Plant secondary metabolite diversity reflects both phylogeny and ecological adaptation

Simon Pierce^{1*}, Wen-Yong Guo², Bruno. E.L. Cerabolini³, Daniel Negreiros⁴, Franco Faoro¹, Giulia Magoga¹, Matteo Montagna^{1,6}, G. Wilson Fernandes⁴, Alberto Spada¹

¹Department of Agricultural and Environmental Sciences – University of Milan, via Celoria 2, 20133 Milano, Italy. ²Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, 200241 Shanghai, P.R. China. ³Department of Biotechnologies and Life Sciences (DBSV), University of Insubria, via J.H. Dunant 3, I-21100 Varese, Italy. ⁴Ecologia Evolutiva e Biodiversidade/DBG, CP 486, ICB/Universidade Federal de Minas Gerais, 30161-970, Belo Horizonte, MG, Brazil. ⁶BAT Center - Interuniversity Center for Studies on Bioinspired Agro-Environmental Technology, University of Napoli "Federico II", Portici, Italy.

*email: simon.pierce@unimi.it

Abstract

A phylogenetic framework explaining plant secondary metabolite diversity is lacking, but metabolite classes could represent adaptations to habitat resource availability. We test the hypothesis that primary adaptive strategies (competitors, C; stress-tolerators, S; ruderals, R) are associated, respectively, with nitrogenous metabolites synthesized in persistent organs (alkaloids), nitrogen-lacking aromatic terpenes and phenolics, and nitrogenous compounds prevalent in reproductive tissues (cyanogenic glucosides and glucosinolates). A matrix was compiled of 1019 species for which secondary metabolite pathways and CSR strategies are known. Accounting for phylogenetic relatedness and native biomes, we found that most phytochemical pathways did not correlate with strategy axes, but certain key associations were evident. C-selection was positively associated with amino acid-derived phenylpropanoids (low phylogenetic relatedness; $\lambda < 0.5$) and pyrrolizidine alkaloids and galloyl derivatives (high λ), and negatively with N-lacking linear monoterpenes (low λ). Nitrogenous cyanogenic glucosides positively correlated with R-selection (low λ). Terpenoids were widely distributed, but correlated positively with S- and negatively with R-selection (low λ). Twenty-six correlations between phytochemicals and biomes (low λ) were evident. Most secondary metabolite synthesis pathways are widespread, reflecting common roles and obligate defence, and strong phylogenetic effects are often evident. However, the character of phytochemical/adaptive strategy associations agrees with ecological theory and thus reflects adaptation.

Keywords: CSR theory, Grime, plant defence/defense, plant functional type, plant toxin, universal adaptive strategy theory

Introduction

Many plant products and secondary metabolites are perceived as characteristic of specific taxonomic groups, but secondary metabolite diversity is difficult to explain solely in terms of phylogenetic relationships because no overarching phylogenetic pattern is evident¹⁻⁷. Even when specific metabolites are apparently restricted to particular taxa this often reflects a lack of wider investigation and depends on the identity of the species from which the compound was first isolated. For example, Cucurbitacins, typically isolated from the eponymous Cucurbitaceae, are now known to occur throughout a range of disparate families including Elaeocarpaceae, Rosaceae, Rubiaceae, Scrophulariaceae and Sterculiaceae⁸⁻¹⁶. Even the well-studied mustard-oil system (i.e., employing glucosinolates with a myrosinase ‘activator’ compartmentalised in myrosin cells), often viewed as typical of the Brassicaceae, has been recorded from 16 families encompassing nine angiosperm orders and has evolved at least twice in two major ‘mustard-oil’ clades¹⁷. Cyanogenic glucosides, also often associated with Brassicaceae, are even more widely distributed across major eudicot clades, monocots, gymnosperms and ferns^{18,19}. Specific compounds can sometimes be used as chemotaxonomical markers to distinguish taxa, reflecting micro-cladogenic events (e.g., certain iridoid glucosides²⁰⁻²¹). However, widespread occurrence of secondary metabolite classes is understandable because of their derivation from common precursors in primary metabolism, following minor modification of pre-existing biosynthesis pathways²². For example, pathways involved in terpenoid synthesis were apparently present in the ancestor of all land plants^{3,5}.

As our view of secondary metabolites shifts from one of firm phylogenetic constraints and taxonomic characters²³ to one of plants as ‘Jacks-of-all-trades, masters of all’²⁴, it is perhaps accurate to say that some taxa tend to favour or emphasise production of certain secondary metabolite classes. Some can specialise in production of very specific molecules within these classes, the result being a ‘patchy’ distribution of secondary metabolites across the plant kingdom⁵. Such apparent trends and specific secondary metabolite roles may reflect a variety of processes, including ancient horizontal gene transfer from bacteria and fungi (and possibly viral recombination), widespread but silent genes, and direct lateral transfer of the metabolites themselves from other organisms, particularly symbiotic endophytes and mycorrhizal fungi^{3,5,6}.

The functions and chemical composition of secondary metabolites also potentially help explain their patchy distribution within the plant kingdom. They can be broadly classified as ‘nitrogen containing’ and ‘without nitrogen’⁵, and their production is thus potentially driven by adaptation to different edaphic resource availabilities across habitats, reflecting resource

investment trade-offs between growth (involving primary metabolism) and functions such as the attraction of pollinator and seed dispersal vectors and defence against herbivory, pathogens, or abiotic stressors (involving secondary metabolism). Indeed, as primary and secondary metabolism depend largely on the output of the shikimate, polyketide and mevalonic pathways²⁵ this potentially leads to competition between primary and secondary metabolism for carbon skeletons and nitrogen. For example, alkaloid synthesis competes with protein synthesis because both use amino acid precursors such as lysine or phenylalanine. The concepts of primary/secondary metabolism trade-offs and of differential allocation to defence at the cost of growth are inherent to various plant defence hypotheses, such as *optimal defence theory*²⁶, the *resource availability hypothesis*²⁷, and the *growth-differentiation balance hypothesis* (GDBH)²⁵.

