Interaction diversity explains the maintenance of phytochemical diversity

Susan R. Whitehead[1,2], Ethan Bass[2,3], Alexsandra Corrigan[1,4], André Kessler[3], and Katja Poveda[2]

[1]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA USA

[2]Department of Entomology, Cornell University, Ithaca, New York, USA

[3]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA

[4]Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA USA

Author emails: swhitehead@vt.edu, eb565@cornell.edu, alex7@vt.edu, ak357@cornell.edu, kap235@cornell.edu

Running Title: Maintenance of phytochemical diversity

Keywords: secondary metabolites, specialized metabolites, plant-insect interactions, plant-pathogen interactions, synergy, interaction diversity hypothesis, screening hypothesis, *Malus*, phenolics

Article Type: Letters

Length: Abstract (147), Main Text (5000), References (44), Figures (4), Tables (1), Text Boxes (0)

Correspondence: Susan R. Whitehead; 220 Ag Quad Ln (MC 0390), Blacksburg, VA 24061; 434-841-1512; swhitehead@vt.edu

Statement of Authorship: S.R.W., E.B., A.K., and K.P. developed the initial conceptual framework and methods for insect experiments; A.C. and S.R.W. developed methods for fungi experiments; E.B., A.C., and S.R.W. collected data; K.P. and S.R.W. provided access to resources and equipment; S.R.W. conducted analyses and wrote the first draft of the manuscript, and all authors contributed to edits and revisions.

Data Accessibility Statement: All data associated with the manuscript are accessible in Github (WhiteheadLabVT/Phytochemical-Diversity-Experiment) and archived in Zenodo (DOI: 10.5281/zenodo.4586758).

Abstract

The production of complex mixtures of secondary metabolites is a ubiquitous feature of plants. Several evolutionary hypotheses seek to explain how phytochemical diversity is maintained, including the synergy hypothesis, the interaction diversity hypothesis, and the screening hypothesis. We experimentally tested predictions derived from these hypotheses by manipulating the richness and structural diversity of phenolic metabolites in the diets of eight plant consumers. Across 3940 total bioassays, there was clear support for the interaction diversity hypothesis over the synergy or screening hypotheses. The number of consumers affected by a particular phenolic composition increased with increasing richness and structural diversity of compounds. Furthermore, the bioactivity of phenolics was consumer-specific. All compounds tested reduced the performance of at least one consumer, but no compounds affected all consumers. These results show how phytochemical diversity may be maintained in nature by a complex selective landscape exerted by diverse communities of plant consumers.

Introduction

Individual plants can produce hundreds of secondary metabolites from diverse chemical classes. Although many metabolites have demonstrated roles in plant ecology (Iason *et al.* 2012), a central question remains unanswered: why do plants produce so many different metabolites simultaneously? Recent studies have shown links between phytochemical diversity and herbivore community structure in several systems (e.g. Richards *et al.* 2015; Salazar *et al.* 2018; Volf *et al.* 2018). Still, the mechanisms through which herbivores and other plant consumers may select for complex phytochemical phenotypes remain unclear. This may be due, in part, to difficulties in untangling the multiple dimensions of phytochemical diversity, including the richness, evenness, structural complexity, and spatiotemporal variation of compounds. These different components of diversity have unique evolutionary origins and ecological consequences, but addressing them independently in natural systems is challenging because they do not vary independently (Wetzel & Whitehead 2020). Here, we use a novel experimental approach to manipulate different dimensions of phytochemical diversity and test predictions derived from three major hypotheses that explain why plants produce so many compounds. Under all of these hypotheses, diversity may be generated initially by various underlying biochemical mechanisms (Gershenzon *et al.* 2012). However, the three hypotheses differ in their explanations for how phytochemical diversity is maintained by natural selection.

First, the synergy hypothesis suggests that phytochemical diversity is maintained because greater-than-additive interactions among compounds increase the effectiveness of plant defense. Under this scenario, selection is acting on suites of metabolites that represent complex integrated phenotypes (Pigliucci & Preston 2004). Synergisms that increase herbivore resistance have been demonstrated in a number of systems (Richards *et al.* 2016). Most studies have focused on interactions among closely related compounds, but there are also examples of interclass synergy (Berenbaum & Neal 1985; Steppuhn & Baldwin 2007), and evidence from the pharmaceutical literature suggests synergies are

more likely to occur among structurally dissimilar compounds (Liu & Zhao 2016). Thus, synergies could be a main driver of compound diversification within and across classes, but this is still unclear because most synergy studies have involved only two or very few compounds. Compounds can also interact antagonistically (Diawara *et al.* 1993; Whitehead & Bowers 2014; Liu *et al.* 2017), and antagonisms could also increase with phytochemical diversity. If compound synergies drive the evolution of phytochemical diversity, the major prediction is not just that beneficial synergies exist, but that they accumulate as the diversity (including richness, evenness, and structural diversity) of a mixture increases, providing a selective advantage to plants that produce complex mixtures over those that produce just a few potent compounds.

Second, the interaction diversity hypothesis posits that plants produce numerous compounds because they interact simultaneously with numerous organisms. No single metabolite or class of metabolites can protect plants against all enemies simultaneously because different organisms vary in their susceptibility to different molecular mechanisms of bioactivity. Additional metabolites may function to attract or manipulate mutualists, including pollinators, seed dispersers, and microbial symbionts (Schiestl 2015; Choi *et al.* 2018; Nevo & Valenta 2018). Under the interaction diversity hypothesis, selection acts independently on different compounds that confer different advantages in different pairwise interactions, and phytochemical diversity is an emergent property of that process. This basic hypothesis has long been discussed in the literature (Berenbaum & Zangerl 1996; Iason *et al.* 2011), but has often been treated somewhat concurrently with the synergy hypothesis because both are united by the idea that most compounds in a mixture serve an adaptive role. However, we distinguish between these hypotheses (also see Gershenzon et al. 2012) because they have different underlying mechanisms and predict different effects of secondary metabolites on plant consumers. If interaction diversity drives the evolution of phytochemical diversity, then more diverse mixtures will not necessarily provide increased

resistance to any one consumer (as predicted by the synergy hypothesis), but they will provide increased overall resistance against diverse communities of antagonists.

Finally, the screening hypothesis suggests that phytochemical diversity is maintained, even though most compounds provide no fitness benefits, because selection acts on biochemical mechanisms that generate diversity (e.g. enzymes with broad substrate tolerance). A key tenet is that potent biological activity is a rare property among all possible molecular structures (Jones *et al.* 1991; Firn & Jones 2003). Thus, individual plants that make and "screen" a large number of metabolites will have a selective advantage because they are more likely to produce a few novel compounds that provide resistance. Selection for biochemical mechanisms that generate diversity is not, by itself, mutually-exclusive with the synergy or interaction diversity hypotheses. However, the screening hypothesis posits that this selective pressure drives the maintenance of phytochemical diversity even though most compounds provide no fitness benefits (Jones *et al.* 1991; Firn & Jones 2003). Thus, a major assumption of the screening hypothesis–that most compounds are not bioactive–can be considered a null alternative to the synergy and interaction diversity hypotheses, both of which assume that most compounds are adaptive. Because it assumes most compounds are not active, the screening hypothesis also predicts that the presence of a few potent compounds, rather than the full multivariate plant chemotype, should explain most variation in consumer performance.

The three hypotheses described above do not provide an exhaustive list of all possible evolutionary mechanisms that could explain phytochemical diversity. For example, diverse mixtures of phytochemicals may slow the evolution of consumer counter-adaptations (Palmer-Young *et al.* 2017), provide more specific communication with mutualists (Gershenzon *et al.* 2012; Nevo *et al.* 2018), or be generated by fluctuating selection during co-evolutionary arms races (Speed *et al.* 2015). We focus on the synergy, interaction diversity, and screening hypotheses because they are the most widely-

considered hypotheses that explain phytochemical diversity and were experimentally tractable in our system.

To compare support for the three hypotheses, we manipulated the richness and structural diversity of phenolic metabolites in the diets of four species of insect herbivores and four strains of plant-pathogenic fungi. We focused on phenolics because they are ubiquitous in plants, have known roles in defense, and are highly diverse in terms of both richness and structural variation (Lattanzio *et al.* 2009). We generated mixtures from a pool of 14 compounds and compared their structural diversity using several metrics that describe the dissimilarity of structural features in pairs of metabolites (Cao *et al.* 2008; Kim *et al.* 2016). In total, we tested 59 unique phytochemical compositions across six levels of compound richness and three levels of structural diversity in a series of bioassays with 2,756 caterpillars and 1,024 fungal cultures. For each of the three hypotheses described above, we formulated a set of predictions for how phytochemical diversity should affect the plant consumers in our experiments (Table 1).

Material and Methods

Compound Mixtures

We focused on phenolic metabolites, defined here as phytochemicals that possess an aromatic ring with one or more hydroxyl groups and arise from the shikimate/phenylpropanoid pathway and/or the polyketide/malonate pathway (Lattanzio *et al.* 2009). Fourteen compounds were selected from different sub-classes, including benzoic acids (gallic acid, syringic acid, and gentistic acid), hydroxycinnamic acids (chlorogenic acid, caffeic acid, p-coumaric acid, and ferulic acid), flavonols (quercetin, rutin, and hyperin), flavan-3-ols (catechin and epicatechin), and dihydrochalcones (phloridzin and phloretin). All co-occur in the fruits of wild and domesticated apples (accounting for ~66% of total phenolics), a system

where we previously detected a correlation between phytochemical diversity and herbivore resistance (Whitehead and Poveda 2019). Compounds (> 95% purity) were purchased commercially from Sigma-Aldrich (St. Louis, MO, USA), Chengdu-Push Biotechnology (Chengdu, China), and Toronto Research Chemicals (Toronto, Canada).

We tested each compound individually and also created random mixtures of 2, 4, 6, 8, and 10 compounds across three levels of structural diversity (45 total mixtures; Table S1). To choose mixtures, we first assigned a value for structural diversity to all possible mixtures in each richness level using three metrics (atom pairs, maximum common substructure, and PubChem fingerprints) in the R package ChemmineR (Cao et al. 2008). We designated the lowest 10% of values as "low", the middle 10% as "medium", and the highest 10% as "high". We then semi-randomly selected three mixtures from each richness/structural diversity combination using several *a priori* criteria to minimize overlap of individual compounds across mixtures. Full details on mixture selection are provided in Appendix S1a.

Phenolics were dissolved in dimethyl sulfoxide (DMSO) and added to growth media. The concentration of total phenolics in the media was 0.2% fresh mass for all insect trials and 0.1% for fungi, and DMSO was always 0.5%. Control media was prepared with 0.5% DMSO only. Phenolic concentration was chosen to mimic apple fruits (mean: 0.2%, range: 0.06% - 0.43% across 31 cultivars; Neveu et al. 2010) where the 14 compounds co-occur. We used concentrations from the lower end of the range for fungi because preliminary trials showed that many compounds fully inhibited growth at 0.2%, limiting our ability to assess variation among treatments. In the main experiments, all compounds were included in the mixture in equal relative mass abundance (not equimolar concentrations; see Appendix S1a for a discussion of this choice and its implications). A second experiment, conducted only for insects, manipulated the relative mass abundances (evenness) of compounds and is described in Appendix S1c.

Study organisms

We chose four lepidopteran species, including two generalists, *Helicoverpa zea* Boddie (corn earworm; Noctuidae) and *Spodoptera frugiperda* (J.E. Smith) (fall armyworm; Noctuidae), one adapted specialist, *Cydia pomonella* (Linnaeus) (the codling moth, an apple fruit-feeding specialist; Tortricidae), and one non-adapted specialist, *Plutella xylostella* (Linnaeus) (diamondback moth; Plutellidae), which feeds on phenolic-rich plants, but with (mostly) different individual compounds than those tested here. We chose four fungal strains from a culture collection of taxa isolated from diseased apples in the laboratory of Dr. Anton Baudoin (Virginia Tech). These were tentatively identified based on morphology and disease symptoms as *Botryosphaeria dothidea* (Moug. ex Fr) Ces. & De Not (Botryosphaeriaceae; white rot), *Colletotrichum* sp. (Glomerellaceae; bitter rot), *Penicillium expansum* Link (Trichocomaceae; blue mold), and *Sclerotinia sclerotiorum* (Lib.) de Bary (Sclerotiniaceae; calyx end rot).

Insect bioassays

Insect bioassays were conducted between October 2016 and March 2017 at Cornell University. Between 4 and 43 individual caterpillars per species (typically 8 or 9; Table S2) were reared on each unique treatment in 1 oz plastic cups, totalling 2,756 individual caterpillars. The trials were conducted in multiple rounds with controls included in each round and on each tray of cups. Full details of the diet preparation and handling of larvae are provided in Appendix S1a. For each caterpillar, we recorded survival, pupal mass (mg), and days to pupation as metrics of performance. For larvae that pupated, we also determined the sex.

Fungal bioassays

Fungal bioassays were conducted between March and July 2019 at Virginia Tech. Growth assays were conducted in 96-well plates with four replicate wells per treatment. Controls were included on each plate, for a total of 1,024 growth assays (59 phenolic compositions x 4 species x 4 replicates = 944, plus 80 controls). Full details of the plate preparation are provided in Appendix S1a. Once a plate was prepared, we took an initial reading of the absorbance for each well using a plate reader (Biotek Synergy HT) at 600nm. Three additional readings were taken per plate over 72 hours, with 22-26 hours between readings.

