Paul, R. & Ode, P.J. (2018) Variation in plant defense suppresses herbivore performance. *Current Biology* **28**, 1981–1986.

Petrén, H., Anaia, R.A., Aragam, K.S., Bräutigam, A., Eckert, S., Heinen, R. & et al. (2023) Understanding the phytochemical diversity of plants: Quantification, variation and ecological function. *bioRxiv* .

Pichersky, E., Sharkey, T. & Gershenzon, J. (2006) Plant volatiles: a lack of function or a lack of knowledge? *Trends in Plant Science* **11**, 421.

Piola, F., Bellvert, F., Meiffren, G., Rouifed, S., Walker, V., Comte, G. & et al. (2013) Invasive *Fallopia* x *Bohemica* interspecific hybrids display different patterns in secondary metabolites. *Ecoscience* **20**, 230–239.

Rhoades, D.F., Cates, R.G. & Mansell, R.L. (1976) *Toward a General Theory of Plant Antiherbivore Chemistry*, pp. 168–213. Springer US, Boston, MA.

Richards, L., Glassmire, A., Ochsenrider, K., Smilanich, A., Dodson, C., Jeffrey, C. & et al. (2016) Phytochemical diversity and synergistic effects on herbivores. *Phytochemistry Reviews* **15**, 1153–1166.

Rokas, A., Mead, M., Steenwyk, J., Raja, H. & Oberlies, N. (2020) Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Natural Product Reports* **37**, 868–878.

Salazar, D. & Marquis, R.J. (2022) Testing the role of local plant chemical diversity on plant–herbivore interactions and plant species coexistence. *Ecology* **103**, e3765.

Sandoval-Castellanos, E. & Núñez-Farfán, J. (2023) The joint evolution of herbivory defense and mating system in plants: A simulation approach. *Plants* **12**, 555.

Sato, Y., Ito, K. & Kudoh, H. (2017) Optimal foraging by herbivores maintains polymorphism in defence in a natural plant population. *Functional Ecology* **31**, 2233–2243.

Schrieber, K. & Lachmuth, S. (2016) The genetic paradox of invasions revisited: The potential role of inbreeding environment interactions in invasion success. *Biological Reviews* **92**, 939–952.

Schweiger, R. & Müller, C. (2015) Leaf metabolome in arbuscular mycorrhizal symbiosis. *Current Opinion in Plant Biology* **26**, 120–126.

Servedio, M.R., Brandvain, Y., Dhole, S., Fitzpatrick, C.L., Goldberg, E.E., Stern, C.A. & et al. (2014) Not just a theory—the utility of mathematical models in evolutionary biology. *PLOS Biology* **12**, 1–5.

Shoji, T. (2019) The recruitment model of metabolic evolution: Jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. *Frontiers in Plant Science* **10**, 560.

Smakowska, E., Kong, J., Busch, W. & Belkhadir, Y. (2016) Organ-specific regulation of growth-defense tradeoffs by plants. *Current Opinion in Plant Biology* **29**, 129–137, growth and development.

Speed, M.P., Fenton, A., Jones, M.G., Ruxton, G.D. & Brockhurst, M. (2015) Coevolution can explain defensive secondary metabolite diversity in plants. *New Phytologist* **208**, 1251–1263.

Speed, M.P., Ruxton, G.D., Mappes, J. & Sherratt, T.N. (2012) Why are defensive toxins so variable? An evolutionary perspective. *Biological Reviews* **87**, 874–884.

Stamp, N. (2003) Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology* **78**, 23–55.

Stone, M.J. & Williams, D.H. (1992) On the evolution of functional secondary metabolites (natural products). *Molecular Microbiology* **6**, 29–34.

Tewes, L.J., Michling, F., Koch, M.A. & Müller, C. (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.

Thompson, J.D., Chalchat, J.C., Michet, A., Linhart, Y.B. & Ehlers, B. (2003) Qualitative and quantitative variation in monoterpene co-occurrence and composition in the essential oil of *Thymus vulgaris* chemotypes. *Journal of Chemical Ecology* **29**, 859–880.

Thompson, J.N. (1999) Specific hypotheses on the geographic mosaic of coevolution. *The American Naturalist* **153**, S1–S14.

Thorsteinson, A. (1953) The chemotactic responses that determine host specificity in an oligophagous insect (*Plutella maculipennis* (Curt.) Lepidoptera). *Canadian Journal of Zoology* **31**, 52–72.