In experimental studies, the operation of growth/defence trade-offs is suggested by limitation of relative growth rates^{28,29}. Trade-offs involving biomass production, flowering, fruiting, and seed production are evident in 82% of controlled studies that investigate defence costs³⁰. A recent review³¹ suggests that while growth/defence trade-offs are prevalent across biological systems, they sometimes do not manifest due to allocation to processes additional to growth and defence, such as storage or reproduction. Trade-offs can occur for one class of compounds but not for others^{32,33}, and are sometimes only evident under nutrient limiting conditions³⁴. Plant cultivation can facilitate both growth and defence, masking the fitness value of trade-offs operating in the wild³¹. Thus, allocation trade-offs are evidently complex and depend on the simultaneous action of multiple factors, both internal and external¹, which has led to difficulty placing the cost of secondary metabolite production in the context of overall plant fitness, although the absolute energetic cost of metabolite production is likely to be significant⁵.

In light of the complexity and context-dependency of growth/defence trade-offs and secondary metabolite production, is a general explanation of trends in secondary metabolite production possible?

A general explanation of plant functioning is attempted by plant ecological strategy theories. A “strategy” is the regime of resource investments between numerous adaptive traits in the face of multiple natural selection pressures³⁶, or “an instantaneous process of resource allocation between competing functions, maximising fitness across contrasting niches during development”³⁷. Secondary metabolites are potentially functional traits (i.e., characters that affect fitness and survival³⁸), with interspecific variability in defence metabolites and defence/herbivory relationships playing a significant role in species coexistence³⁹. The concept

of ‘plant defence syndromes’ comprising a range of traits and multiple defence compounds^{4,24} is consistent with that of plant ecological strategies, and indeed can be conceptualized as a component of the general plant strategy. The principal components of plant functional variability world-wide are summarized by the ‘global spectrum of plant form and function’: an observation of the importance of plant resource economics (acquisitive vs. conservative management of matter and energy) and stature (organ and whole-plant size)⁴⁰. The resource economics spectrum essentially represents a nitrogen allocation trade-off between structural (cell wall) investment and metabolism⁴¹, highlighting the competing roles of mineral resources in plant structural and metabolic adaptations and the potential integration of trade-offs in both primary and secondary functioning, and does appear to be involved in plant defence trade-offs^{42,43}.

Currently the only theory that can explain resource economics and stature as the principal evolutionary responses to environmental selection pressures is Grime’s⁴⁴⁻⁴⁶ competitor, stress-tolerator, ruderal (CSR) theory, for which economics and size traits delimit a common frame of reference across vascular plant life forms⁴⁷⁻⁵¹. Notably, CSR theory was cited by, and thus evidently inspired, subsequent growth/defence trade-off theories^{25,27}, and thus CSR theory promises to provide a general plant eco-evolutionary context in which to understand adaptive/defence specialisations.

Which specific specialisations can we expect?

S-selected ‘stress-tolerators’, no matter what type of stressor or limitation to metabolic performance, tend to invest in constitutive physical defence (inherently tough and persistent tissues) forming part of a conservative strategy in which resources and biomass are gradually accrued^{52,53}. S-selection is thus associated with high tissue dry matter content, high C:N, and a greater extent of mechanical and parenchyma tissues^{47,54} which probably reflects greater investment in cell wall materials and reserves that can be used to support metabolism during periods of limitation. The secondary metabolites associated with S-selection could thus be expected to respect this high C:N and be based predominantly on compounds lacking N, such as phenolics and terpenoids.

Alternatively, C- and R-selected species occupy relatively nutrient-rich habitats and can be expected to experience greater availabilities of mineral resources and thus N-containing secondary metabolites. Alkaloids are typically synthesized in roots or rhizomes and are subsequently translocated to and accumulated in other plant parts⁵⁵⁻⁵⁷. Occasionally alkaloid biosynthesis occurs in leaves (e.g., caffeine accumulates in seeds of *Coffea arabica* from biosynthesis in younger leaves⁵⁸) or in stem sieve elements adjacent to laticifers⁵⁹, but rarely

does alkaloid synthesis occur in fruits and seeds. Thus, alkaloid production is more prevalent in larger perennial herbs with persistent root systems and rhizomes, and can be expected to be associated with C-selection. N-containing compounds such as cyanogenic glucosides and glucosinolates can be synthesized in any plant part but are typically found in elevated concentrations in reproductive structures, particularly fruits and seeds, for both annual^{60,61} and perennial⁶² species. This emphasis on the support of reproductive function is also a key concept for R-selection, and we can reasonably expect a tendency towards cyanogenic glucoside and glucosinolate production in relatively R-selected species. However, exceptions to these possible trends exist (e.g., alkaloids in the annual opium poppy, *Papaver somniferum*, cyanogenic glucosides in rosaceous stone fruits), and the extent to which these different types of secondary metabolite defences are embroiled in primary adaptive strategies remains to be tested. This was the main objective of the present study.