Data standardization

Prior to analyses, we standardized all data relative to controls. For insect pupal weights, we divided the weight of each experimental insect by the mean weight of the relevant controls (insects of the same species and sex from the same experimental round). For days to pupation, because lower values indicate better performance, we did the inverse calculation (control divided by experimental; hereafter, "development speed"). For insect survival, data were first summarized as the proportion of survivors per treatment, then divided by the proportion of survivors for all relevant controls. For fungal growth, we used linear models to determine the slope of the relationship between time (0, 24, 48, and 72 hours) and mean absorbance of the well. This was standardized by dividing the mean slope for each treatment by the mean slope for all controls for that fungal strain. Because there were a few cases where slopes were negative, we added one as a constant to all slope values prior to this calculation to make all values positive.

Statistical analyses

General. All analyses were conducted in R v. 4.0.2 using the packages lme4 (Bates *et al.* 2020), multcomp (Hothorn *et al.* 2020), MuMIn (Bartoń 2019), and MASS (Ripley *et al.* 2020) and figures were prepared using ggplot2 (Wickham *et al.* 2020). All analyses are described in further detail in Appendix S1b and data and associated R scripts are archived in Zenodo (DOI: 10.5281/zenodo.4586759). We used diverse approaches to test predictions derived from our three hypotheses (Table 1), and it is important to note that statistical power varies across these analyses. In several cases (Predictions 1d, 1e, 2a, 2b), we used approaches in which coefficients from one set of models are used as derived metrics in further models. In these cases, we performed null models with 1000+ iterations of randomized data to confirm that the intrinsic distributions in derived metrics did not bias the results toward support or rejection of hypotheses (see Appendix S1b for more detail).

Synergy hypothesis. To test whether consumer performance was affected by compound richness (Prediction 1a), structural diversity (Prediction 1b), or their interaction (Prediction 1c), we used linear mixed models (LMMs) conducted separately for each performance metric. We first tested models that included richness, structural diversity, consumer species, sex (for pupal mass and development speed only), and all possible interactions as fixed effects. Based on numerous strongly supported interactions involving species and sex, we followed these analyses with LMMs conducted separately for each of the eight taxa and (where relevant) each sex. Next, to test whether the occurrence of synergisms increased with richness (Prediction 1d), we calculated an interaction index for each mixture as the difference between its observed bioactivity and its expected bioactivity based on additive interactions among compounds (Tallarida 2000). We then used linear models (LMs) to test whether the index increased with richness. These analyses were conducted separately for two performance metrics (pupal weight and development speed) for each insect species and for growth rate for each fungal species. Finally, to test whether increasing richness increased the probability that a mixture's bioactivity would exceed the average bioactivity of singletons (Prediction 1e), we created a binary variable that indicated whether a

mixture had a stronger negative effect on performance than the average effect of all singletons in that mixture. We then used a binomial GLM to test whether this probability increased with increasing compound richness in the mixture.

Interaction diversity hypothesis. To test whether the number of consumers affected by phenolics increased with richness (Prediction 2a) or structural diversity (Prediction 2b), we first counted the number of consumers (out of 8 taxa) that were negatively affected by each diet (N = 59 compositions). Then we tested whether that count increased with increasing richness or structural diversity using LMs. A consumer species was considered negatively affected if any metric of its performance was reduced relative to controls, based on 95% confidence intervals for the standardized performance metric (for pupal weights, development speed, and fungal growth) or Fisher's exact tests (for survival). Next, to examine whether most compounds were biologically active against one or a few, but not all, consumers (Predictions 2c & 2d), we used confidence intervals and Fisher's exact tests (as above) to examine the bioactivity of each of the individual compounds. Finally, to test whether the effects of phenolics depend on consumer identity (Prediction 2e), we used LMMs with species, treatment (the specific diet composition), and their interaction as fixed effects. These models were conducted separately for each performance metric. A significant interaction between species and treatment would indicate that the effects of a compound mixture depend on the consumer identity.

Screening hypothesis. To assess support for the prediction that most compounds have no apparent biological activity (Prediction 3a), we examined results from analyses of the effects of individual compounds on consumer performance (described above for Prediction 2c). Next, to test whether variation in consumer performance could be explained by one or a few compounds (Prediction 3b), we constructed LMs for each performance metric that included richness, structural diversity, consumer species, and the presence/absence of each individual compound as fixed effects. We then used stepwise AIC-based model selection (Ripley *et al.* 2020) to determine the minimum adequate model to describe

consumer performance. If variation in performance can be explained by one or a few compounds, then most individual compounds should not be retained in these models. Finally, to test whether the specific mixture of compounds in the diet explains a significant portion of the variance in herbivore performance (Prediction 3c), we used multilevel random intercepts models that partitioned variance in performance among the specific diet treatment, experiment number, and residual variance. We compared these models with and without treatment using likelihood ratio tests and also calculated the variance partition coefficients (i.e. the proportion of total variance in performance explained by the specific treatment).

Results

Synergy hypothesis

Richness and structural diversity do not decrease consumer performance (Predictions 1a-c not supported).

We found no evidence that the performance of plant consumers decreases with increasing richness or structural diversity (Figure 1-2; Figure S1; Tables S3-S5). In fact, in contrast to Prediction 1a, model-averaged coefficients for the relationship between richness and performance were all positive, though not supported statistically (Table S4). The multi-model averaging approach did reveal strongly supported effects of species, sex, and multiple two and three-way interactions involving these variables (Table S4). For fungi, there were also two-way interactions between richness and structural diversity and strong three-way interactions between these two variables and species (Table S4). To further explore these interactions, we used separate models for each species and (for insects) males and females. This included 24 separate models, across which the only supported effect of richness was in the opposite direction than predicted by the synergy hypothesis: as compound richness increased, there was a corresponding increase in *S. frugiperda* survival ($P = 0.043$; Table S5; Figure S1A). There were two cases

with support for an interaction between richness and structural diversity (for *P. xylostella* male development speed and *B. dothidea* growth rates), and one case with support for a main effect of structural diversity (*P. xylostella* male development speed); however, none of these effects were in the direction predicted by the synergy hypothesis (Figure S1, Table S5).

Synergies among compounds do not increase with increasing richness (Prediction 1d not supported)

Across 12 separate LMs (different species/performance metric combinations), we found no evidence that synergies increase with increasing richness (Figure S2; Table S6). There were no models that supported an effect of richness on the interaction index, and trends were in the opposite direction in 10 out of 12 cases (Figure S2; Table S6).

Probability of a mixture's bioactivity exceeding that of singletons does not increase with increasing richness (Prediction 1e not supported)

We found no evidence that increasing richness increased the probability that a mixture's bioactivity would exceed the average bioactivity of its singletons (Figure S3, Table S7). To the contrary, out of 12 tests, there were two cases where the opposite pattern was supported (Fig. S3, Table S7).

Interaction diversity hypothesis

The number of herbivores affected by phenolics increases with increasing richness and structural diversity (Predictions 2a & 2b supported)

As compound richness increased from 1-10 compounds, the average number of consumers negatively affected by phenolics increased by 37% from 3.64 to 5 organisms (t = 3.22; P = 0.0021; Figure 3A). Increasing structural diversity also increased the number of consumers affected, with a significant

difference between low (average 3.83 organisms) versus medium (average 4.8 organisms, + 25%) or high (average 4.73 organisms, +23%) structural diversity (t = -2.52; P = 0.015; Figure 3B).

All individual compounds have some bioactivity, but none are "silver bullets" (Predictions 2c & 2d supported)

Across the 14 phenolic metabolites, all had detectable bioactivity in interactions with insects and/or fungi (Figure 4). However, there were no single compounds that had negative effects on all consumers (Figure 4). Importantly, although all compounds had negative effects on at least one aspect of consumer performance, most (10 of 14) also had some positive effects. Across 224 tests of compound bioactivity (separate tests for each compound, consumer, and metric), there were 46 cases of negative effects and 23 of positive effects on performance (Figure 4). There were clear differences across species in their responses to phenolics; the two specialized insect species, in particular, largely experienced positive effects of phenolics on performance (Figure 4).

Effects of phenolics depend on consumer identity (Prediction 2e supported)

For all performance metrics, there were strong interactions between the specific diet treatment and the consumer identity ($P < 0.001$), strong effects of diet treatment ($P < 0.001$), and strong effects of consumer species ($P < 0001$) (Figure S4; Table S8).

Screening hypothesis

All compounds have some bioactivity (Prediction 3a not supported)

As described above (see Results for Predictions 2c & 2d), all 14 phenolic compounds tested had negative effects on performance for at least one of the consumers, and many compounds (12 of 14) also had some positive effects on performance (Figure 4).

All compounds are necessary to explain variation in performance (Prediction 3b not supported)

All compounds were retained in one or more of the minimum adequate models as significant predictors of performance, and at least half of the compounds were included in each model (Table S9). In all cases, some compounds had negative effects and others had positive effects. For pupal weights, there were four compounds with negative effects ($P < 0.01$), four compounds with positive effects ($P < 0.05$), and six compounds with no or marginal effects on performance. For development speed, there were three compounds with negative effects (the same that negatively affected pupal weights: gentistic acid, hyperin, and quercetin; $P < 0.001$; Table S9), four compounds with positive effects ($P < 0.05$), and seven compounds with no or marginal effects. For survival, there were six compounds with negative effects ($P < 0.002$), five compounds with positive effects ($P < 0.0001$), and three compounds with no or marginal effects. For fungal growth rate, there were six compounds with negative effects ($P < 0.03$), six compounds with positive effects ($P < 0.0005$), and two compounds with no or marginal effects.

Specific mixture of compounds in diet explains substantial variation in performance (Prediction 3c not supported)

For three of four insect species (*H. zea*, *S. frugiperda*, and *P. xylostella*) and all four fungi, all models containing treatment as a random effect were supported over those that contained experiment number only (Table S10). For *C. pomonella* (the apple specialist), models containing treatment as a random effect were strongly supported for development speed ($P = 0.0025$), but not for pupal weight ($P = 0.099$) or survival ($P = 0.544$). The percentage of variation in performance explained by the specific diet treatment was highly variable across species and performance metrics (Table S10) but was generally high for fungal growth (45-59%) as well as the development speed and survival of the two generalists (*H. zea* and *S. fugiperda*; 26-73%). Treatment explained a relatively small percentage of variance in pupal weights (3-9% across species) and in performance for *C. pomonella*, the apple specialist (3.2-6.7%).

352

353 **Discussion**

354 Our results provide clear support for the interaction diversity hypothesis over the synergy hypothesis or
355 the screening hypothesis (Table 1). Despite highly variable bioactivities of different randomly selected
356 mixtures, we saw a clear signal indicating that more diverse mixtures are effective against a larger
357 diversity of consumers (37% increase in the number of consumers affected by 10-compound mixtures vs
358 single compounds; Figure 3). Furthermore, all compounds showed some bioactivity (Figure 4), but the
359 relative bioactivities of individual compounds and mixtures were highly dependent on the consumer
360 species being tested (Figure 4, Figure S4). These experimental results provide new insight into the
361 mechanisms that drive patterns in natural systems, where phytochemical diversity can be correlated
362 with plant resistance at multiple scales: among individuals (Whitehead *et al.* 2013; Whitehead & Poveda
363 2019), among species (Richards *et al.* 2015; Salazar *et al.* 2018), and among communities (Salazar *et al.*
364 2016; Bustos-Segura *et al.* 2017).

365 In contrast to expectation, we found no evidence in support of the synergy hypothesis (Figure 1-2, S1-
366 S3; Table S3-S7). It is intuitive to assume that novel functions should emerge in complex mixtures, just as
367 classic experiments have shown that increasing species diversity in a plant community increases
368 productivity (Tilman *et al.* 2001). However, our results suggest that, at least among phenolic
369 metabolites, strong synergies may be a relatively rare occurrence (also see Stamp & Osier 1998). Out of
370 540 total tests for synergy, we found only 31 cases that were supported at 95% confidence, and
371 antagonisms were nearly as common, with 23 cases (Figure S2). Most critically for our hypothesis, the
372 probability of detecting a synergy did not change, or, in several cases, decreased (Fig. S2-S3) with
373 increasing phytochemical richness. Certainly synergies could have been more common in a different
374 compound class (e.g. among furanocoumarins; Berenbaum & Zangerl 1993) or across compound classes

(e.g. alkaloids and protease inhibitors; Steppuhn & Baldwin 2007). It is also possible that synergies would have been more common if compounds had been tested in equimolar rather than equal mass concentrations (see Appendix S1a), or if we had assessed synergies with more robust methods that rely on dose-response curves for individual compounds and mixtures (Tallarida 2000). Overall, our results should not be interpreted as robust evidence (or lack thereof) for any specific compound interaction, and they do not rule out synergy as an important factor shaping phytochemical phenotypes. However, the lack of relationship between diversity and synergy in our well-replicated experiment suggests that, at least for phenolics, synergies are not the major driver of phytochemical diversity.

Our results are also inconsistent with the screening hypothesis, which assumes that secondary metabolite bioactivity is rare (Firn & Jones 2003). Instead, every compound we examined had multiple modes of bioactivity (Figure 4). It is possible that these compounds represent a biased subset of phenolics. Wild apples contain at least 80 phenolics at detectable quantities (Whitehead & Poveda 2019), and we selected 14 based on commercial availability. These tended to be compounds produced in high abundance in apples and/or distributed widely across plants, which may have a higher probability of bioactivity than rare or low abundance compounds. Thus, our results cannot disprove the screening hypothesis. Certainly its main evolutionary mechanism—selection for promiscuous enzymes and other biochemical mechanisms that generate diversity—is likely occurring for phenolics and other metabolites. For example, arbutin synthase, a glucosyltransferase, can accept at least 35 highly diverse phenolic metabolites as substrates (Hefner *et al.* 2002). However, the screening hypothesis posits that this selection pressure maintains phytochemical diversity even though most compounds are not bioactive. In contrast, our results suggest that, in interactions with biologically relevant organisms, secondary metabolite bioactivity is common (also see Thompson *et al.* 1985).