Till-Bottraud, I. & Gouyon, P.H. (1992) Intra-versus interplant batesian mimicry? a model on cyanogenesis and herbivory in clonal plants. *The American Naturalist* **139**, 509–520.

van der Meijden, E. (1996) Plant defence, an evolutionary dilemma: Contrasting effects of (specialist and generalist) herbivores and natural enemies. *Entomologia Experimentalis et Applicata* **80**, 307–310.

Vannette, R.L. & Hunter, M.D. (2011) Plant defence theory re-examined: Nonlinear expectations based on the costs and benefits of resource mutualisms. *Journal of Ecology* **99**, 66–76.

Volf, M., Segar, S.T., Miller, S.E., Isua, B., Sisol, M., Aubona, G. & et al. (2018) Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in ficus. *Ecology letters* **21**, 83–92.

Weng, J.K., Philippe, R.N. & Noel, J.P. (2012) The rise of chemodiversity in plants. *Science* **336**, 1667–1670.

Wetzel, W.C. & Whitehead, S.R. (2020) The many dimensions of phytochemical diversity: linking theory to practice. *Ecology Letters* **23**, 16–32.

Whitehead, S.R., Bass, E., Corrigan, A., Kessler, A. & Poveda, K. (2021) Interaction diversity explains the maintenance of phytochemical diversity. *Ecology Letters* **24**, 1205–1214.

Wolf, V., Berger, U., Gassmann, A. & Müller, C. (2011) High chemical diversity of a plant species is accompanied by increased chemical defence in invasive populations. *Biological Invasions* **13**, 2091–2102.

Wright, S. (1937) The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences* **23**, 307–320.

Yadav, A.N., Kour, D., Kaur, T., Devi, R. & Yadav, A. (2022) Endophytic fungal communities and their biotechnological implications for agro-environmental sustainability. *Folia Microbiologica* **67**, 203–232.

Yamamura, N. & Tsuji, N. (1995) Optimal strategy of plant antiherbivore defense: Implications for apparency and resource-availability theories. *Ecological Research* **10**, 19–30.

Zangerl, A.R. & Berenbaum, M.R. (2003) Phenotype matching in wild parsnip and parsnip webworms: Causes and consequences. *Evolution* **57**, 806–15.