Specifically, our hypotheses were that: 1). the extent of S-selection is correlated with production of secondary metabolites based on N-lacking molecules, such as terpenoids or phenolics, 2). the degree of C-selection is associated with capacity to synthesize N-containing alkaloids, 3). R-selection is associated with N-containing secondary metabolites typically found in greater quantities in reproductive tissues, such as cyanogenic glucosides and glucosinolates. Additionally, we expect these compounds to reflect the environment and thus be differentially represented in distinct biomes. Here, we compare the extent of C-, S- and R-selection with the ability to synthesize different classes of phytochemicals evident from a global database of vascular plant species for which both CSR strategies and secondary metabolite compounds are known. We use phylogenetic linear regression to account for phylogenetic relatedness, and for relationships between phytochemical clusters and biomes world-wide. While this approach may include limitations (such as not being able to differentiate between obligate or facultative use of secondary metabolites, and restriction to a small portion of the world flora) it is presently the most comprehensive way of placing secondary metabolites in a general ecological context, using the largest globally relevant dataset currently available.

Materials and methods

CSR strategies for thousands of plant species world-wide are available as supplementary material to the global CSR strategy analysis of Pierce *et al.*⁶³, which was employed here. For each species, we searched the global phytochemical interactions database (PCIDB;

www.genome.jp/db/pcidb) and found a total of 1019 vascular plant species present in both databases, encompassing 52 orders and 139 families of vascular plants, and including 892 dicotyledonous species (basal and eudicots), 87 monocotyledons, 19 gymnosperms and 21 pteridophytes. Thus, while each separate database includes many thousands of species, the current study was restricted to the species present in both databases (i.e., not all species known to produce, for example, glucosinolates, were included and not all species known to be, for instance, strongly R-selected were present).

The PCIDB has the particular advantage of a standardised classification system of secondary metabolites, employing 83 phytochemical cluster classes which include information on chemical structure, including KNApSAC metabolite codes⁶⁴ and KEGG Chemical Function-and-Substructures codes (KCF-S)⁶⁵ (the full list of clusters is presented in Table S1[**Supplementary Information**]). Species in the PCIDB are classified according to their National Center for Biotechnology Information (NCBI) taxonomic name and ID code (www.ncbi.nlm.nih.gov/taxonomy)⁶⁶, and in the present study the NCBI open chemistry database (<https://pubchem.ncbi.nlm.nih.gov>) was also referred to in order to double-check chemical formulas and structures, specifically with regard to N-containing vs. N-lacking compounds. Species nomenclature was checked according to the Plant List (www.theplantlist.org) to account for synonyms between NCBI and the nomenclature used by the CSR strategy database of Pierce *et al.*⁶³.

Associations between secondary metabolite use and the native biome(s) of each species were also investigated. Fourteen biomes were defined according to Olson *et al.*⁶⁷. Information regarding the biome(s) within the native range of each species was obtained from the comparison of the global distribution of biomes with the georeferenced records of each species (obtained from GBIF; www.gbif.org/species/search) occurring inside its native range. The native range of each species was obtained from sources such as GRIN taxonomy for plants (www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl?language=en), eMonocot (<http://emonocot.org>) and Catalogue of Life (www.catalogueoflife.org/col/search/all). The C-, S-, and R- values of species, the number of compounds in each phytochemical cluster known for each species, and data regarding presence/absence in biomes were collated into a matrix.

As leaf economics traits are subject to phylogenetic effects⁶⁸, phylogenetic relatedness was accounted for by constructing a phylogenetic tree for the study species based on the dated supertree of Zanne *et al.*⁶⁹, corrected and extended by a recent tree including over 31,000 species based on several genetic markers⁷⁰. All analyses were performed in the R environment⁷¹. An R function, *S.PhyloMaker*⁷¹ was used to produce phylogenies for subsets of

species using Qian and Jin's⁷⁰ "Scenario 3" approach (i.e., adding absent species to their families or genera following Phylomatic and BLADJ⁷²). Phylogenetic linear regression was applied to analyse the relationships between the 83 clusters, C-, S- and R-values and biome classes while accounting for phylogenetic relatedness, using the R package *phylolm*⁷³. The phylogenetically corrected coefficient of correlation was denoted by α (using the $p \leq 0.05$ level), and the strength of phylogenetic relatedness was denoted by λ (on a scale of 0 to 1, with $\lambda < 0.5$ interpreted as relatively low phylogenetic relatedness).

Results

After phylogenetic correction, only seven out of 83 clusters were significantly correlated with C-, S- or R-selection, and only five of these with relatively low phylogenetic relatedness between species ($\lambda < 0.5$). The extent of C-selection was significantly positively correlated with the production of pyrrolizidine alkaloids (cluster 2: $\alpha = 0.099$, $p = 0.004$) and galloyl derivatives and ellagitannins (cluster 81: $\alpha = 0.062$, $p = 0.015$) but with strong phylogenetic relatedness ($\lambda = 0.782$ and 0.853 , respectively) (Fig. 1A,D). C-selection was significantly positively correlated with coumarin and furanocoumarin phenylpropanoids (cluster 25; $\alpha = 0.092$, $p = 0.036$) with a low phylogenetic effect ($\lambda = 0.395$; Fig. 1B), and negatively correlated with linear monoterpenes (cluster 34: $\alpha = -0.079$, $p = 0.004$), with a negligible phylogenetic effect ($\lambda = 0.000$; Fig. 1C). The ten species in the database that produced the greatest number of pyrrolizidine alkaloids were all classed as C, C/CR, C/CSR or CR-selected: for example, *Symphytum officinale* (tertiary CSR strategy = C/CR; C:S:R = 69.7 : 0.0 : 30.3 %) is known to produce eight distinct pyrrolizidine alkaloids. The species exhibiting the greatest diversity in aromatic coumarin and furanocoumarin phenylpropanoids were the C-selected Apiaceae *Heracleum sphondylium* (C:S:R = 82.8 : 6.8 : 10.5 %), *Angelica archangelica* (C:S:R = 81.4 : 8.8 : 9.8 %) and *Peucedanum ostruthium* (C:S:R = 66.2 : 17.1 : 16.7 %).