Importantly, our results emphasize that bioactivity is not a singular property of a molecule, but can vary across consumers and different aspects of physiology. Many of the compounds or mixtures we tested

had negative effects on generalist herbivores, but neutral or positive effects on specialists (Figure 1, Figure 4; Figure S4). For example, there were no individual compounds that decreased the performance of the specialist herbivore, *C. pomonella*, but six compounds that increased its performance, including several that negatively affected generalist herbivores or fungi (Figure 4). There were also conflicting effects on different aspects of a single consumer's performance. For example, quercetin decreased the development speed of *H. zea* but increased its pupal mass, and gentistic acid did the opposite (Figure 4). In natural systems, these same compounds may also affect mutualists, in some cases leading to trade-offs between attraction of mutualists and defense against antagonists (Cipollini & Levey 1997; Adler 2000). Furthermore, the bioactivity of any given compound or mixture is likely strongly dependent on the biochemical environment, including pH, nutrient availability, or the presence of detoxification enzymes (Appel 1993; Wu *et al.* 2015). The nutrient contents of our diet, especially nitrogen and iron, were high relative to typical plant material (Table S15). Because some phenolics act by limiting availability of these nutrients (Lattanzio *et al.* 2009), their bioactivities in our experiment may differ from that in natural systems. These variable bioactivities across consumers and environments, combined with a complex relationship between consumer performance and plant fitness (Erb 2018), likely lead to highly variable selection pressures on plant chemical traits in a complex and continuously changing biotic environment.

The multifarious concept of phytochemical diversity can encompass the complexity of phytochemical composition and its spatiotemporal variation (Wetzel & Whitehead 2020). Here, we focused on two dimensions of phytochemical diversity—richness and structural diversity. We also tested how the evenness of compounds affected the performance of the four insects (Appendix S1c). Similar to the patterns for richness, increasing evenness had no effect on the performance of individual herbivores, but increased the total number of herbivores affected (Figure S5-S7; Tables S12-S14). Taken together, our results suggest that the interaction diversity hypothesis can provide an adaptive explanation for

multiple dimensions of phytochemical diversity, including richness, evenness, and structural diversity. Future work should continue to combine observational and comparative approaches with experimentation to further disentangle how the different components of phytochemical diversity, including spatiotemporal variation, provide adaptive benefits to plants.

Conclusion

What is the raison d'être of secondary metabolite diversity? We show that multiple components of phytochemical diversity, including richness, evenness, and structural complexity, can benefit plants in multispecies interactions by increasing the diversity of potential bioactivities against different antagonists. Every phenolic metabolite we tested showed bioactivity in interactions with one or more consumers, but bioactivity was consumer-specific and no compounds were effective against all consumers. These experimental results support the intuition of many ecologists. Plants are embedded in complex biotic communities in which they must simultaneously deter multiple enemies and attract multiple mutualists. It is the selective landscape exerted by this interaction diversity that likely maintains the rich array of metabolites that fundamentally shape the life of a plant.

Acknowledgements

We are grateful to Lindsay Fennell, Angela Renhui Lu, Anna Wilson, Molly Atkinson, Jacob Turner, and Grace Florjancic for assistance with data collection. Anton Baudoin provided fungal cultures and advice on the design of fungal bioassays. Helpful comments on the manuscript were provided by Anurag Agrawal and members of the Agrawal Lab. Funding was provided by start-up funds to K.P. (Cornell University) and S.R.W. (Virginia Tech), an Agriculture and Food Research Initiative Competitive Grant no.

2018-07366 from the USDA National Institute of Food and Agriculture to S.R.W., and National Science Foundation Grant no. 1856776 to S.R.W.

References

Adler, L.S. (2000). The ecological significance of toxic nectar. *Oikos*, 91, 409–420.

Appel, H.M. (1993). Phenolics in ecological interactions: the importance of oxidation. *J. Chem. Ecol.*, 19, 1521–1552.

Bartoń, K. (2019). MuMIn: multi-model inference. R package version 1.43.15.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., *et al.* (2020). lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-23.

Berenbaum, M. & Neal, J.J. (1985). Synergism between myristicin and xanthotoxin, a naturally cooccurring plant toxicant. *J. Chem. Ecol.*, 11, 1349–1358.

Berenbaum, M. & Zangerl, A. (1993). Furanocoumarin metabolism in *Papilio polyxenes*: biochemistry, genetic variability, and ecological significance. *Oecologia*, 95, 370–375.

Berenbaum, M.R. & Zangerl, A.R. (1996). Phytochemical diversity: adaptation or random variation? In: *Phytochemical Diversity and Redundancy in Ecological Interactions*, Recent Advances in Phytochemistry (eds. Romeo, J.T., Saunders, J.A. & Barbosa, P.). Springer US, Boston, MA, pp. 1–24.

Bustos-Segura, C., Poelman, E.H., Reichelt, M., Gershenzon, J. & Gols, R. (2017). Intraspecific chemical diversity among neighbouring plants correlates positively with plant size and herbivore load but negatively with herbivore damage. *Ecol. Lett.*, 20, 87–97.

Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T. & Girke, T. (2008). ChemmineR: a compound mining framework for R. *Bioinformatics*, 24, 1733–1734.

Choi, J., Summers, W. & Paszkowski, U. (2018). Mechanisms underlying establishment of arbuscular mycorrhizal symbioses. *Annu. Rev. Phytopathol.*, 56, 135–160.

Cipollini, M.L. & Levey, D.J. (1997). Secondary metabolites of fleshy vertebrate-dispersed fruits: adaptive hypotheses and implications for seed dispersal. *Am. Nat.*, 150, 346–372.

Diawara, M.M., Trumble, J.T., White, K.K., Carson, W.G. & Martinez, L.A. (1993). Toxicity of linear furanocoumarins to *Spodoptera exigua*: evidence for antagonistic interactions. *J. Chem. Ecol.*, 19, 2473–2484.

Erb, M. (2018). Plant defenses against herbivory: closing the fitness gap. *Trends Plant Sci.*, 23, 187–194.

Firn, R.D. & Jones, C.G. (2003). Natural products – a simple model to explain chemical diversity. *Nat. Prod. Rep.*, 20, 382–391.

Gershenzon, J., Fontana, A., Burow, M., Wittstock, U. & Degenhardt, J. (2012). Mixtures of plant secondary metabolites: metabolic origins and ecological benefits. *Ecol. Plant Second. Metab. Genes Glob. Process.*, 56.

Hefner, T., Arend, J., Warzecha, H., Siems, K. & Stöckigt, J. (2002). Arbutin synthase, a novel member of the NRD1β glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg. Med. Chem.*, 10, 1731–1741.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A. & Scheibe, S. (2020). Multcomp: simultaneous inference in general parametric models. R package version 1.4-13.

Iason, G.R., Dicke, M. & Hartley, S.E. (2012). *The Ecology of Plant Secondary Metabolites: From Genes to Global Processes*. Cambridge University Press.

Iason, G.R., O'Reilly-Wapstra, J.M., Brewer, M.J., Summers, R.W. & Moore, B.D. (2011). Do multiple herbivores maintain chemical diversity of Scots pine monoterpenes? *Philos. Trans. R. Soc. B Biol. Sci.*, 366, 1337–1345.

Jones, C.G., Firn, R.D., Malcolm, S.B., Chaloner, W.G., Harper, J.L. & Lawton, J.H. (1991). On the evolution

of plant secondary chemical diversity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 333, 273–280.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., *et al.* (2016). PubChem substance and compound databases. *Nucleic Acids Res.*, 44, D1202–D1213.

Lattanzio, V., Kroon, P., Quideau, S. & Treutter, D. (2009). Plant phenolics – secondary metabolites with diverse functions. In: *Recent Advances in Polyphenol Research*. pp. 1–35.

Liu, X., Vrieling, K. & Klinkhamer, P.G.L. (2017). Interactions between plant metabolites affect herbivores: a study with pyrrolizidine alkaloids and chlorogenic acid. *Front. Plant Sci.*, 8.

Liu, Y. & Zhao, H. (2016). Predicting synergistic effects between compounds through their structural similarity and effects on transcriptomes. *Bioinformatics*, 32, 3782–3789.

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

Nevo, O. & Valenta, K. (2018). The ecology and evolution of fruit odor: implications for primate seed dispersal. *Int. J. Primatol.*, 39, 338–355.

Palmer-Young, E.C., Sadd, B.M. & Adler, L.S. (2017). Evolution of resistance to single and combined floral phytochemicals by a bumble bee parasite. *J. Evol. Biol.*, 30, 300–312.

Pigliucci, M. & Preston, K. (2004). *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*. Oxford University Press.

Richards, L.A., Dyer, L.A., Forister, M.L., Smilanich, A.M., Dodson, C.D., Leonard, M.D., *et al.* (2015). Phytochemical diversity drives plant–insect community diversity. *Proc. Natl. Acad. Sci.*, 112, 10973–10978.

Richards, L.A., Glassmire, A.E., Ochsenrider, K.M., Smilanich, A.M., Dodson, C.D., Jeffrey, C.S., *et al.* (2016). Phytochemical diversity and synergistic effects on herbivores. *Phytochem. Rev.*, 15, 1153–1166.

Ripley, B., Venables, B., Bates, D., M., Hornik, K., Gebhardt, A. & Firth, D. (2020). Support functions and

datasets for Venables and Ripley's MASS. R package version 7.3-51.6.

Salazar, D., Jaramillo, A. & Marquis, R.J. (2016). The impact of plant chemical diversity on plant–herbivore interactions at the community level. *Oecologia*, 181, 1199–1208.

Salazar, D., Lokvam, J., Mesones, I., Vásquez Pilco, M., Ayarza Zuñiga, J.M., de Valpine, P., *et al.* (2018). Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.*, 2, 983–990.

Schiestl, F.P. (2015). Ecology and evolution of floral volatile-mediated information transfer in plants. *New Phytol.*, 206, 571–577.

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

Stamp, N.E. & Osier, T.L. (1998). Response of five insect herbivores to multiple allelochemicals under fluctuating temperatures. *Entomol. Exp. Appl.*, 88, 81–96.

Steppuhn, A. & Baldwin, I.T. (2007). Resistance management in a native plant: nicotine prevents herbivores from compensating for plant protease inhibitors. *Ecol. Lett.*, 10, 499–511.

Tallarida, R.J. (2000). *Drug synergism and dose-effect data analysis*. CRC Press.

Thompson, J.E., Walker, R.P. & Faulkner, D.J. (1985). Screening and bioassays for biologically-active substances from forty marine sponge species from San Diego, California, USA. *Mar. Biol.*, 88, 11–21.

Tilman, D., Reich, P.B., Knops, J., Wedin, D., Mielke, T. & Lehman, C. (2001). Diversity and productivity in a long-term grassland experiment. *Science*, 294, 843–845.

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. *Ecol. Lett.*, 21, 83–92.

Wetzel, W.C. & Whitehead, S.R. (2020). The many dimensions of phytochemical diversity: linking theory

to practice. *Ecol. Lett.*, 23, 16–32.

Whitehead, S.R. & Bowers, M.D. (2014). Chemical ecology of fruit defence: synergistic and antagonistic interactions among amides from Piper. *Funct. Ecol.*, 28, 1094–1106.

Whitehead, S.R., Jeffrey, C.S., Leonard, M.D., Dodson, C.D., Dyer, L.A. & Bowers, M.D. (2013). Patterns of secondary metabolite allocation to fruits and seeds in *Piper reticulatum*. *J. Chem. Ecol.*, 39, 1373–1384.

Whitehead, S.R. & Poveda, K. (2019). Resource allocation trade-offs and the loss of chemical defences during apple domestication. *Ann. Bot.*, 123, 1029–1041.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., *et al.* (2020). ggplot2: create elegant data visualisations using the grammar of graphics. R package version 3.3.2.

Wu, K., Zhang, J., Zhang, Q., Zhu, S., Shao, Q., Clark, K.D., *et al.* (2015). Plant phenolics are detoxified by prophenoloxidase in the insect gut. *Sci. Rep.*, 5, 16823.

Figure Legends

Figure 1: Phenolic richness has no effect on herbivore performance (A-C) or fungal growth (D). Response variables are standardized performance metrics; values less than one indicate that performance decreased on phenolic-supplemented diet relative to controls and vice versa. For pupal mass (A) and development speed (B), each point represents an individual insect (N = 2164; see Table S2). For survival (C), each point represents the proportion survival across all insects reared on a particular diet composition (N = 236 [59 compositions x 4 species] based on 2753 individuals; see Table S2). For fungal growth (D), each point represents the fungal growth in a single replicate well (N = 944; 4 wells per species per composition). See Table S1 for a description of compound compositions at each richness level and Table S3 for statistical results.

Figure 2: Phenolic structural diversity has no effect on herbivore performance (A-C) or fungal growth (D). Response variables are standardized performance metrics; values less than one indicate that performance decreased on phenolic-supplemented diet relative to controls and vice versa. For pupal mass (A) and development speed (B), each point represents an individual insect (N = 2164; see Table S2). For survival (C), each point represents the proportion of survivors across all insects reared on a particular diet composition (N = 236 [59 compositions x 4 species] based on 2753 individuals; see Table S2). For fungal growth (D), each point represents the fungal growth in a single replicate well (N = 944; 4 wells per species per composition). See Table S1 for a description of compound compositions at each richness level and Table S3 for statistical results.

Figure 3: The number of consumers (out of 8 possible insect and fungal species) that were negatively affected by phenolics increased with increasing richness (A) and structural diversity (B) of phenolic mixtures. Each point represents a unique diet composition (N = 59; Table S1). A consumer was considered negatively affected if there was statistical support (at 95% confidence) that any aspect of

performance (i.e. survival, pupal mass, and/or development speed for insects or growth rate for fungi) was reduced on phenolic-containing diets compared to controls.