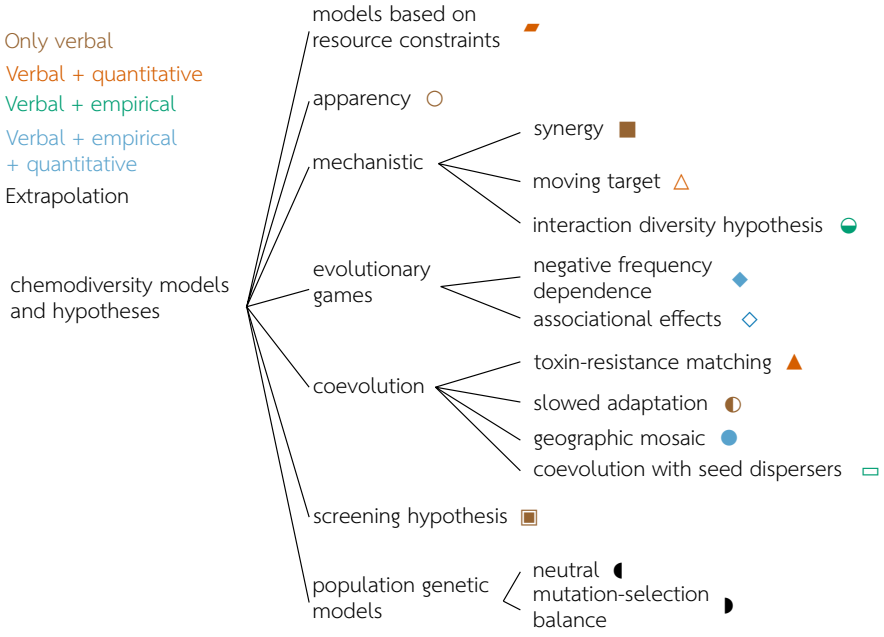

Figure 1

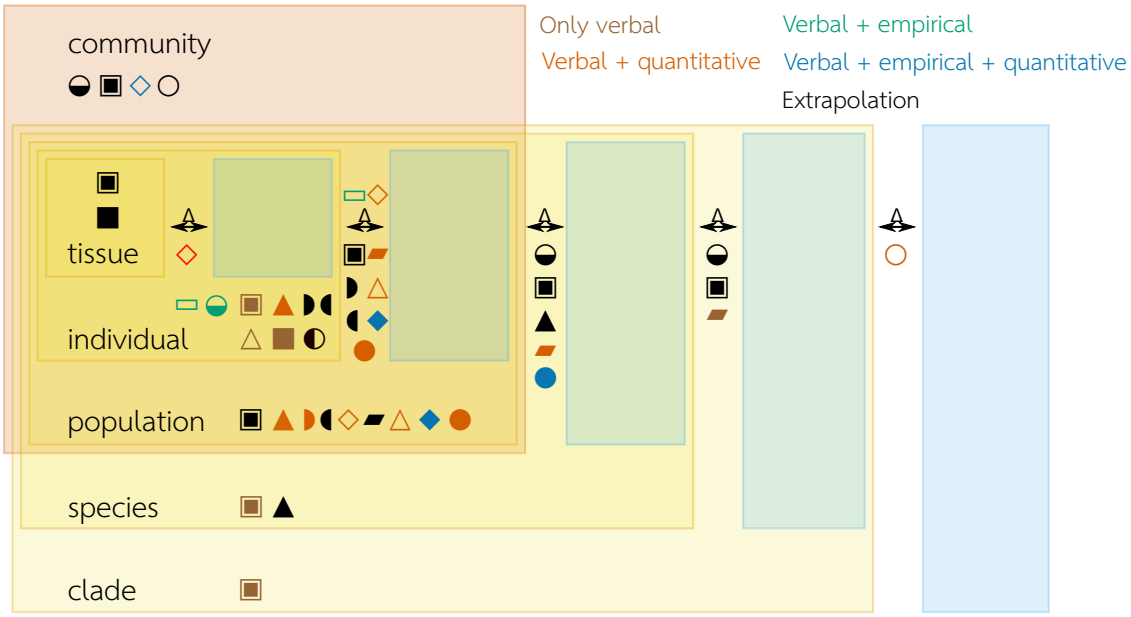

Figure 2

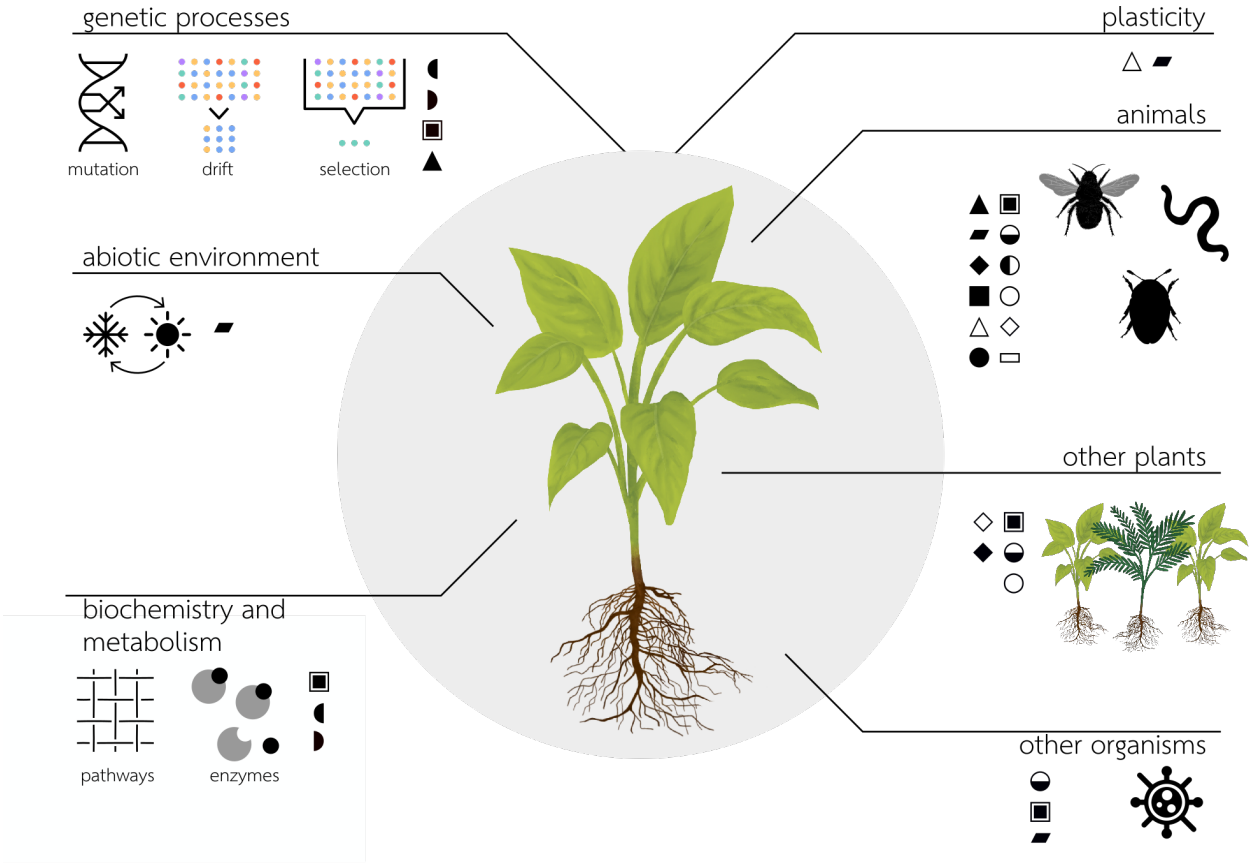
genetic processes
mutation
drift
selection
plasticity
animals
abiotic environment
other plants
biochemistry and
metabolism
pathways
enzymes
other organisms