The extent of S-selection was positively correlated with production of terpenoids (cluster 38; $\alpha = 0.287$, $p = 0.045$), with low phylogenetic relatedness ($\lambda = 0.399$; Fig. 2A), and with cadinanes (cluster 39; $\alpha = 0.044$, $p = 0.041$, $\lambda = 0.759$; Fig. 2B) and labdanes (cluster 46; $\alpha = 0.019$, $p = 0.038$, $\lambda = 0.235$; Fig. 2C). Species that produced ten or more different kinds of terpenoids included the strongly S-selected *Cistus creticus* (C:S:R = 11.8 : 88.2 : 0.0 %), *Acorus calamus* (C:S:R = 5.8 : 61.1 : 33.1 %), *Rosmarinus officinalis* (C:S:R = 3.5 : 96.5 : 0.0 %) and *Pinus halepensis* (C:S:R = 2.4 : 97.6 : 0.0 %), but also some C-selected species (e.g., *Valeriana*

officinalis; C:S:R = 80.4 : 0.0 : 19.6 %) and broadly R-selected species (e.g., *Artemisia annua*; C:S:R = 24.9 : 28.3 : 46.8 %). The S-selected species mentioned here also produced the greatest variety of cadinanes. Species producing the greatest numbers of labdanes included, again, S-selected angiosperms such as *Cistus creticus* and *Marrubium vulgare* (C:S:R = 15.9 : 66.8 : 17.3 %), extremely S-selected gymnosperms (e.g., *Platycladus orientalis*, C:S:R = 0.0 : 100.0 : 0.0 %; *Juniperus communis*, C:S:R = 0.0 : 86.6 : 13.4 %), but also some domesticated cereals such as *Zea mays* (C:S:R = 83.0 : 8.6 : 8.4 %) and *Oryza sativa* (C:S:R = 37.9 : 37.1 : 25.0 %).

The extent of R-selection was negatively correlated with flavanone flavonoids (cluster 14; $\alpha = -0.127$, $p = 0.047$, $\lambda = 0.448$; Fig. 3A) and terpenoids (cluster 38; $\alpha = -0.339$, $p = 0.034$, $\lambda = 0.419$; Fig. 2A), but positively correlated with production of cyanogenic glucosides derived from valine or isoleucine (cluster 73; $\alpha = 0.032$, $p = 0.029$, $\lambda = 0.000$; Fig. 3B). Species exhibiting production of multiple flavanone flavonoids included a wide range of perennial species encompassing C-selected woody angiosperms (*Paulownia tomentosa*, C:S:R = 84.2 : 2.7 : 13.2 %; *Vitis vinifera*, C:S:R = 72.5 : 0.0 : 27.5 %), S-selected gymnosperms (*Cedrus deodara*, C:S:R = 0.9 : 99.1 : 0.0 %), a range of both woody and herbaceous Fabaceae exhibiting intermediate CSR strategies (*Amorpha fruticosa*, C:S:R = 40.4 : 39.1 : 20.5 %; *Lupinus luteus*, C:S:R = 58.5 : 0.0 : 41.5 %; *Phaseolus vulgaris*, C:S:R = 57.7 : 19.4 : 22.9 %; *Robinia pseudoacacia*, C:S:R = 47.9 : 41.4 : 10.7 %), but also some R-selected species such as *Arabidopsis thaliana* (C:S:R = 4.3 : 0.0 : 95.7 %). The ten species exhibiting the greatest diversity of cyanogenic glucosides included extremely R-selected *Linum usitatissimum* (Linaceae; Fabid eudicot clade; C:S:R = 1.9 : 12.9 : 85.2 %) and *Arabidopsis thaliana* (Brassicaceae; Malvid eudicot clade), R-selected herbaceous Fabaceae such as *Trifolium repens* (C:S:R = 36.2 : 1.6 : 62.2 %), *Lotus corniculatus* (C:S:R = 14.2 : 19.0 : 66.8 %), *Ornithopus perpusillus* (C:S:R = 8.0 : 23.7 : 68.4 %), *Lotus maritimus* (C:S:R = 23.8 : 30.8 : 45.4 %), and CR-selected Ranunculaceae (i.e., basal eudicots) such as the forb *Caltha palustris* (C:S:R = 59.1 : 0.0 : 40.9 %) and the vine *Clematis flammula* (C:S:R = 41.1 : 24.5 : 34.4 %). However, tree species were also able to produce cyanogenic glucosides, such as *Euonymus japonicus* (S/CS), *Fagus sylvatica* (S/CS) and *Robinia pseudacacia* (CS).