Figure 4: Average performance of four insect species (Cp = *Cydia pomonella*, Hz = *Helicoverpa zea*, Px = *Plutella xylostella*, and Sf = *Spodoptera frugiperda*) and four fungal taxa (Bd = *Botryosphaeria dothidea,* C = *Colletotrichum* sp.*,* Pe = *Penicillium expansum,* and Ss = *Sclerotinia sclerotiorum*) on growth media containing one of 14 individual phenolic compounds. Performance metrics are standardized relative to controls, such that values less than one indicate a negative effect of phenolics on performance and vice versa. Stars inside squares indicate statistical support for a difference in performance between the compound and controls (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). The dendrogram indicates structural relatedness of compounds based on an average of three structural similarity metrics (atom pairs, maximum common substructure, and PubChem fingerprints).

Table 1: Summary of predictions derived from three hypotheses that explain the maintenance of phytochemical diversity

Hypotheses	Predictions	Supported?	Evidence
I. Synergy hypothesis	**a.** consumer performance decreases with increasing richness	NO	Figs. 1, S1; Tables S3-S5
	b. consumer performance decreases with increasing structural diversity	NO	Figs. 2, S1; Tables S3-S5
	c. the effects of richness on consumer performance are more pronounced at high levels of structural diversity	NO	Fig. S1; Tables S3-S5
	d. the occurrence and strength of synergisms in mixtures increases with increasing richness	NO	Fig. S2; Table S6
	e. the probability of a mixture's bioactivity exceeding that of its average component single compound increases with increasing richness	NO	Fig. S3; Table S7
II. Interaction diversity hypothesis	**a.** number of consumers affected increases with increasing richness	YES	Fig. 3
	b. number of consumers affected increases with increasing structural diversity	YES	Fig. 3
	c. most or all compounds are effective against one or more consumers	YES	Fig. 4
	d. few or no compounds are effective against all consumers (no silver bullets)	YES	Fig. 4
	e. effect of compound(s) depends on consumer identity	YES	Fig. S4; Table S8
III. Screening hypothesis	**a.** most compounds have no apparent effects on the performance of any consumer	NO	Fig. 4
	b. one or a few compounds are sufficient to explain variation in consumer performance	NO	Table S9
	c. the specific mixture of compounds in diet is not a strong predictor of performance	NO	Fig. S4; Table S10

(A)

Pupal Mass

1.5
1.0
0.5

1 2 4 6 8 10

(B)

Development Speed

1.4
1.0
0.6

1 2 4 6 8 10

(C)

Survival

1.6
1.2
0.8
0.4

1 2 4 6 8 10

Richness

(D)

Growth Rate

1.3
1.1
0.9
0.7

1 2 4 6 8 10

Richness

Insect Species (Plots A, B, C)

C.pomonella *H. zea* *P. xylostella* *S. frugiperda*

Fungal Species (Plot D)

B. dothidea *Colletotrichum* *P. expansum* *S. sclerotiorum*

(A)

Pupal Mass

1.5
1.0
0.5

low med high

(B)

Development Speed

1.6
1.2
0.8
0.4

low med high

(C)

Survival

1.6
1.2
0.8
0.4

low med high

Structural Diversity

(D)

Growth Rate

1.3
1.1
0.9
0.7

low med high

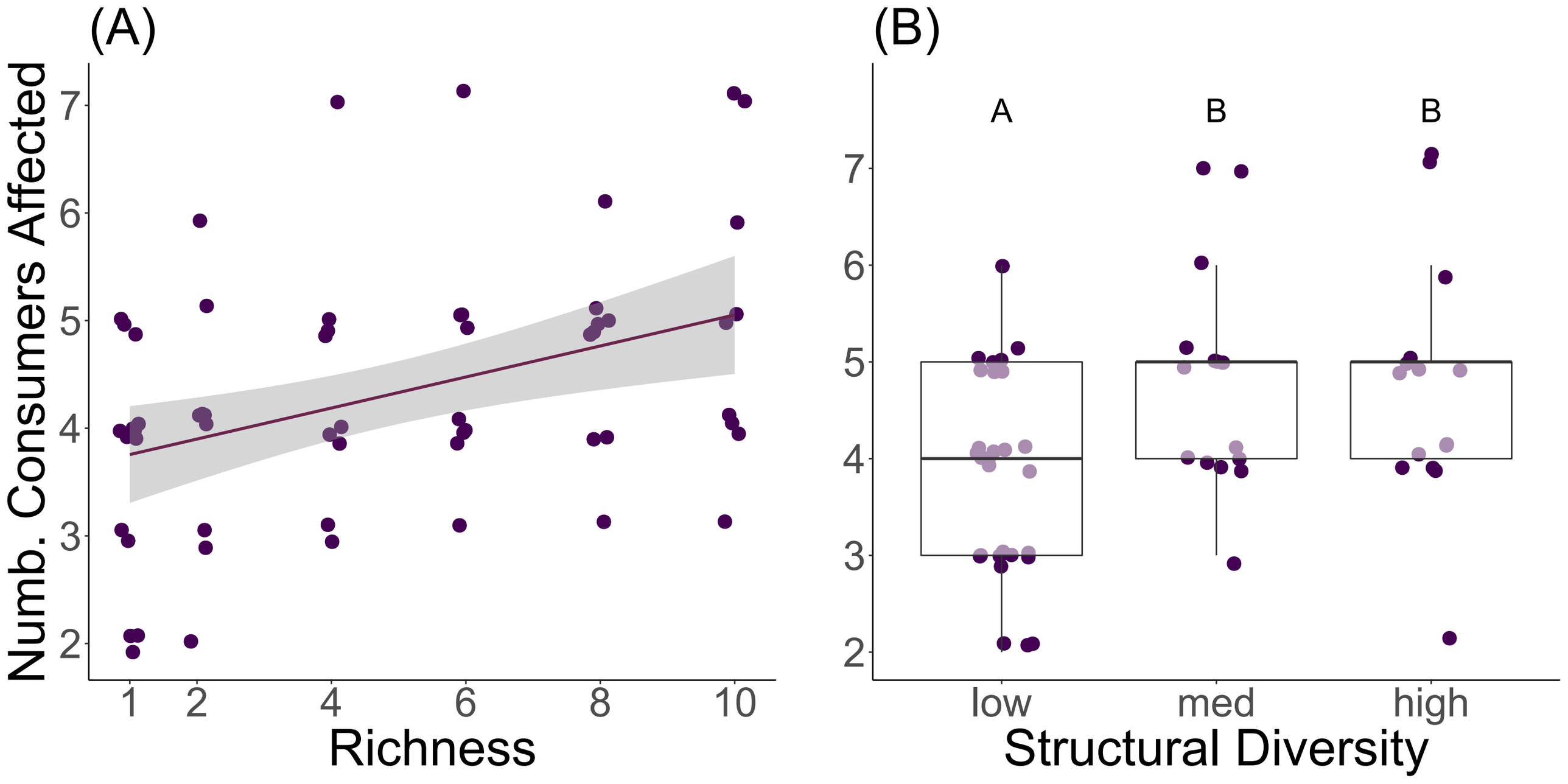
Structural Diversity

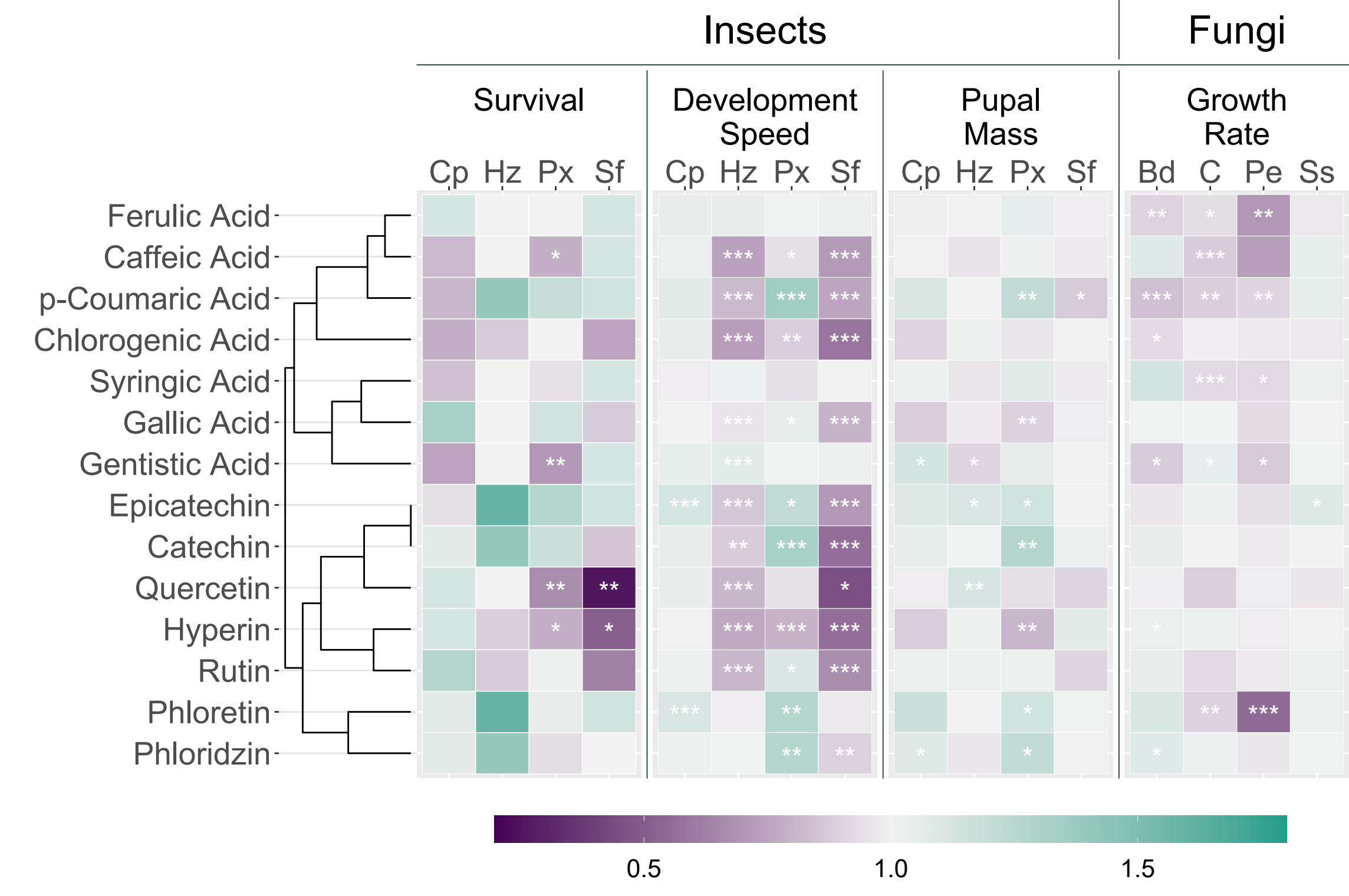
Insect Species (Plots A, B, C)

C.pomonella *H. zea* *P. xylostella* *S. frugiperda*

Fungal Species (Plot D)

B. dothidea *Colletotrichum* *P. expansum* *S. sclerotiorum*


(A)
Numb. Consumers Affected
7
6
5
4
3
2
1
2
4
6
8
10
Richness
(B)
A
B
B
7
6
5
4
3
2
low
med
high
Structural Diversity


Insects
Fungi
Survival
Development Speed
Pupal Mass
Growth Rate
Cp Hz Px Sf
Cp Hz Px Sf
Cp Hz Px Sf
Bd C Pe Ss
Ferulic Acid
Caffeic Acid
p-Coumaric Acid
Chlorogenic Acid
Syringic Acid
Gallic Acid
Gentistic Acid
Epicatechin
Catechin
Quercetin
Hyperin
Rutin
Phloretin
Phloridzin
0.5
1.0
1.5

APPENDIX S1: Supporting information to accompany: "Interaction diversity explains the maintenance of phytochemical diversity"

Susan R. Whitehead, Ethan Bass, Alexsandra Corrigan, André Kessler, and Katja Poveda

Contents

Appendix S1a: Experimental Methods

Selection of compound mixtures

To select the compound mixtures, we first generated a list of all possible mixtures of 2, 4, 6, 8, and 10 compounds out of our pool of 14 total compounds. For each possible mixture, we calculated the structural diversity based on three existing metrics that provide a measure of the structural similarity between any two known structures: atom pairs, PubChem fingerprints, and maximum common substructure (Cao et al. 2008; Wang et al. 2013). All of these metrics consider shared and unique molecular features between two compounds and can be summarized by the Tanimoto coefficient. For any two compounds *A* and *B*, the Tanimoto coefficient takes the form $\frac{c}{a+b+c}$, where *a* is the number of unique features in compound *A*, *b* is the number of unique features in compound *B*, and *c* is the number of shared features. Thus, Tanimoto values are between 0 and 1, with 0 indicating no shared features and 1 indicating complete overlap. For each pairwise combination of compounds, we took the average Tanimoto coefficient calculated across the three metrics of structural similarity. Then, to assess the structural diversity of any complex mixture, we used the average Tanimoto coefficient across all pairwise comparisons of individual compounds in a mixture. Note that this averaging means that coefficients are only comparable within a certain level of richness, as the range of possible values decreases with increasing richness. Thus, once this metric was calculated for all possible mixtures of 2, 4, 6, 8, and 10 compounds, we designated the highest 10% of values within a richness level as "low" structural diversity, the middle 10% as "medium", and the lowest 10% as "high".