Figure 3

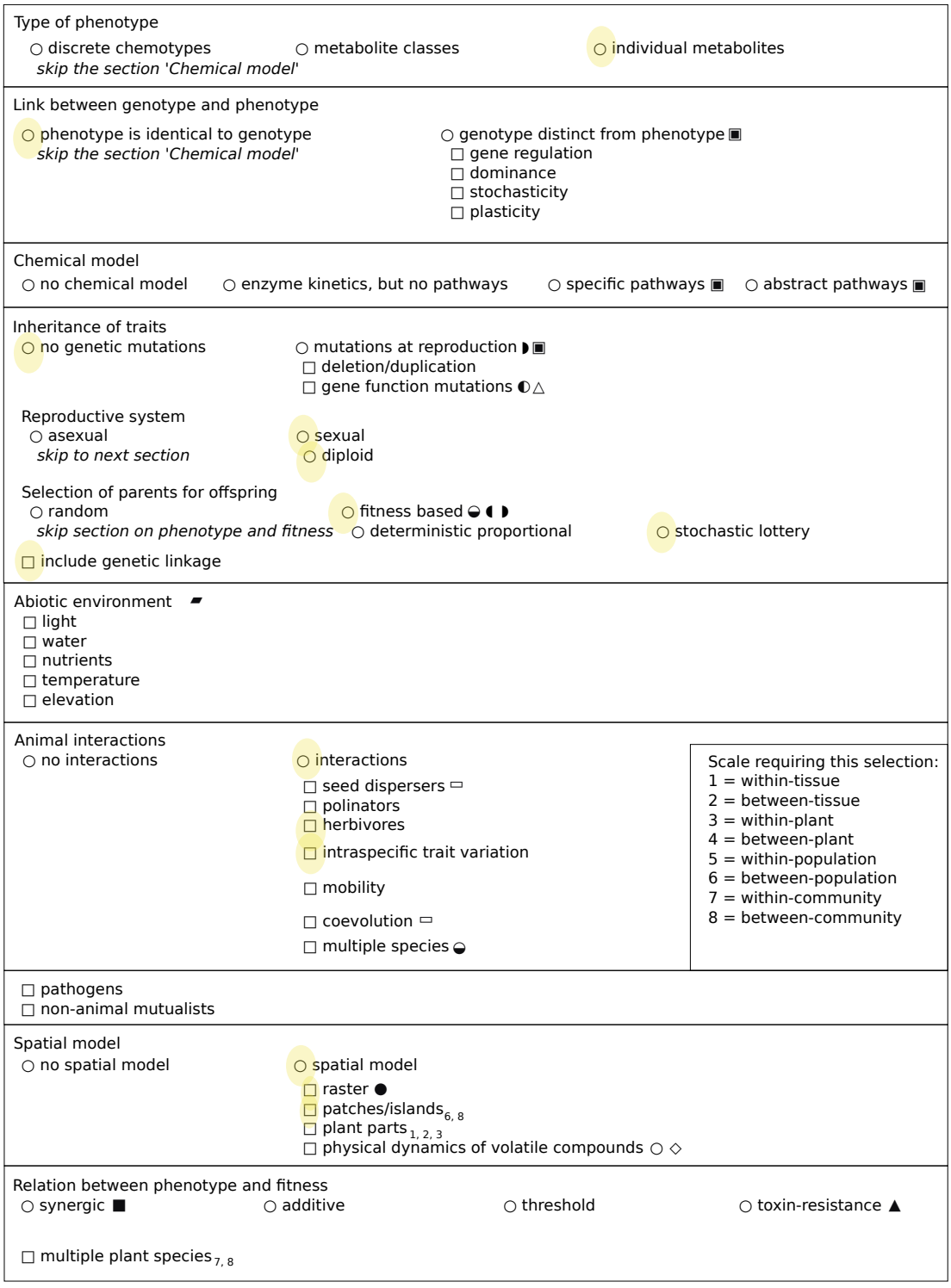
Type of phenotype
○ discrete chemotypes
○ metabolite classes
○ individual metabolites
skip the section 'Chemical model'
Link between genotype and phenotype
○ phenotype is identical to genotype
skip the section 'Chemical model'
○ genotype distinct from phenotype ▣
☐ gene regulation
☐ dominance
☐ stochasticity
☐ plasticity
Chemical model
○ no chemical model
○ enzyme kinetics, but no pathways
○ specific pathways ▣
○ abstract pathways ▣
Inheritance of traits
○ no genetic mutations
○ mutations at reproduction ◗▣
☐ deletion/duplication
☐ gene function mutations ◐△
Reproductive system
○ asexual
skip to next section
○ sexual
○ diploid
Selection of parents for offspring
○ random
skip section on phenotype and fitness
○ fitness based ◒◖◗
○ deterministic proportional
○ stochastic lottery
☐ include genetic linkage
Abiotic environment ▰
☐ light
☐ water
☐ nutrients
☐ temperature
☐ elevation
Animal interactions
○ no interactions
○ interactions
☐ seed dispersers ▭
☐ polinators
☐ herbivores
☐ intraspecific trait variation
☐ mobility
☐ coevolution ▭
☐ multiple species ◒
Scale requiring this selection:
1 = within-tissue
2 = between-tissue
3 = within-plant
4 = between-plant
5 = within-population
6 = between-population
7 = within-community
8 = between-community
☐ pathogens
☐ non-animal mutualists
Spatial model
○ no spatial model
○ spatial model
☐ raster ●
☐ patches/islands 6, 8
☐ plant parts 1, 2, 3
☐ physical dynamics of volatile compounds ○ ◇
Relation between phenotype and fitness
○ synergic ■
○ additive
○ threshold
○ toxin-resistance ▲
☐ multiple plant species 7, 8

Figure 4

Box 1: Ingredients of chemodiversity models

Here we explain the biological processes that act as key ingredients (Fig. 3) in current models and hypotheses for chemodiversity and, without aspiring to completeness, give some exemplary empirical studies supporting their role in shaping chemodiversity. How the ingredients are used in the various theoretical models is explained in section 3 and Box 4.

- **Genetic processes**: SMs are produced by enzymes, which are coded for in genes, which in turn have regulatory genes. Therefore, genetic processes can give rise to chemodiversity. Genes for enzymes in the same metabolic pathway are often organized as operons, i.e. located closely together in the genome (Hamberger & Bak, 2013). Within these operons, gene duplication and subsequent divergence is common, which facilitates a flexible gene network that can undergo rapid evolution (Hamberger & Bak, 2013).

- **Mutations** may change the function of the enzyme coded for by a gene, so it has a different substrate profile or catalyzes different reactions and therefore produces different SMs than its ancestor. Meanwhile changes in gene regulation may change the abundance of SMs, and duplication and deletion may open new pathways of SM synthesis or close them off. Random mutations thus bring about genetic and consequently chemical diversity.