Twenty-six significant correlations between clusters and global biomes were evident with low phylogenetic relatedness (Table 1). Notably, N-lacking compounds such as coumestanes and neoflavonoids (cluster 17) and aphidicolane terpenoids (cluster 41) were positively correlated with deserts and xeric shrublands ($\alpha = 0.145$, $p = 0.003$ and $\alpha = 0.530$, $p = 0.012$, respectively; Table 1A). The biome associated with the greatest phytochemical diversity, with low phylogenetic effects, was that of montane grasslands and shrublands. These

were positively associated with diversity in caffeate and tyramine derivative alkaloids (cluster 6), steroid alkaloids (cluster 11), lignans (cluster 21), terpenoids (cluster 38), carotenoids (cluster 59) and galloyl derivatives/ellagitannins (cluster 81) (see Table 1A for respective α and p values). A range of clusters were significantly negatively correlated with temperate forests: purine alkaloids (cluster 12) with temperate broadleaf/mixed forest, while cadinanes (cluster 39), aphidicolane terpenoids (cluster 41) and cyanogenic glucosides derived from valine or isoleucine (cluster 73) correlated negatively with temperate coniferous forests (Table 1A). Tropical and subtropical moist broadleaf forests were particularly associated with diversity in iridoids (cluster 36), and negatively with pyrrolidine alkaloids (cluster 1) (Table 1A). Boreal forests and taiga, Mediterranean forests, woodlands and scrub, and tundra biomes were not significantly correlated with any particular phytochemical cluster (Table 1B).

Discussion

Analysis of a global dataset demonstrates that while most secondary metabolites are widespread (reflecting roles as essential pigments, membrane components and general constitutive defences) and strong phylogenetic effects are often evident, the distribution of some secondary metabolites is associated with primary ecological strategies and native biomes. In these cases, the specific trends are also consistent with explanations suggested by ecological theory^{44-46,53}. In extreme summary: 1). certain nitrogenous compounds synthesized mainly in perennial plant parts (or some aromatics also derived from amino acid synthesis) are correlated with C-selection and thus adaptation to resource-rich, stable habitats, 2). some N-lacking aromatic compounds are associated with S-selection and thus adaptation to resource-poor habitats, and 3). N-containing cyanogenic glucosides are associated with R-selection and thus adaptation to resource-rich but unstable habitats. This statement is extremely general and the correlations, while statistically significant, are imperfect (e.g., not all competitors produce pyrrolizidine alkaloids, while some stress-tolerators do; Fig. 1A), but this can potentially offer a general explanation as to why certain groups of secondary metabolites seem to be favoured by particular species: species that share a general approach to survival can share a general approach to defence, even when they are not closely related phylogenetically.

Specifically, hypothesis 1 suggested that the extent of S-selection is correlated with production of secondary metabolites based on N-lacking molecules, such as terpenoids or phenolics. Positive correlations with the production of terpenoids, cadinanes and labdanes bears this out. The character of the species producing these secondary metabolites also agrees with

the hypothesis: most are strongly S-selected perennial species and represent a range of gymnosperms and angiosperms, including a number of Mediterranean taxa (e.g., *Cistus creticus*, *Marrubium vulgare*, *Pinus halepensis*, *Rosmarinus officinalis*). The positive association of the desert and xeric shrubland biome with N-lacking coumestanes and aphidicolanes (but not with alkaloids or other N-containing secondary metabolites) also agrees with this hypothesis.

Hypothesis 2 suggested that the degree of C-selection is associated with capacity to synthesize N-containing alkaloids, which are usually produced in perennial organs. The findings were broadly supportive of the hypothesis but outlined a slightly more complicated picture. The positive correlation of C-selection with pyrrolizidine alkaloids suggests that these may reflect adaptation to productive habitats, but some classes of alkaloids appear to be extremely widespread (pyrrolidine alkaloids, isoquinoline alkaloids). C-selection was also associated with production of certain aromatic compounds, particularly coumarin and furanocoumarin phenylpropanoids and galloyl derivatives. Many C-selected Apiaceae, such as *Cicuta virosa*, *Chaerophyllum hirsutum* and *Oenanthe fistulosa*, also produced polyynes, as did numerous CS- to CR/CSR-selected Fabaceae and Araliaceae. C-selected species producing coumarin and furanocoumarin phenylpropanoids were large Apiaceae such as *Angelica archangelica*, *Heracleum sphondylium* and *Pastinaca sativa*, Asteraceae such as *Inula helenium*, and other broadly C/CS- or C/CR-selected members of Solanaceae (*Atropa belladonna*), Calophyllaceae (*Calophyllum brasiliense*), Sapindaceae (*Aesculus hippocastanum*) and Moraceae (*Ficus* spp., *Morus* spp.). Crucially, although the final coumarin and furanocoumarin phenylpropanoid products are N-lacking aromatic compounds, they are derived from amino acid synthesis pathways that rely on N availability. Polyynes are also relatively complex molecules characterised by carbon chains with alternating triple bonds, and often including multiple aromatic sub-groups. Thus, it appears that C-selection is not necessarily associated with N incorporated directly in the final defence compound *per se*, but with N availability during synthesis.

The third hypothesis, that R-selection is associated with nitrogenous secondary metabolites typically used to protect reproductive structures, such as cyanogenic glucosides and glucosinolates, was found to be supported in the case of cyanogenic glucosides only. Cyanogenic glucosides are associated with R-selection across a range of families including Brassicaceae, Fabaceae, Linaceae, Phytolaccaceae, Ranunculaceae, Resedaceae and Tropeolaceae, with negligible phylogenetic effects (Fig. 3). Some CS- and S/SC-selected trees of various families, notably Fabaceae, Rosaceae and Myrtaceae were also able to produce

cyanogenic glucosides, in agreement with the known production of these compounds in genera such as *Acacia*, *Prunus* and *Eucalyptus*, respectively^{19,62}.