From these subsets of high, medium, and low structural diversity mixtures, we then randomly selected three combinations for each richness level. For each set of three, we also used two additional criteria to minimize the chance that our study would be biased by the bioactivity of any one compound: 1) the maximum number of times any one compound could occur across a set of three mixtures was set at 1, 2, 2, 2, and 3 for mixtures at richness levels 2, 4, 6, 8, and 10, respectively, and 2) the maximum number of compounds that could be shared between any two mixtures in a set of three was 0, 2, 4, 5, and 7 for mixtures at richness levels of 2, 4, 6, 8, and 10, respectively. Thus, to generate, for example, a set of

three mixtures with low structural diversity and a richness level of four, we randomly drew sets of three from the 4-compound, low-diversity subset until we achieved a set that met the two additional selection criteria. The final details of the compound compositions for each mixture are provided in Table S1.

The calculation of structural diversity and selection of mixtures was conducted in R v 3.6.1 using the packages ChemmineR and fmcsR (see R scripts at Zenodo DOI: 10.5281/zenodo.4586759).

Selection of compound concentrations

We maintained the total concentration of phenolics in the growth media constant at 0.2% fresh mass for all insect trials and 0.1% for fungi. We chose these concentrations to mimic the total phenolic fresh weight concentration in ripe apple fruits (mean: 0.2%, range: 0.06% - 0.43% across 31 cultivars; Neveu et al. 2010). For organisms in our study that commonly consume apple fruits (*Cydia pomonella* and all four fungal strains), this concentration would be a conservative estimate of what is typically experienced in their diet, because higher concentrations are found in unripe apple fruits and in wild apples (Whitehead and Poveda 2019).

Concentrations of individual compounds in the diet mixtures were included in equal mass concentration. We focused on mass concentration over molar concentration for several reasons. First, the hypotheses tested in this study do not focus on specific chemical reactions between two compounds, for which equimolar concentrations would typically lead to maximum bioactivity. Instead, they focus on the broad potential for functional diversity in a mixture, which could include multiple modes of bioactivity and molecular interactions for any given phenolic compound (Lattanzio *et al.* 2009; Shahidi & Yeo 2018). In general, larger, more structurally complex phenolic metabolites may have more potential active sites (e.g. multiple hydroxyl groups) and complex metabolic fates in the consumer, and therefore may have a higher total functional potential (Xiao *et al.* 2013, 2015; Chen *et al.* 2018). A second reason we chose to use mass-based calculations is that a mass-based concentration of a compound may be more representative of the total investment made by a plant in its biosynthesis, both in terms of raw materials and the enzymes and other biosynthetic machinery involved in its production. This may be especially true of phenolics, which are often constructed from repeated addition of core units and skeletal modification of phenolic ring systems (Gershenzon *et al.* 2012). A final reason is that we wanted to focus on concentration as it is generally reported and understood in the chemical ecology literature—on a mass basis. We modeled our concentrations off of mass-based values (0.2% fresh mass total phenolics) reported in the literature and kept that value constant even in single compound trials. Evenness has also been typically calculated in the literature based on mass concentrations (though certainly calculating evenness based on molarity or bioactivity are important alternatives that should be explored further; Wetzel and Whitehead 2020).

The above reasons led us to choose mass-based over molar-based calculations, but this choice does also come with several caveats to the interpretation of results. First, when directly comparing the bioactivity of individual compounds, it is important to consider that they were tested in unequal molar concentrations. Compounds used in this study vary in molecular weight from 154.12 g/mol (gentisic acid) to 610.5 g/ml (rutin) (Table S1); thus there are nearly 4x as many molecules of gentisic acid as compared to rutin when present in equal mass concentration. One might expect that this would lead to increased apparent bioactivity of the smaller molecules, but we did not find that molecular weight was a

major predictor of bioactivity in our study (see R scripts on Github: WhiteheadLabVT/Phytochemical-Diversity-Experiment). In fact, the only case we found of a relationship between molecular weight and bioactivity was a negative relationship between the molecular weight of a compound and the survival of *Spodoptera frugiperda* on diets containing that compound (P = 0.003). This generally supports our argument above that mass concentration may, at least in some cases, be a better predictor of total functional potential than molar concentration, but still it is critical to interpret the comparative bioactivity of the compounds through the lens of mass concentration.

A second important consequence of our choice of using mass concentrations is that it could have reduced the total potential for non-additive interactions to occur. For any given synergy or antagonism that relies on direct interactions between two compounds, equimolar concentrations would likely yield the maximum bioactivity. We do not think that this would have biased our main conclusions regarding the potential for synergies to accumulate with increasing chemical diversity, which focused on comparing the relative bioactivities of mixtures composed in a random fashion without regard to molecular weight. However, it may have reduced the total likelihood of detecting non-additive compound interactions across the entire study. Overall, our results should not be interpreted as clear evidence (or lack thereof) for any specific non-additive interactions, which would ideally be examined by testing compounds individually and in mixtures across a range of concentrations (Tallarida 2000).

Insect bioassays

Eggs of *H. zea*, *S. frugiperda*, and *C. pomonella* were obtained from Benzon Research (Carlisle, PA, USA), and eggs of *P. xylostella* were obtained from a colony maintained in the laboratory of Anthony Shelton (Cornell University). All caterpillars were reared on a multispecies artificial diet in a walk-in growth chamber maintained at 20°C, 50% relative humidity, and a 12:12 light:dark cycle. The pH of the midgut of these or closely related herbivore species is generally basic and has been reported in the literature as *H. zea* (8.0), *Spodoptera litura* (8.2-8.5), *Cydia molesta* (8.5-8.7), *Plutella xylostella* (7.4-8.29) (Berenbaum 1980).

To conduct the trials, we prepared growth media using a multispecies insect diet (Southland Products Incorporated, Arkansas, USA). The diet consists of soyflour, wheat germ, mineral mix, sugar, aureomycin, vitamins, agar, methyl paraben, sorbic acid, and calcium propionate and has a final pH of 4.6. A complete nutritional profile is available in Appendix S2; Table S15. To prepare diets, multispecies insect diet was mixed with boiling water and blended using an immersion blender according to manufacturer's directions in batches of ~150mL (24.3g of diet mix and 138.7 g water [added by mass using a balance]) and equilibrated to a temperature of 60°C in a water bath. Concentrated stock solutions were prepared for each phenolic compound in DMSO (440.5 mg/ml) and a total of 744µL [=0.819g] of these solutions (evenly divided among the number of compounds in the mixture; e.g. 93µL from each stock solution for an 8 compound mixture) were then added to the diet to achieve each of the randomly selected compound compositions (Table S1). This gave a final total concentration of ~0.2% fresh diet mass for phenolics and ~0.5% fresh diet mass for DMSO (138.8g of water, 24.3g of diet mix, 0.8g DMSO and 0.3g total phenolics). This concentration of DMSO in the diets allowed full dissolution of all compounds in stock solutions and, according to our preliminary experiments and past studies (e.g. Zeng et al. 2009), has minimal effects on insect performance. Once the compounds were added, the diet was re-mixed with an immersion blender and transferred with a 25mL syringe to 1oz plastic deli cups.

We prepared approximately 22 cups of diet from each 150mL batch. These included eight cups each for *S. frugiperda* and *H. zea* (6mL per cup) and three cups each for *C. pomonella* and *P. xylostella* (5mL per cup). More cups were prepared for the two generalists because each individual caterpillar needed to be reared in a separate cup (to avoid cannibalism and ensure enough diet for these larger caterpillars), whereas the two specialists can be easily reared with multiple individuals in one cup. We prepared control diets (0.5% DMSO) in the same manner, adding DMSO only. The trials were conducted in three experimental rounds with different diet treatments in each. The order in which different diet treatments were prepared and included in rounds was randomized. On a few occasions, we repeated a particular treatment in a second round due to low survival rates of larvae in the first round that limited our ability to assess pupal weights and days to pupation. Deleting these observations (107 out of 2756 total individual insects) did not have any qualitative effect on the outcome of our analyses, so all final results are based on the full dataset.

For *H. zea* and *S. frugiperda*, we added one newly-hatched neonate larvae to each of eight cups. Larvae were carefully transferred from egg sheets using a paintbrush within 24-hours of hatching. Forty-eight hours after the transfers, we checked all cups, removed any dead larvae, and replaced them with a new set of newly-hatched neonates. For *C. pomonella*, we added three or four newly-hatched neonates to each of three cups, but did not replace any dead larvae. For *P. xylostella*, due to very low survival of newly-hatched larvae that were handled with a brush, we placed egg sheets over a larger deli-cup with diet and allowed newly-hatched larvae to feed for approximately 48 hours on standard diet after hatching before transferring to the experimental diets with a paintbrush. Approximately eight two-day old larvae were transferred to each of three experimental diet cups. In total, between 4 and 43 individual caterpillars per species were reared on each unique treatment, for a total of 2,756 individual caterpillars. Sample sizes (number of individual caterpillars) for each species/treatment combination are provided in Table S2. These numbers include only larvae that began feeding and established on the diet (i.e. we did not count neonates that did not survive the transfer to diet as “non-survivors” in the experiment, only those that began feeding and then later died). Once caterpillars were added to cups, cups containing different diet treatments in a given experimental round were haphazardly mixed and added to trays (11" x 21" greenhouse seedling flats), with eight control cups added per tray. Cups for each species were maintained separately. Trays were transferred to a growth chamber, and we began monitoring larvae for survival and pupation at 13 days following transfer for *H. zea and S. frugiperda*, 20 days following transfer for *C. pomonella*, and 7 days following transfer for *P. xylostella* and continued monitoring daily until all larvae had pupated or died.

Fungal growth media preparation

Fungal cultures were obtained from a culture collection of taxa isolated from diseased apples and maintained in the laboratory of Dr. Anton Baudoin (Virginia Tech). These were tentatively identified based on morphology and disease symptoms as *Botryosphaeria dothidea* (Botryosphaeriaceae; white rot)*, Colletotrichum* sp. (Glomerellaceae; bitter rot), *Penicillium expansum* (Trichocomaceae; blue mold)*,* and *Sclerotinia sclerotiorum* (Sclerotiniaceae; calyx end rot). These identifications should be considered tentative, as many fruit rots and diseases have multiple causal agents. Fungi were maintained over the course of the experiment in mycological broth (M264; Thomas Scientific) in an incubator at 30^{o}C, transferring stock cultures weekly to fresh broth using sterile technique.

For fungal trials, we prepared sterile broth (HiMedia mycological broth M264, Thomas Scientific). The broth consists of papaic digest of soybean meal (10g/L) and dextrose (40g/L) and has a final pH of 7.0 (+/- 0.2). Broth was prepared in an autoclave in batches of 500mL according to the manufacturer's directions. Concentrated stock solutions were prepared for each individual compound in DMSO (28.215 mg/ml [=2.5% w/w]) and a total of 8.87 µL of these solutions (evenly divided among the number of compounds in the mixture) was added to 96-well plates to achieve the randomly selected compositional mixtures (Table S1). The DMSO was evaporated to dryness in a vacuum chamber, leaving a total mass of 0.2503mg of phenolics per well. To each well, we then added 1.14 µL [= 1.25mg] DMSO, vortexed, then added 239 µL of fresh mycological broth and 10 µL of broth containing homogenized mycelia from a stock fungal culture (see above) to achieve a final total concentration 0.1% of the broth mass for phenolics and 0.5% for DMSO [249 mg broth, 0.25 mg phenolics, 1.25mg DMSO). Stock cultures in 50mL culture tubes were homogenized prior to the transfer of aliquots using a FastPrep-24 (MP Biomedicals). Each assay was replicated in four wells per plate. Each plate also contained four replicate negative control wells with 0.5% DMSO only (no phenolic metabolites) as well as four positive controls that contained the fungicide Captan at 0.1%. Once the plate was prepared, we took an initial reading of the absorbance for each well using a plate reader (BioTek Synergy HT) at 600nm. Additional readings were taken every 22-26 hours for an additional 72 hours, giving four readings per plate. We used the slope of the relationship between time and absorbance as a measure of fungal growth.

Appendix S1b: Statistical methods

Statistical analyses: synergy hypothesis

Prediction 1a-1c. To test whether consumer performance was affected by the richness and structural diversity of chemical compounds in their diet, we used linear mixed models (LMMs) and a multimodel inference approach to assess support for predictor variables (Grueber et al. 2011). These models were conducted separately for herbivores and for fungi and for each performance metric (survival, pupal weights, and development speed for herbivores and growth rate for fungi). For standardized herbivore pupal weights and development speed, the global LMM included richness, structural diversity, herbivore species, sex, and all possible interactions as fixed effects. Random effects were: 1) the compound mixture (the identity of the randomly generated mixture within each level of richness x structural diversity) and 2) the cup identity (when multiple insects were reared in a single cup) nested within the tray identity nested within the experiment number where the herbivore was tested. For standardized survival data, we used a similar approach but with simplified models. Sex was not included because we did not have sex information for larvae that did not survive to pupation, and cup identity and tray identity were not included as random effects because we summarized survival data across all caterpillars in a treatment. For standardized fungal growth, we used a global model that contained richness, structural diversity, species, and all possible interactions as fixed effects, and the compound mixture and plate identity as random effects. Once these global models were constructed, we used the dredge function in the R package "MuMin" (Bartoń 2019) to examine all possible subsets of the full model. Models that were within $\Delta AIC < 4$ of the top model were averaged to obtain parameter estimates and assess support for individual predictor variables.

Based on numerous strongly supported interactions involving species and sex in the multimodel averaging (see results), we followed these analyses with a series of linear mixed models (LMMs) conducted separately for each of the eight taxa. For herbivores, separate analyses were also conducted for each performance metric (survival, pupal mass, and development speed) and for males and females (for pupal mass and development speed only; sex could not be determined when larvae died prior to pupation). In all models, we included richness, structural diversity, and their interaction as fixed effects and a random effects structure as described above. Support for richness and structural diversity as predictors of consumer performance was assessed using likelihood ratio tests comparing models to simplified versions with single terms deleted.