- **Genetic drift** is the phenomenon of changing allele frequencies through chance effects on reproduction and survival rather than natural selection based on fitness. For example, a gene might become fixed or lost randomly within one population but not another, resulting in chemodiversity between populations.

- **Natural selection** occurs when SMs which confer net fitness benefits become more common, and SMs that confer net fitness costs become more rare. Similarly, if there is a fitness benefit to lack of specificity in the production of SMs, this trait may be selected for.

- **The abiotic environment**, such as water and nutrient availability, temperature, or elevation, can determine how much a plant can invest in growth and defense (Bryant *et al.*, 1983; Coley *et al.*, 1985; Herms & Mattson, 1992; Hamilton *et al.*, 2008; Monson *et al.*, 2021; Smakowska *et al.*, 2016). Moreover, changes in the abiotic environment may cause stress responses that lead to changes in the concentrations of certain SMs or induce biosynthesis of new SMs (Agati & Tattini, 2010).

- **Biochemistry and metabolism**: Metabolic pathways for the production of SMs can involve multiple enzyme-catalyzed reactions and can be branched. Which new SMs can be produced by changes in these

pathways is constrained by the extant pathways. Many of the enzymes involved in SM synthesis are "promiscuous", meaning that they accept multiple substrates and thus can produce multiple different SMs at low cost (Aharoni *et al.*, 2005). Thus, small modifications to the enzyme suite can have a profound impact on the resulting metabolite composition (Moghe & Last, 2015; Shoji, 2019).

- **Phenotypic plasticity**: Phenotypic plasticity can explain chemodiversity even among genetically identical individuals in a population. For example, an individual in a nutrient-poor micro-environment may invest these resources differently from an individual in a nutrient-rich micro-environment, leading to variation in chemical composition between individuals in a population (Defossez *et al.*, 2021; Stamp, 2003).

- **Interactions with animals** play a role in the majority of hypotheses and models for the evolution of chemodiversity and their importance has broad empirical support (Calf *et al.*, 2018; Hambäck *et al.*, 2014; Li *et al.*, 2020; Whitehead *et al.*, 2021). For example, in the bittersweet nightshade *Solanum dulcamara*, slugs showed a preference for leaves from populations which produced fewer glycoalkaloids (Calf *et al.*, 2018). These plants belonged to populations in which few slugs were present, hinting at local adaptation of plant populations. SMs can also serve as visual, olfactory or gustatory signals to attract pollinators and seed dispersers, and influence seed dispersion by influencing gut retention time (Baldwin *et al.*, 2020; Borghi *et al.*, 2021; Cipollini & Levey, 1997; Nevo *et al.*, 2018).

- **(Indirect) interactions with other plants:** Plants can interact directly with each other, for example through SMs that inhibit the germination or growth of heterospecific competitors (i.e allelopathy). In the black mustard *Brassica nigra* for example, the production of sinigrin was found to be advantageous when among heterospecifics but not when among conspecifics (Lankau & Strauss, 2007, 2008). Interactions with other plants can also be indirect, often through interactions with other species in the environment such as herbivores (associational effects, Hambäck *et al.*, 2014; Salazar & Marquis, 2022).

- **Interactions with other organisms**: Apart from animal interaction partners, numerous microbial mutualists are guided by plant SMs to their host, and the formation of symbiotic structures is induced by SMs (De la Peña & Loyola-Vargas, 2014), for example in interaction with rhizobia (Cooper, 2004) and arbuscular mycorrhizal fungi (AMF) (Akiyama *et al.*, 2005). Differences in root exudates among populations within species have been found to lead to distinct soil chemical communities (Mueller *et al.*, 2020). Mutualists can in turn also affect patterns of chemodiversity. For example, endophytes

and a symbiosis with AMF resulting in functional arbuscular mycorrhiza can modify the chemical composition of different plant tissues or organs (Schweiger & Müller, 2015; Yadav *et al.*, 2022).

Box 2: Scales of chemodiversity

Chemodiversity can be quantified at various scales of organization (Wetzel & Whitehead, 2020). Terms such as richness and evenness, α,β and γ diversity that are familiar from species diversity are frequently used. However, different authors use different definitions for these terms (see e.g. discussion in Kessler & Kalske, 2018; Li *et al.*, 2020; Wetzel & Whitehead, 2020), depending on the scales their work focuses on. We argue that using the same terms to speak about different scales can be confusing, especially when comparing studies, and when there are more than two scales that may be of interest. Therefore, rather than redefining the terms α, $\beta$and γ depending on the scales a particular study or model works with, we name these scales explicitly.