These associations and correlations may not be perfect, but imperfect correlations are an expected outcome of natural variability. Additionally, some of this variability may not be entirely natural. Many of the notable exceptions, such as relatively C-selected cereals that produce a range of aromatic labdanes (*Zea mays* and *Oryza sativa*) or the excessively chalcone-rich *Vitis vinifera*, are domesticated species. These could embody Züst and Agrawal's³¹ observation that cultivation alters the pressures selecting for specific growth/defence trade-offs with respect to the wild. Indeed, in the present study no wild graminoids produced labdanes, and Schmelz *et al.*⁷⁴ confirm that these are known within Poaceae only from domesticated cereals. *Vitis vinifera* has been artificially selected for a rich palate of organoleptic qualities that reflect, in part, secondary metabolite diversity. Indeed, domestication of *V. vinifera* involved selection of a range of genes affecting berry development, sugar transport and secondary metabolite production, especially monoterpenoid indole alkaloid biosynthesis, proanthocyanidin accumulation and stilbenoid biosynthesis⁷⁵. Domestication is a process of adaptation to symbiosis with the human species, in which some of the work (notably plant defence) is carried out by the human symbiont and for which the plant is exposed to fewer natural selection pressures^{6,76,77}. Thus, traits such as secondary metabolites that are strongly selected for in the wild can become less decisive for survival under cultivation and may be prone to genetic drift or artificial selection. From a wild situation in which plants maintain lasting resistance using a complex mixture of defence compounds, domestication involves human preference for monoculture and small numbers of externally applied defence compounds, potentially leading to rapid herbivore and pathogen adaptation⁶. Experimental studies have found that, across a range of cultivated taxa, domestication does alter plant defences and rates of herbivory, although changes in growth/defence allocation trade-offs are inconsistent and depend on the plant/herbivore system⁷⁸. There is unlikely to be a simple 'domestication = less innate defence' relationship as each species has a different propensity to allocate resources to additional functions such as storage or stress resistance⁷⁸. Nonetheless, the original ecological strategies of wild ancestral species are likely to influence defensive capacity following domestication.

Although some specific correlations between phytochemical clusters and biomes were found (such as the relationship between certain N-lacking metabolites and the desert biome) no overarching general trend emerged, probably reflecting the fact that each biome can include a wide range of habitats differing in productivity, causing 'noise' in the analysis. Indeed,

biome/CSR strategy associations have been observed, but are highly variable for this reason⁶³. It is likely that comparison of specific habitats and the presence of species synthesizing distinct phytochemicals would yield more specific insights regarding the relationship between strategies, secondary metabolites and environmental resource availabilities. In the present study, complications also arose from the highly generalised nature of the available phytochemical data (i.e., simply the presence/absence of compounds recorded for each taxon, revealing little of whether metabolites are constitutive or induced), lack of information regarding the site of synthesis and/or accumulation, and the relative energetic costs of each defence mechanism. Indeed, secondary metabolite contents may increase in response to abiotic stresses⁷⁹, and the current dataset can reveal nothing of responses or absolute concentrations of metabolites. Again, it could be hypothesized that metabolite concentration and the relative energetic cost of each metabolite pathway is related to habitat and adaptive strategy. Unfortunately, due to the complexity of metabolite synthesis pathways, currently a precise indication of the synthesis costs of major secondary metabolite classes is unachievable³⁵. Data in the present study are also extremely general in the sense that they represent mixed records – i.e., detections of distinct compounds in different samples by different laboratories – for ‘the species’. Almost nothing is currently known of the variability of the secondary metabolite garniture and concentrations of each compound between individuals within populations, and the extent to which this is genetically determined or reflects responses to soil nutrient availabilities.

As with other plant functional traits, intra-specific diversity in secondary metabolites and population-level variability is a prerequisite for natural selection, and almost certainly occurs. Thus, at fine scales, such as within habitats or along environmental gradients of productivity or stress, it is likely that growth/defence trade-offs occur, are important for the structuring of plant communities, and involve strategy-dependent differences in phytochemical production between species and between individuals within populations. Indeed, just as plant species diversity is tightly correlated with strategy diversity and functional trait variance⁸⁰, phytochemical diversity is likely to have been extended via the same eco-evolutionary feedbacks (*sensu* Grime and Pierce⁵³). Hopefully, using the applied CSR analysis method used here, we can construct an evolutionary and ecological context for plant secondary metabolites as functional traits that contribute to fitness when plants live together and with other organisms.

In conclusion, most secondary metabolite synthesis pathways are widely evident, reflecting common roles as membrane components, pigments, and general obligate defences. However, additional secondary metabolic specializations appear to have emerged as an integral part of

plant primary adaptive strategies in response to fundamental constraints of habitat productivity and stability. This represents further evidence in agreement with the concepts of secondary metabolites as functional characters affecting fitness, and of secondary metabolite involvement in complex growth/defence trade-offs occurring during plant adaptation.

Acknowledgements

This project was funded with the aid of grant number 19628 (RV_RIC_AT16SPIER; *Piano di Sostegno alla Ricerca* 2015-2017) awarded to S.P. from the Department of Agricultural and Environmental Sciences (DiSAA), University of Milan, Italy.

Author contributions

S.P., A.S., F.F. and B.E.L.C. conceived the work and compiled the database. S.P. also coordinated between work groups and wrote the manuscript. G.M., M.M. and W.-Y.G. performed the phylogenetic logistic regressions, D.N. and G.W.F. analysed associations between secondary metabolite pathways and biomes.

Competing interests

The authors declare no competing interests.