Prediction 1d: To test whether synergistic effects of compounds in mixtures increased with increasing phytochemical richness, we calculated an interaction index for each diet mixture and used linear models to examine how that index changed with increasing richness. These analyses were conducted separately for two performance metrics (pupal weights and development speed) for each insect species and for growth rate for each fungal species (12 separate analyses; $N = 45$ mixtures for each). Our interaction index was modified from established methods from the pharmacology literature (Tallarida 2000) that compare the "observed" bioactivity of a compound mixture to its "expected" bioactivity based on additive interactions among compounds. Our metric of observed bioactivity was based on linear models comparing consumer performance on a particular compound mixture to that on comparable controls. Coefficients (slopes) and their 95% confidence intervals from these models were collated for each compound mixture as the "observed" effect sizes. To calculate the "expected" additive effects, we used the data from single compound trials and assumed that each compound would contribute a proportion of the total effect size according to its proportional abundance in the compound mixture (note that these methods assume a linear dose response curve for each compound, which we did not explicitly test in this study). As our interaction index, we used the difference between expected minus observed effects, such that positive values would indicate synergy and negative values indicate antagonism. If the synergy hypothesis explains the maintenance of phytochemical diversity, then values for this index should increase with increasing richness in the diet mixture (Prediction 1d).

To confirm that the intrinsic distribution of the derived interaction index did not lead to biased rejection/support of hypotheses, we also performed a set of null models using 1000 iterations of the complete analysis. Each iteration involved 12 separate tests as above (resulting in 12,000 total models), randomizing values for the response variable (pupal weight, days to pupation, or fungal growth) within a given species. At $\alpha = 0.05$, our overall measured Type I error rate (i.e. false discovery of a relationship between richness and the interaction index) in these models was similar to null expectations at 0.0395.

Prediction 1e: To test whether increasing compound richness in a mixture increased the probability that the mixture's bioactivity would exceed that of the average bioactivity of its singletons, we used generalized linear models with the binomial distribution. The response variable was a binary outcome of whether the average bioactivity of a mixture against a particular consumer was higher than the average bioactivity of all of its component single compounds. These analyses were conducted separately for three standardized performance metrics (pupal weights, development speed, and survival) for each herbivore species and for standardized growth rate for each fungal species (16 separate analyses). For

each diet mixture, we compared the mean standardized performance of the consumer species on that mixture to the mean standardized performance across all component compounds in that mixture. We assigned values of 0 if the average performance on the singletons was lower (i.e. the singletons had higher bioactivity), and a value of 1 if the average performance on the mixture was lower (i.e. the mixture had higher bioactivity). This binary outcome was used as the response variable in a generalized linear model with a binomial distribution and richness as the predictor.

To confirm that the intrinsic distribution of the derived binomial response variable did not lead to biased rejection/support of hypotheses in the subsequent linear models, we also performed a set of null models using 1000 iterations of the complete analysis. Each iteration involved 16 separate tests as above (resulting in 16,000 total models), randomizing values for the response variable (pupal weight, days to pupation, survival, or fungal growth) within a given species. At $\alpha = 0.05$, our overall measured Type I error rate in these models was 0.044, in line with null expectations.

Statistical analyses: interaction diversity hypothesis

Predictions 2a and 2b: To test whether the number of consumers affected by phenolics increased with increasing richness or structural diversity, we first determined the number of consumers (out of 8 taxa) that were negatively affected by each individual diet mixture ($N = 54$ mixtures). For herbivores, a species was considered negatively affected if there was statistical support (at 95% confidence) that any single performance metric was affected. For pupal weights and development speed, this meant that the 95% confidence intervals for the standardized performance metrics did not cross one. For survival, this meant that there was a difference in survival (at $\alpha = 0.05$) between control and treatment diets based on Fisher's exact tests of the counts of dead and alive insects. For fungi, a strain was considered negatively affected if the 95% confidence intervals for the standardized performance metrics did not cross one. Once we obtained a count of the number of consumers affected for each unique diet mixture, we used this as a response variable in linear models to assess whether the number of consumers negatively affected by phenolics increases with increasing richness (Prediction 2a) or structural diversity (Prediction 2b). Based on support for an overall effect of structural diversity, we followed this analysis with a Tukey HSD multiple comparisons test among low, medium, and high diversity groups.

Our approach of counting an insect species as negatively affected by a treatment if any of the three performance metrics was reduced may have increased the likelihood that negative effects of insects would be detected relative to fungi, for which we only measured one performance metric. We used this approach because many treatments affected just one aspect of insect performance, so counting any of these increased the number "affected consumers" across all treatments and therefore our power to detect effects of richness and structural diversity. Still, we were also able to detect an effect of richness ($P = 0.01$) and a marginal effect of structural diversity ($P = 0.076$) on the number of consumers affected if we based our counts of affected insect species on just one performance metric—days to pupation—that was the most commonly reduced by phenolics. Thus, we do not think that our approach led to bias in our conclusions regarding the significance of phytochemical diversity, but results should not be interpreted in the context of comparing the likelihood of negative effects of specific treatments on insects versus fungi.

To further confirm that our analytical approach did not lead to biased rejection/support of hypotheses, we also performed a set of null models using 10,000 iterations of randomized data. For each iteration, values for the standardized response variables were randomized across treatments within a given species. We then calculated the number of organisms affected as above and used this metric as a response variable in subsequent linear models with richness and structural diversity as predictors. At $\alpha = 0.05$, our overall measured Type I error rate (i.e. false discovery of a relationship between chemical diversity and the number of consumers affected) was 0.048 for richness and 0.059 for structural diversity. Out of 10,000 null iterations, there were no cases for richness and only four cases for structural diversity where the effect sizes reached the magnitude estimated by our data.

Predictions 2c and 2d: To examine whether most compounds were biologically active against one or more herbivores (Prediction 2c) and, simultaneously, whether few or no compounds were active against all consumers (Prediction 2d), we tested the effects of each of the 14 individual compounds on each performance metric and species. For insect pupal mass, insect development speed, and fungal growth rate, we considered there to be support for bioactivity if the 95% confidence intervals for the standardized performance metrics crossed one. For insect survival, we used Fisher's exact tests (at $\alpha = 0.05$) of the counts of dead and alive insects on control and treatment diets.

Prediction 2e: To test whether the effects of phenolics depend on consumer identity, we used linear mixed models with species, treatment (the specific diet mixture), and their interaction as fixed effects. These models were conducted separately for each standardized performance metric (survival, pupal mass, and development speed for insects and growth rate for fungi). For pupal mass and development speed, random effects included the tray identity nested within the experiment number in which the herbivore was tested. For survival, we used the random effect of experiment number only. For fungal growth rate, we used the random effect of plate identity. A significant interaction between species and treatment would indicate that the effects of a compound mixture depend on the consumer identity.

Statistical analyses: screening hypothesis

Prediction 3a: To assess support for the prediction that most compounds have no apparent biological activity, we examined results from analyses of the effects of individual compounds on consumer performance (described above for Prediction 2c, which predicted the opposite based on the interaction diversity hypothesis).

Prediction 3b: To test whether variation in consumer performance could be explained by one or a few compounds, we began by constructing linear models with the following predictors: the presence/absence of each individual compound, richness, structural diversity, and consumer species. These models were constructed separately for each standardized performance metric for insects (pupal mass, development speed, and survival) and fungi (growth rate). We then used stepwise (both backward and forward) AIC-based model selection in the R package 'MASS' (Ripley et al. 2019) to determine the minimum adequate model to describe consumer performance. If variation in performance can be explained by one or a few compounds, then most individual compounds should not be retained in these models.

Prediction 3c: To test whether the specific mixture of compounds in the diet explains a significant portion of the variance in herbivore performance, we used multilevel random intercepts models that partitioned variance in performance among the specific diet treatment, the experiment number, and residual variance. The models included diet treatment and experiment number as random effects and no fixed effects. These were conducted separately for each consumer species and performance metric. To examine the importance of treatment in explaining variation in performance, we compared models with and without treatment as a random effect using likelihood ratio tests and also calculated the variance partition coefficients (i.e. the proportion of total variance in performance explained by variation within treatments).

All analyses were conducted in R v. 3.6.1 using the packages lme4 (Bates et al. 2020), multcomp (Hothorn et al. 2020), MuMIn (Bartoń 2019), and MASS (Ripley et al. 2019), and figures were prepared using ggplot2 (Wickham et al. 2020). All data and associated R scripts are available on Github (WhiteheadLabVT/Phytochemical-Diversity-Experiment) and archived in Zenodo (DOI: 10.5281/zenodo.4586759).

Appendix S1c: Complementary experiment on compound evenness

Motivation

Although theory on phytochemical diversity generally discusses diversity as a uni-dimensional concept, diversity is inherently multi-dimensional and depends on the richness, relative abundances, and structural complexity of compounds and how they vary in space and time (Wetzel and Whitehead 2019). One of the least explored dimensions is evenness, which can be defined as the extent to which compounds in a mixture are present in equal abundances. Although little is known about the explicit role of phytochemical evenness in species interactions, many studies of phytochemical diversity include evenness implicitly by utilizing diversity metrics such as Simpson's or Shannon's to describe phytochemistry (e.g. Iason et al. 2005; Richards et al. 2015; Whitehead and Poveda 2019). Considering that different axes of phytochemical diversity are likely to co-vary in natural systems, and that each facet may have unique consequences for plant interactions, experimental approaches in model systems are critical to disentangle the relative contributions of these different facets of diversity to the adaptive nature of phytochemical phenotypes. Thus, we also conducted an additional complementary experiment to the one described in the main text in which we manipulated compound evenness in insect diets.

Methods

The evenness experiment included five different evenness levels (0.2, 0.4, 0.6. 0.8, and 1.0), where values of 1.0 are perfectly even and values closer to zero are highly uneven, crossed with three levels of structural diversity (high, medium, and low), for a total of 45 different diet treatments. Richness was held constant with six different compounds in all mixtures. The total concentration of phenolics was held constant at 0.2% of diet fresh mass as in the richness experiment (see main text). Between eight and 28 individual caterpillars from each species were reared on each diet treatment, for a total of 2340 experimental insects, plus 101 controls. Of these insects, 1436 experimental and 70 controls survived. This was a much lower overall survival rate compared to the richness experiments, which may have

been due to pathogen incidence or a difference in the starting populations obtained from commercial egg sources.

To select the diet mixtures for the evenness experiment, we first randomly selected compound identities for three replicate mixtures at each level of evenness and structural diversity (using methods described in Supplement 1 for the richness experiment). Next, to achieve different levels of evenness, we randomly assigned relative concentration values to a vector of six compounds by choosing six random values that total 100%, with a minimum relative concentration of 1%. The evenness of this mixture was then calculated using the Simpson's Equitability index, and the mixture composition was retained only if it was within +/- 0.01 of the particular evenness level desired (0.2, 0.4, 0.6, 0.8, 1.0). This process was repeated until we had three randomly chosen unique compound combinations for each level of evenness and structural diversity. The final details of the compound compositions for each diet treatment are provided in Table S11.

Insects were reared in conditions identical to the richness experiment, with control insects (diet with DMSO only) included on each tray in each experimental round, and we measured survival, development speed, and pupal mass as metrics of performance (see main text).

Statistical Analysis

To test whether herbivore performance decreases with decreasing evenness and structural diversity (as predicted by the synergy hypothesis), we analyzed data from the evenness experiment using linear mixed models and a model averaging approach identical to that described for the richness experiments (Predictions 1a-1c), but with evenness instead of richness as a predictor.

To test whether the number of herbivores affected by phenolics increased with increasing evenness (as predicted by the interaction diversity hypothesis), we first assessed the effects of each diet mixture in the evenness experiment on each herbivore species and performance metric. We used methods identical to those in the richness experiment (see main text, Prediction 2a and b) to determine whether particular diet mixtures had significant negative effects on herbivore performance (pupal weights, development speed, and/or survival). Next, we used linear models to assess whether the number of herbivores negatively affected by phenolics increases with increasing evenness.

Results

Evenness and structural diversity have little to no effect on herbivore performance (synergy hypothesis not supported).

In our assessment of the effects of evenness and structural diversity on standardized performance metrics, we found no clear evidence that performance decreases with increasing phytochemical diversity (Figure S5). For pupal weights, the top model included evenness, species, sex, structural diversity, and interactions including evenness by sex, evenness by species, sex by species, and evenness by sex by species (Table S12). However, in contrast to predictions of the synergy hypothesis, when parameters were averaged across four component models (Table S13), there was no overall effect of evenness (P = 0.391), no effects of structural diversity, and no interactions between evenness and structural diversity (Table S13). We did find a significant three-way interaction of evenness by sex by

species (Table S13), thus, these results should be interpreted alongside additional analyses (reported below) that are conducted separately for sex and species. For development speed, the top model included evenness, sex, species, structural diversity, and interactions including evenness by structural diversity, evenness by species, structural diversity by species, and sex by species (Table S12). Averaged across nine component models, in contrast to predictions based on the synergy hypothesis, there were no effects of evenness (P = 0.59), no effects of structural diversity, and no evenness x structural diversity interactions (Table S13). Again, we did find significant three-way evenness by sex by species interactions (Table S13), necessitating further analyses separated by sex and species to interpret evenness (see below). For survival, the top model included evenness, species, and evenness by species interactions (Table S12). Averaged across two component models (Table S13), there were no overall effects of evenness (P = 0.37), no effects of structural diversity, and no evenness by structural diversity interactions (Table S13).