We argue that instead of using α, β, and γ diversity, it is more helpful to explicitly state the scales under consideration (tissue, individual, population, species, clade, or community) and additionally specify whether diversity within units at those scales or differences between units at those scales are considered (Fig. 2). For example, one can quantify chemodiversity within individuals as the average number of SMs produced per individual, and between-population chemodiversity as number of SMs that are produced only by one of the populations. Note that between-unit chemodiversity is not the same as within-unit chemodiversity at the next higher scale. For example, the average number of metabolites that are not shared between randomly picked individuals in the population is not the same as the total number of metabolites in the whole population. At each scale, chemodiversity could be quantified using different metrics, e.g. metrics focusing on richness, evenness, or disparity (Petrén *et al.*, 2023). In Fig. 2, we place all hypotheses at the scale or scales at which they have been shown to explain, claim to explain, or can be logically inferred to explain chemodiversity.

Box 3: Biological invasions as testing ground for chemodiversity models

Connecting models for the evolution of chemodiversity to empirical data can be challenging because in natural populations we can usually only observe a snapshot of the evolutionary process at the current time, but the long evolutionary history remains hidden. Biological invasions of plants or interacting animals offer exciting opportunities for testing these hypotheses because here the interaction partners often experience sudden large changes in selection regime, with different locations providing 'replicate experiments', sometimes resulting in rapid evolution on ecological time scales (Cox, 2004). Moreover, understanding the mechanisms facilitating invasions may be important to develop conservation strategies.

Invasive species may particularly benefit from high intraspecific variation in chemical composition, i.e. chemodiversity on a population level. For example, tansy *Tanacetum vulgare* (Asteraceae) individuals are highly diverse in their terpenoid profiles, with different terpenoids occurring in different concentrations (Wolf *et al.*, 2011). Terpenoids have various ecological functions (Cheng *et al.*, 2007), and effects on antagonists are highly species-specific. Therefore, a high chemodiversity within invasive plant populations may impede the adaptation of herbivores and microorganisms native to the habitat (Tewes *et al.*, 2018; Wolf *et al.*, 2011), as predicted by the slowed adaptation hypothesis. Furthermore, the simple unfamiliarity of native antagonists with the novel SMs is thought to provide a competitive advantage to *T. vulgare*. This is called the 'novel weapons hypothesis' (Callaway & Ridenour, 2004) in the context of biological invasions and similar to the moving target hypotheses described above in that both confront antagonists with random and unfamiliar SMs.

Furthermore, multiple introductions from different source populations to areas where the plant became invasive may have led to hybridisation of individuals of different chemotypes, producing novel mixtures of terpenoid profiles and thus novel chemotypes. Indeed, a few chemotypes were only found in the invasive range of *T. vulgare* (Wolf *et al.*, 2011). Moreover, potential inbreeding in invasive populations and environmental conditions different from the originating environment may lead to rapid evolutionary changes (Schrieber & Lachmuth, 2016), which may result in novel SM variation. Overall, changes in chemodiversity and the functions in species interactions in non-native populations have been largely neglected (Tewes *et al.*, 2018), although chemical composition can highly vary between native and non-native populations (Pankoke *et al.*, 2019; Tewes *et al.*, 2018) and due to hybridization (Piola *et al.*, 2013), offering great opportunities to develop theoretical models and test them with empirical studies. The novel-weapons hypothesis has already

been mathematically modeled (Allstadt *et al.*, 2012), albeit not as a chemodiversity model, and is ripe for expanding upon (see Box 4).

Box 4: Constructing a generalized chemodiversity individual-based model

When developing an individual-based model to investigate chemodiversity, one needs to carefully consider what to include in the model, based on what one wants to investigate with it. Different questions require different models with different ingredients. However, the parts that different models are built from can be re-used across models. Considering different mechanisms that could describe how a particular ingredient (Box 1) influences chemodiversity also means that you can construct your model in a modular way so that you can compare these mechanisms to each other and see which model explains empirical results best.