References

1. Koricheva, J. Meta-analysis of sources of variation in fitness costs of plant antiherbivore defences. *Ecology* **83**(1), 176-190 (2002).
2. Wink, M. Phytochemical diversity of secondary metabolites in *Encyclopedia of Plant and Crop Science* (ed. Goodman, R.M.) 915-919 (Marcel Dekker, 2004).
3. Wink, M., Botschen, F., Gosmann, C., Schäfer, H., Waterman, P.G. Chemotaxonomy seen from a phylogenetic perspective and evolution of secondary metabolites in *Annual Plant Reviews volume 40* (ed. Wink, M.) 364-433 (Wiley, 2009).
4. Agrawal, A.A. Current trends in the evolutionary ecology of plant defence. *Funct. Ecol.* **25**, 420-432 (2011).
5. Wink, M. *Evolution of Secondary Plant Metabolism*. (John Wiley and Sons, 2016).
6. Wink, M. Evolution, diversification, and function of secondary metabolites in *Encyclopedia of Evolutionary Biology vol. 4* (ed. Kliman, R.M.) 1-9 (Academic Press, 2016).
7. Pierce, S., Luzzaro, A., Cerabolini, B.E.L. Taxonomy, morphology and ecology of medicinal plants, a botanical perspective in *Herbal Medicines for Human Health* (eds. Bagetta, G., Corasaniti, M.T, Cosentino, M., Sakurada, S.) 7-26 (Taylor & Francis, 2016).

8. Ito, A., Chai, H.B., Lee, D., Kardono, L.B.S., Riswan, S., Farnsworth, N.R., Cordell, G.A., Pezzuto, J.M., Kinghorn, A.D. Ellagic acid derivatives and cytotoxic cucurbitacins from *Elaeocarpus mastersii*. *Phytochemistry* **61**(2), 171-174 (2002).
9. Muñoz, O., Esteyez-Braun, A.M., Rayelo, A.G., Gonzalez, A.G. Cucurbitacin F in seeds of *Kageneckia angustifolia* (Rosaceae). *Z. Naturforsch C* **57**(1-2), 208-209 (2002).
10. Litaudon, M., Gaspard, C., Sevenet, T. Morierinin, a new cytotoxic cucurbitacin from the leaves of *Morierina montana* Vieill. *Nat. Prod. Res.* **17**(4), 229-233 (2003).
11. Rodriguez, N., Vasquez, Y., Hussein, A.A., Coley, P.D., Solis, P.N., Gupta, M.P. Cytotoxic cucurbitacin constituents from *Sloanea zulaensis*. *J. Nat. Prod.* **66**(11), 1515-1516 (2003).
12. Allen, J.G., Colegate, S.M., Mitchell, A.A., Mulder, R.J., Raisbeck, M.F. The bioactivity-guided isolation and structural identification of toxic cucurbitacin steroidal glucosides from *Stemodia kingii* (Scrophulariaceae). *Phytochem. Analysis* **17**(4), 226-235 (2006).
13. Chen, Z.T., Lee, S.W., Chen, C.M. Cucurbitacin B 2-sulfate and cucurbitacin glucosides from the root bark of *Helicteres angustifolia*. *Chem. Pharm. Bull.* **54**(11), 1605-1607 (2006).
14. Bhandari, P., Kumar, N., Singh, B., Kaul, V.K. Cucurbitacins from *Bacopa monnieri* (Scrophulariaceae). *Phytochemistry* **68**(9), 1248-1254 (2007).
15. Kaya, G.I., Melzig, M.F. Quantitative determination of cucurbitacin E and cucurbitacin I in homoeopathic mother tincture of *Gratiola officinalis* L. by HPLC. *Pharmazie* **63**(12), 851-853 (2008).
16. Maloney, K.N., Fujita, M., Eggert, U.S., Schroeder, F.C., Field, C.M., Mitchison, T.J., Clardy, J. Actin-aggregating cucurbitacins from *Physocarpus capitatus* (Rosaceae). *J. Nat. Prod.* **71**(11), 1927-1929 (2008).
17. Rodman, J.E., Soltis, P.S., Soltis, D.E., Sytsma, K.J., Karol, K.G. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *Am. J. Bot.* **85**(7), 997-1006 (1998).
18. Vetter, J. Plant cyanogenic glycosides. *Toxicon* **38**, 11-36 (2000).
19. Gleadow, R.M., Møller, B.L. Cyanogenic glycosides, synthesis, physiology, and phenotypic plasticity. *Ann. Rev. Plant Biol.* **65**, 155-185 (2014).
20. Rank, C., Rasmussen, L.S., Jensen, S.R., Pierce, S., Press, M.C., Scholes, J.D. Cytotoxic constituents of *Alectra* and *Striga* species. *Weed Res.* **44**, 265-270 (2004).
21. Jensen, S.R., Gotfredsen, C.H., Pierce, S. Iridoid glucosides of *Paederota bonarota* and the relationships between *Paederota* and *Veronica*. II. *Biochem. Syst. Ecol.* **35**(8), 501-505 (2007).
22. Tanaka, Y., Sasaki, N., Ohmiya, A. Biosynthesis of plant pigments, anthocyanins, betalains and carotenoids. *Plant J.* **54**, 733-749 (2008).
23. Dahlgren, R.M.T. A revised system of classification of the angiosperms. *Bot. J. Linn. Soc.* **80**, 91-124 (1980).