Based on the multiple interactions involving evenness, species, and sex (Table S13), we also assessed the effects of phytochemical diversity on performance separately for each species, sex, and performance metric (twenty separate models; Table S14; Figure S6). We found three cases where evenness impacted herbivore performance, but results were mixed. In support of the synergy hypothesis, as compound evenness increased, there was a decrease in development speed of male *P. xylostella* (P = 0.032; Table S14; Figure S6B), and a decrease in pupal weights of male *P. xylostella* (P = 0.047; Table S14; Figure S6C). However, in contrast, there was also a significant effect of evenness in the opposite direction than predicted by the synergy hypothesis: as compound evenness increased, there was a corresponding increase in pupal weights of male *S. frugiperda* (P = 0.028; Table S14; Figure S6A). We also found three cases where structural diversity impacted herbivore performance, but patterns were, in all cases, in contrast to the synergy hypothesis. For both female and male *H. zea* and female *S. frugiperda* larvae developed faster on high structural diversity diets compared to low or medium (Figure S6D-F). There were no significant interactions between evenness and structural diversity (Prediction 1d) in any analyses (Table S14).

Increasing evenness increases the number of herbivores negatively affected (interaction diversity hypothesis supported)

As compound evenness increased from 0.2 (low evenness) to 1.0 (perfect evenness), the average number of herbivores negatively affected by phenolics increased by 42% from 2.67 to 3.78 organisms (t = 3.29; P = 0.0020; Fig. S7). However, there were no effects of structural diversity on the number of herbivores affected (F < 0.001; P = 1.0).

Discussion

In general, results from the evenness experiment further support the interaction diversity hypothesis: phytochemical diversity provides an adaptive advantage to plants by increasing the number of consumers that are negatively affected by any given phytochemical composition. Increasing evenness of a mixture increased the number of herbivores affected (Fig. S7) suggesting that the main advantage of more even mixtures is that they are more likely to contain multiple compounds at an effective dose that can provide simultaneous resistance to multiple herbivores. However, in contrast to the synergy hypothesis, increasing evenness did not affect the performance of any single caterpillar species (Fig. S5).

This may be because most six compound mixtures would contain only one or a few compounds that were bioactive against any given herbivore (see Fig. 4, main text).

In this experiment, we calculated evenness based on mass abundance of compounds (i.e. % fresh mass of diet; as is typical in observational studies that incorporate evenness into diversity metrics). An alternative to calculating evenness based on mass abundance is to use molar abundance, but this may not account for the fact that, in the case of phenolics, larger compounds often have multiple hydroxyl groups and other active sites that can increase their bioactivity. In either case, it is critical to note that a mixture that is perfectly even in terms of mass abundance (or molar abundance) may be very uneven in terms of bioactivity of compounds. This would be especially true in studies across different classes of compounds with very different thresholds for bioactivity (e.g. alkaloids vs phenolics). Interpreting the mechanisms behind an effect of evenness depends largely on understanding the minimum threshold for bioactivity of different compounds, which will vary across compounds and also depending on the consumer involved. Thus, future experimental work in model systems can seek to understand evenness both in terms of the relative abundances of compounds and their relative bioactivities in specific interactions.

From the plant perspective, evenness represents a metric of how resources are allocated across biosynthetic pathways. While we have much to learn about evenness as a trait that may be shaped by selection, it is clear that evenness varies across plants. For example, across wild and domesticated apples, on which the phenolic mixtures in this study were based, evenness can vary across tissues and genotypes from 0.24 to 0.98 (Whitehead and Poveda 2019). The extent to which this variation represents a heritable trait that can be shaped by natural selection is unclear. However, our results here suggest that evenness represents a key dimension of phytochemical diversity, along with richness and structural diversity, that may mediate plant-herbivore interactions.

References (also see main text)

Berenbaum, M. (1980). Adaptive significance of midgut pH in larval Lepidoptera. *The American Naturalist*, 115, 138–146.

Chen, L., Teng, H., Xie, Z., Cao, H., Cheang, W.S., Skalicka-Woniak, K., *et al.* (2018). Modifications of dietary flavonoids towards improved bioactivity: an update on structure–activity relationship. *Critical Reviews in Food Science and Nutrition*, 58, 513–527.

Shahidi, F. & Yeo, J. (2018). Bioactivities of phenolics by focusing on suppression of chronic diseases: a review. *International Journal of Molecular Sciences*, 19, 1573.

Xiao, J., Ni, X., Kai, G. & Chen, X. (2013). A review on structure–activity relationship of dietary polyphenols inhibiting α-amylase. *Critical Reviews in Food Science and Nutrition*, 53, 497–506.

Xiao, J., Ni, X., Kai, G. & Chen, X. (2015). Advance in dietary polyphenols as aldose reductases inhibitors: structure-activity relationship aspect. *Critical Reviews in Food Science and Nutrition*, 55, 16–31.

Appendix S1d. Supplementary Figures and Tables

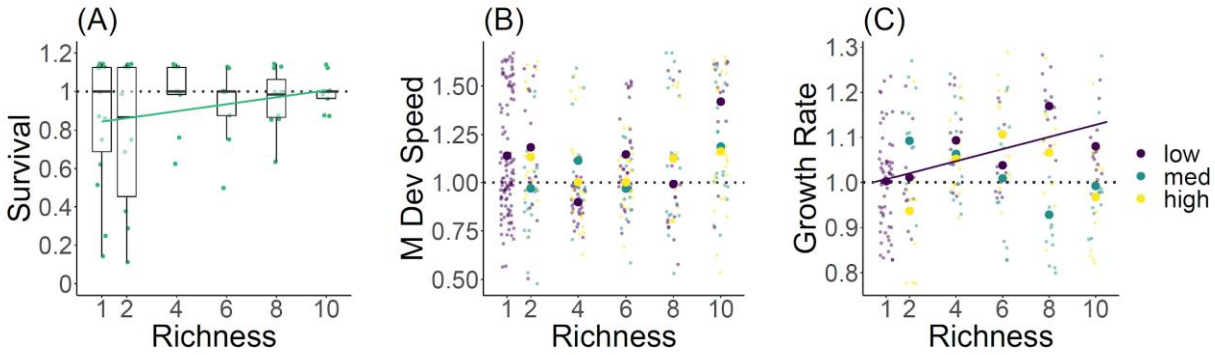


Fig. S1: Across 24 separate analyses, there were four cases of significant effects of richness, structural diversity, or their interaction on consumer performance (Table S5). Survival of *Spodoptera frugiperda* increased with increasing phenolic richness (A). For *P. xylostella* male development speed, there was a significant interaction between richness and structural diversity and a significant overall effect of structural diversity (B). Exploring the interaction, we found no effects of richness at any level of structural diversity. The only potential driver of the interaction was that, for low levels of richness only (two compound mixtures), there was marginal support for an effect of structural diversity (P = 0.085), with herbivores performing slightly worse on medium structural diversity diets compared to high. Finally, for growth rate of the plant pathogenic fungi *B. dothidea*, there was a significant interaction between richness and structural diversity (C), where growth rate increased with phenolic richness at low levels of structural diversity (P = 0.007), but was not affected at medium or high levels of structural diversity.

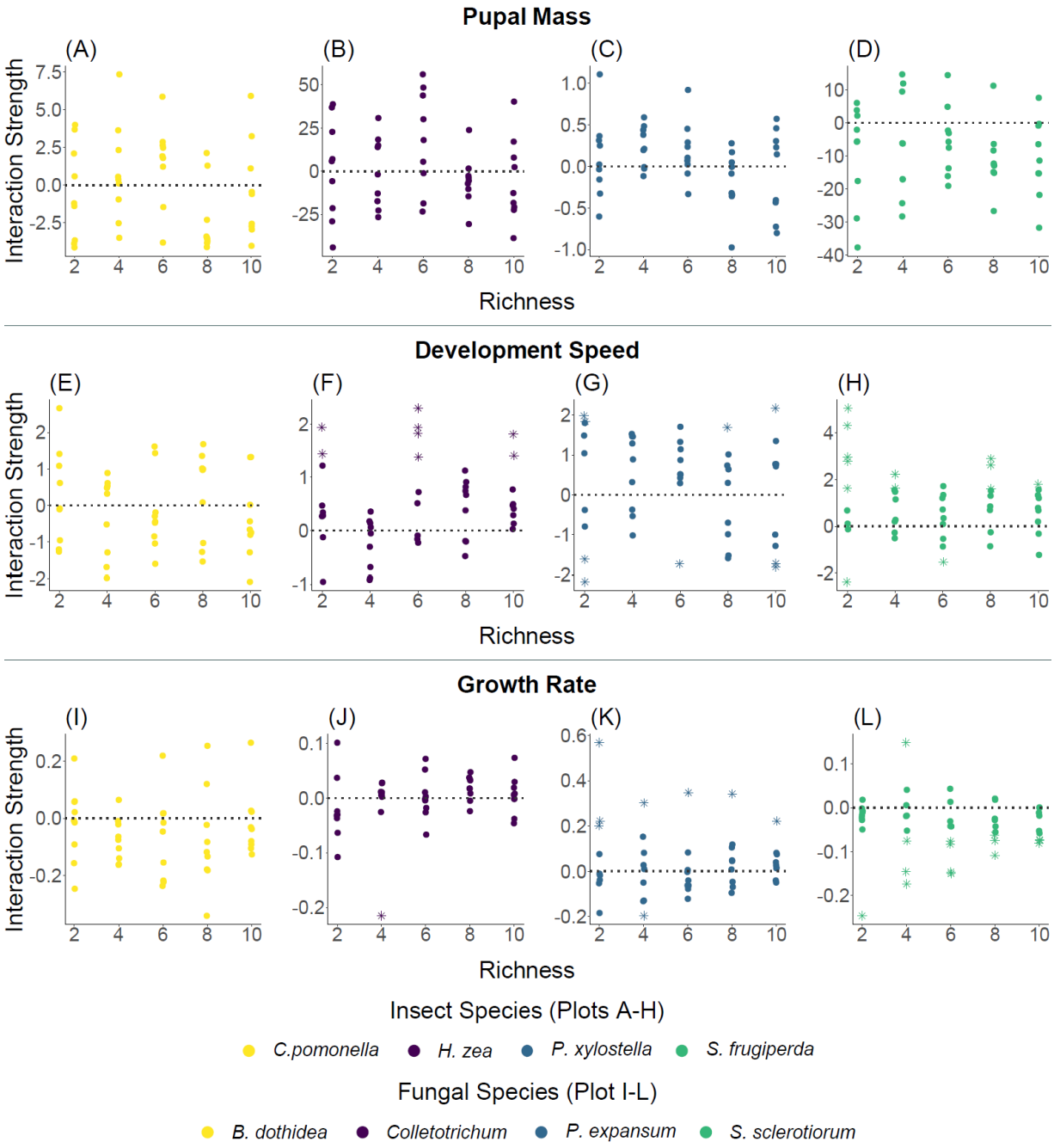


Figure S2: Increasing phenolic richness does not increase the potential for compound synergies that reduce insect performance (Panels A-H) or fungal performance (Panels I-L) (Table S6). Interaction strength is the expected (based on additive interactions) minus the observed effect of compound mixtures in our experiments. Values greater than zero indicate synergy, and values less than zero indicate antagonism. Each point represents a unique diet mixture, with departures from expected additive effects (non-overlapping 95% CI for expected and observed effects) indicated with stars. Counter to predictions for the synergy hypothesis, trends for 10 of 12 models (though not statistically supported; Table S6) were in the opposite direction than predicted, with interaction strength decreasing with richness.

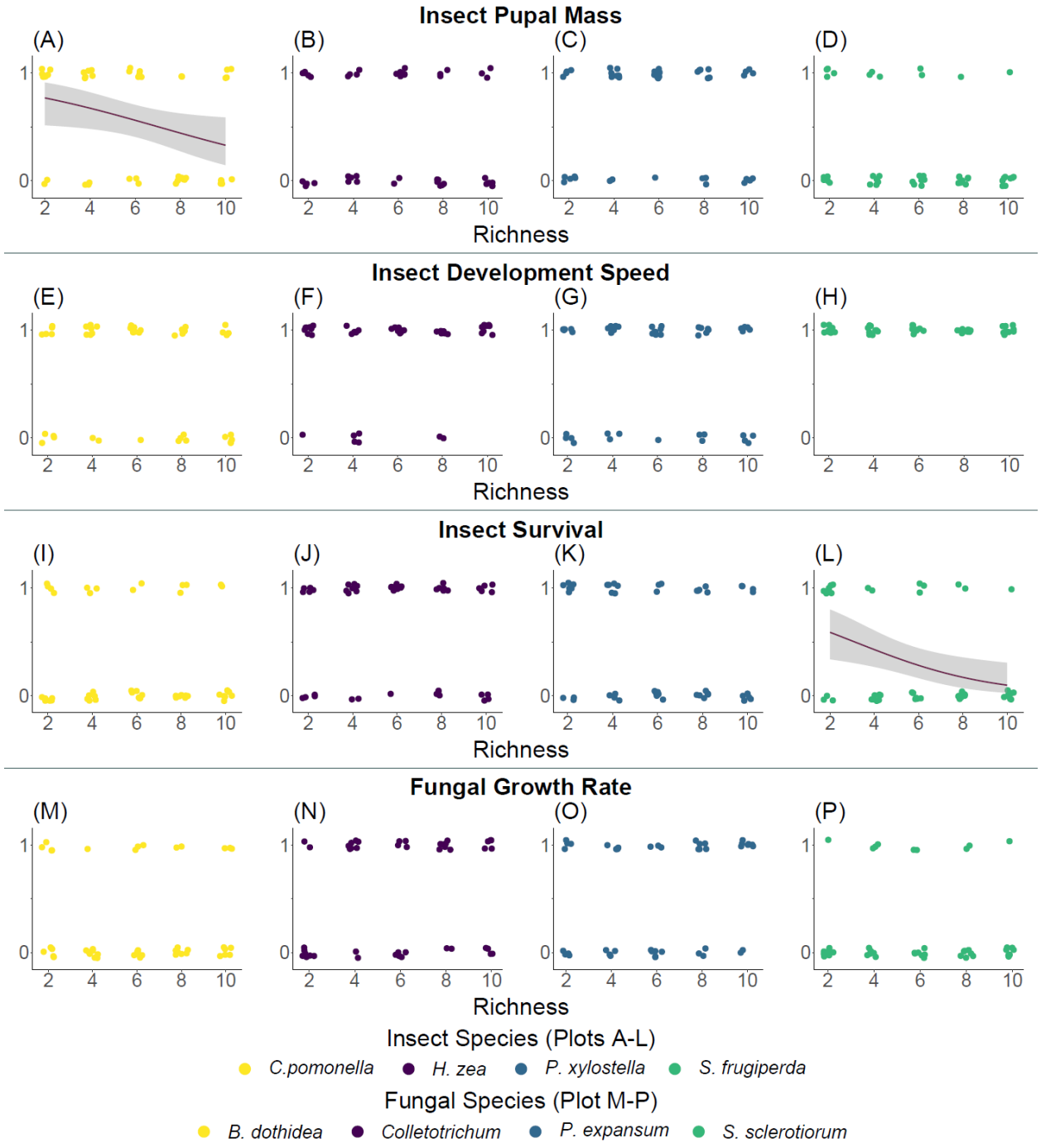


Figure S3: Increasing phenolic richness either decreases or does not affect the probability that a mixture will be more bioactive against insects (A-L) or fungi (M-P) than its average singleton (Prediction 1e not supported). Each point represents a binary outcome for a unique diet mixture (N = 45 mixtures per plot), where 0 indicates that the mixture was less bioactive against a given consumer than the average bioactivity of singletons in that mixture, and 1 indicates that the mixture was more bioactive. Results from binomial GLMMs are provided in Table S7.