This does not mean that every individual-based model of chemodiversity has to be able to test all possible mechanisms involved with the evolution and maintenance of chemodiversity. A model should include no more complexity than it requires (Servedio *et al.*, 2014), and there is also the risk of overfitting in overly complex models. Nevertheless, one should be mindful of alternative explanations when making a model that investigates a particular set of hypotheses. A model can show a verbal hypothesis to be flawed if the stated logic cannot be mathematically reproduced, and a model that can test multiple proposed mechanisms can give clues to which of these mechanisms may have the strongest impact on the origin and maintenance of chemodiversity (Otto & Rosales, 2019).

We propose the workflow described by Fig. 4 and the more elaborate Supplement 1. This is a workflow for creating individual-based models of chemodiversity based on individual plants, and is most useful for models that describe the origin and maintenance of chemodiversity up to the multi-species community or multiple-population level.

For the example from Box 3, a possible set of selections to investigate chemodiversity in biological invasion is marked in Fig. 4. Here, the model of Allstadt *et al.* (2012), which is already an IBM with two types of plants (resident and invader) competing in a raster, is expanded by selecting several options that are not considered in the original model. As empirical studies tend to focus on variation within a metabolite class (here for an example in invasion biology, see Wolf *et al.* (2011)), chemotype in this proposed model is modeled through individual metabolites. The model has sexual reproduction without mutations but with

genetic linkage to focus on the population genetics of biological invasions, and interactions with herbivores are modeled explicitly in this version. Lastly, the origin of the invading plants is modeled as an abstract island that individuals with the invading genotype might arrive from.

Caption Figure 1: Tree showing a rough conceptual grouping of chemodiversity models and hypotheses. Each model or hypothesis has a unique symbol, which is then used in Fig. 3 to indicate which ingredients are part of the model, in Fig. 2 to indicate at what scales the model can explain chemodiversity, and in Fig. 4 to indicate which model components are essential to any model that investigates these hypotheses. Note that similarity in shape between symbols here does not indicate any relationship between models. Models with brown symbols have only a verbal model, green symbols a verbal model and empirical evidence, red a quantitative model, and blue both a quantitative model and empirical evidence. Black symbols stand for models that have, to our knowledge, not been previously used in the context of chemodiversity but that we argue can be extended to modeling chemodiversity (i.e. they are our extrapolation).

Caption figure 2: Schematic overview of the scales at which the different models and hypotheses (potentially) explain the evolution of chemodiversity. Please refer to Fig. 1 for the models each symbol refers to. Models with brown symbols have only a verbal model, green symbols additionally have empirical evidence, red a quantitative model, and blue both a quantitative model and empirical evidence. Black symbols stand for models that have, to our knowledge, not been previously used in the context of chemodiversity but that we argue can be extended to modeling chemodiversity (i.e. they are our extrapolation). For example, the filled blue diamond next to the label "population" indicates that there is a quantitative model and also empirical evidence suggesting that negative frequency-dependence can explain the maintenance of chemodiversity within a population and the filled red circle between the two "individual" boxes indicates that a quantitative model based around a geographic mosaic exists that suggests that it can maintain differences between individuals. For some symbols, we used different colors in different places in the figure to indicate a distinct type of support for chemodiversity at the various scales. The community level is orthogonal to the species level because populations can be considered nested both in their species and in a multi-species community.

Caption figure 3: Overview of the key ingredients of chemodiversity models. To see which hypothesis/model a symbol stands for, please refer to Fig. 1. The ingredients listed in the figure are further explained in Box 1. Models and hypotheses are listed as connected to particular ingredients when these ingredients are key components of the model as described below in section 3 and Box 4.

Caption figure 4: Flowchart of the decision making process for making a individual-based model to investigate chemodiversity. Every section represents one important 'module' in the model, with one or more elements to pick decisions on. Top-level decisions are aligned left, while indentation indicates that a decision should only be considered if the option above was picked. Circles represent mutually exclusive decisions, with the possible options presented in a horizontal row. Squares represent decisions where multiple options can be picked, and are presented in vertical columns. The yellow ovals are a set of possible options for an IBM inspired by (Allstadt *et al.*, 2012) which could investigate the development of chemodiversity under biological invasion (Box 3). Symbols behind options or section names indicate that this option or any option from this section is necessary for a model that aims to investigate the hypothesis corresponding to this symbol (Fig. 1), while numbers indicate this option is necessary to investigate chemodiversity on the corresponding scale (Fig. 2). An interactive R program with a more detailed decision tree can be found in SI. 1.