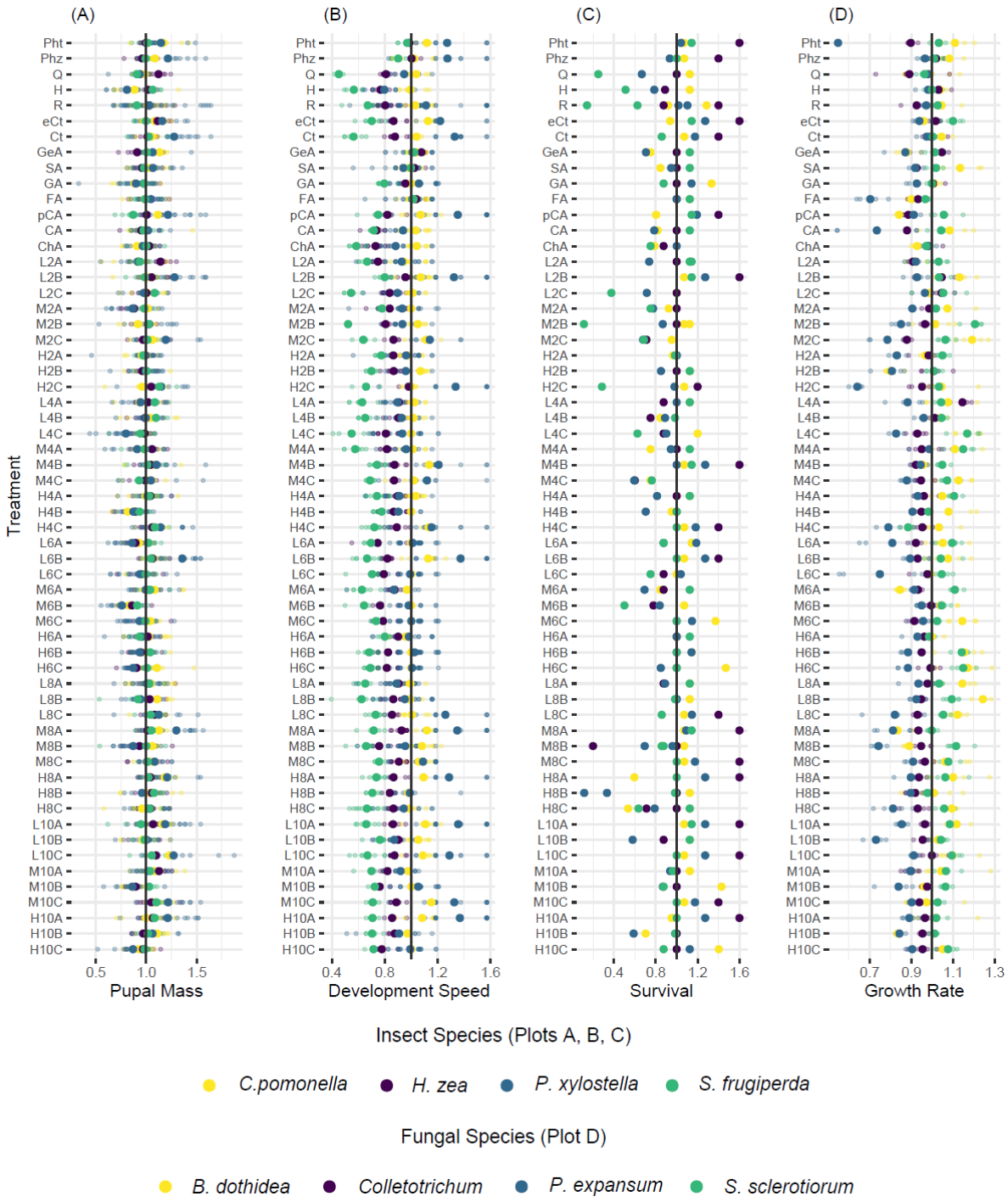


Figure S4: The effects of phenolics on performance are highly variable and depend on the specific diet treatment (mixture) and the consumer identity (Table S8). For pupal mass (A), development speed (B) and fungal growth (D), larger bold points indicate the mean performance across all replicates for a particular species, and the smaller points show each data point. For survival (C) data were first summarized as proportion survival for all individuals in a particular treatment (n > 5 individuals per herbivore per treatment) and then standardized relative to controls. See Table S1 for complete descriptions of diet treatments.

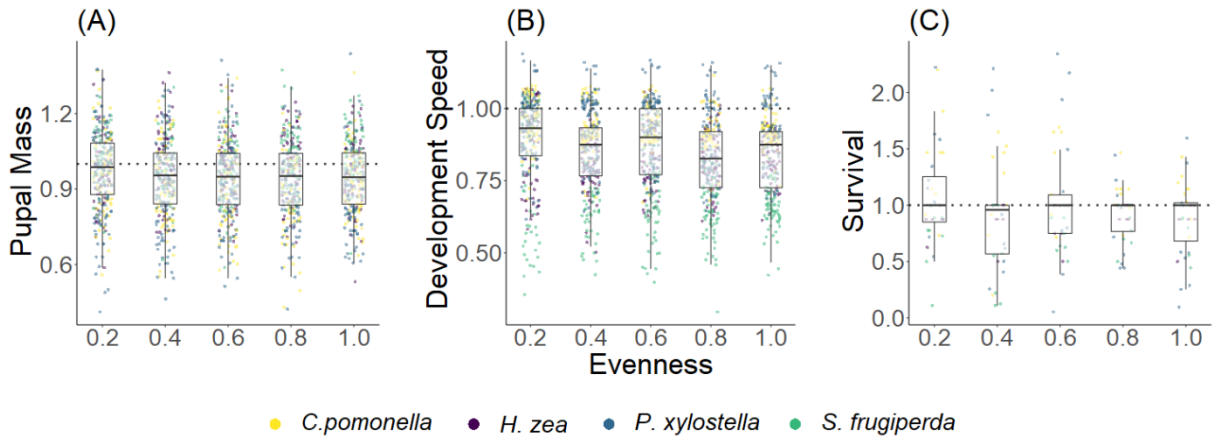


Fig. S5: Phenolic evenness has no effect on herbivore pupal mass (A), development speed (B), or survival (C). All performance response variables are standardized performance metrics; values less than one indicate that performance decreased on phenolic-supplemented diet relative to controls and vice versa. For pupal mass (A) and development speed (B), each point represents an individual insect (N = 1436 survivors). For survival (C), each point represents the proportion of survivors across all insects reared on a particular diet composition (N = 180 [45 compositions x 4 species] based on 2340 individuals). See Table S11 for a description of compound compositions at each evenness level and Table S13 for statistical results.

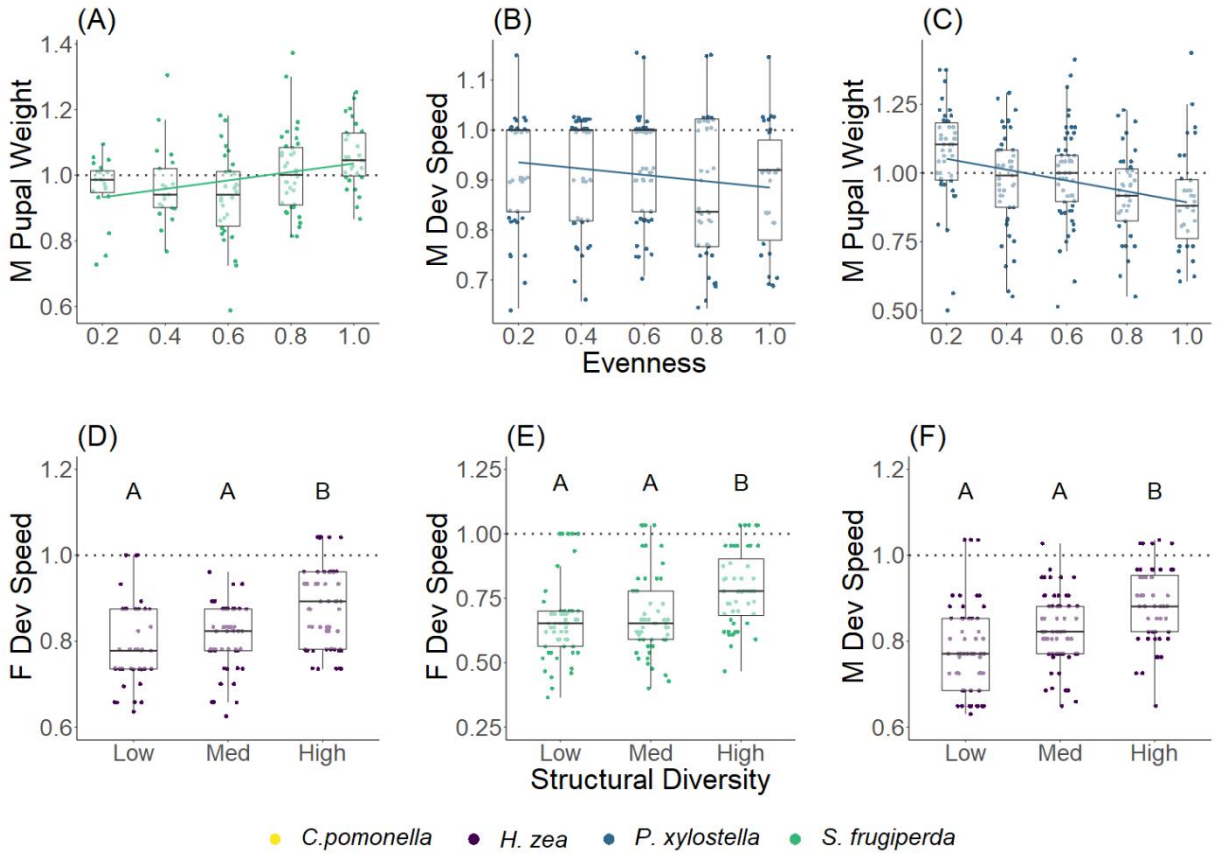


Fig. S6: Across 20 separate analyses, there were six cases of significant effects of evenness or structural diversity on consumer performance (Table S14).

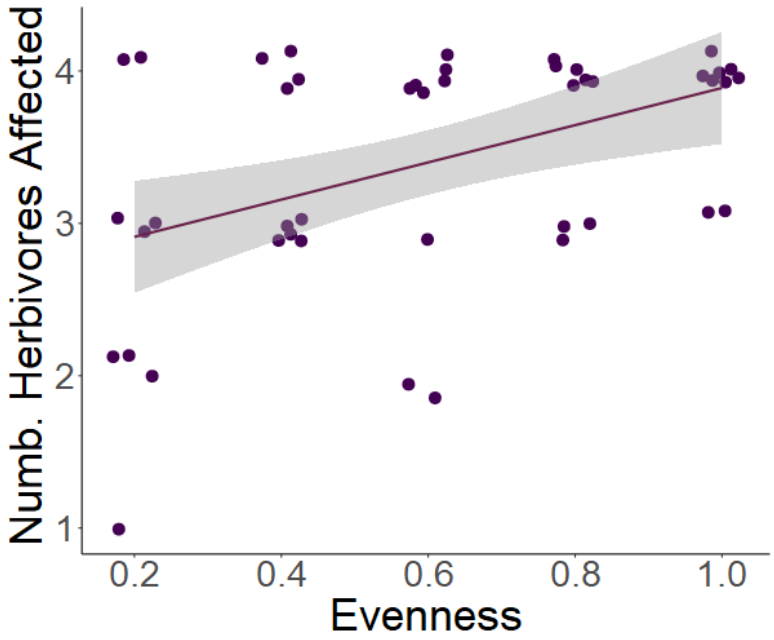


Fig. S7: The number of insect herbivores negatively affected by phenolics increased with increasing evenness of phenolic mixtures. Each point represents a unique diet composition (N = 45; Table S11). An herbivore was considered negatively affected if there was statistical support (at 95% confidence) that any aspect of performance (i.e. survival, pupal mass, and/or development speed) was reduced on phenolic-containing diets compared to controls.