24. Koricheva, J., Nykänen, H., Gianoli, E. Meta-analysis of trade-offs among plant antiherbivore defences, are plants Jacks-of-all-trades, masters of all? *Am. Nat.* **163**(4), E64-E75 (2004).
25. Herms, D.A., Mattson, W.J. The dilemma of plants, to grow or to defend? *Q. Rev. Biol.* **67**, 283-335 (1992).
26. Rhoades, D.F. Evolution of plant chemical defense against herbivores in *Herbivores, their Interaction with Secondary Plant Metabolites*. (eds. Rosenthal, G.A. and Janzen D.H.) 3-54 (Academic Press, 1979).
27. Coley, P.D., Bryant, J.P., Chapin III, F.S. Resource availability and plant antiherbivore defence. *Science* **230**, 895-899 (1985).
28. de la Mata, R., Hood, S., Sala, A. Insect outbreak shifts the direction of selection from fast to slow growth rates in the long-lived conifer *Pinus ponderosa*. *Proc. Natl. Acad. Sci. USA* **114**(28), 7391-7396 (2017).
29. Thompson, K.A., Cory, K.A., Johnson, M.T.J. Induced defences alter the strength and direction of natural selection on reproductive traits in common milkweed. *J. Evolution. Biol.* **30**, 1219-1228 (2017).
30. Strauss, S.Y., Rudgers, J.A., Lau, J.A., Irwin, R.E. Direct and ecological costs of resistance to herbivory. *Trends Ecol. Evol.* **17**(6), 278-285 (2002).
31. Züst, T., Agrawal, A.A. Trade-offs between plant growth and defense against insect herbivory, an emerging mechanistic synthesis. *Annu. Rev. Plant Biol.* **68**, 513-534 (2017).
32. Glynn, C., Herms, D.A., Orians, C.M., Hansen, R.C., Larsson, S. Testing the growth-differentiation balance hypothesis, dynamic responses of willows to nutrient availability. *New Phytol.* **176**, 623-634 (2007).
33. Massad, T.J., Dyer, L.A., Vega, G.C. Costs of defense and a test of the carbon-nutrient balance and growth-differentiation balance hypotheses for two co-occurring classes of plant defense. *PLoS One* **7**(10), e47554 (2012).
34. Sampedro, L., Moreira, X., Zas, R. Costs of constitutive and herbivore-induced chemical defences in pine trees emerge only under low nutrient availability. *J. Ecol.* **99**, 818-827 (2011).
35. Wink, M. Introduction, biochemistry, physiology and ecological functions of secondary metabolites. *Annual Plant Reviews* **40**, 1-19 (2010).
36. Guo, W.-Y., van Kleunen, M., Winter, M., Weigelt, P., Stein, A., Pierce, S., Pergl, J., Moser, D., Maurel, N., Lenzner, B., Kreft, H., Essl, F., Dawson, W., Pyšek, P. The role of adaptive strategies in plant naturalization. *Ecol. Lett.* **21**(9), 1380-1389 (2018).
37. Dayrell, R., Arruda, A.J., Pierce, S., Negreiros, D., Meyer, P.B., Lambers, H., Silveira, F.A.O. Ontogenetic shifts in plant strategies. *Funct. Ecol.* **32**, 2730-2741 (2018).
38. Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E. Let the concept of trait be functional! *Oikos* **116**, 882-892 (2007).

39. Vleminckx, J, Salazar, D., Fortunel, C., Mesones, I., Dávila, N., Lokvam, J., Beckley, K., Baraloto, C., Fine, P.V.A. Divergent secondary metabolites and habitat filtering both contribute to tree species coexistence in the Peruvian Amazon. *Front. Plant Sci.* **9**, 836 (2018).
40. Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I.C., Garnier, E., Bönsch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Chave, J., Wright, S.J., Sheremet'ev, S.N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M.D., Gorné, L.D. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
41. Onoda, Y., Wright, I.J., Evans, J.R., Hikosaka, K., Kitajima, K., Niinemets, Ü., Poorter, H., Tosens, T., Westoby, M. Physiological and structural tradeoffs underlying the leaf economics spectrum. *New Phytol.* **214**, 1447-1463 (2017).
42. Mason, C.M., Donovan, L.A. Does investment in leaf defences drive changes in leaf economic strategy? A focus on whole-plant ontogeny. *Oecologia* **177**, 1053-1066 (2015).
43. Chauvin, K.M., Asner, G.P., Martin, R.E., Kress, W.J., Wright, S.J., Field, C.B. Decoupling dimensions of leaf economic and anti-herbivore defense strategies in tropical canopy tree community. *Oecologia* **186**, 765-782 (2018).
44. Grime, J.P. Vegetation classification by reference to strategies. *Nature* **250**, 26-31 (1974).
45. Grime, J.P. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am. Nat.* **111**, 1169-1194 (1977).
46. Grime, J.P. *Plant Strategies, Vegetation Processes and Ecosystem Properties. Second Edition.* (Wiley Press, 2001).
47. Pierce, S., Ceriani, R.M., De Andreis, R., Luzzaro, A., Cerabolini, B. The leaf economics spectrum of Poaceae reflects variation in survival strategies. *Plant Biosyst.* **141**(3), 337-343 (2007).
48. Pierce, S., Brusa, G., Sartori, M., Cerabolini, B.E.L. Combined use of leaf size and economics traits allows direct comparison of hydrophyte and terrestrial herbaceous adaptive strategies. *Ann. Bot.* **109**(5), 1047-1053 (2012).
49. Pierce, S., Brusa, G., Vagge, I., Cerabolini, B.E.L. Allocating CSR plant functional types, the use of leaf economics and size traits to classify woody and herbaceous vascular plants. *Funct. Ecol.* **27**(4), 1002-1010 (2013).
50. Pierce, S., Bottinelli, A., Bassani, I., Ceriani, R.M., Cerabolini, B.E.L. How well do seed production traits correlate with leaf traits, whole plant traits and plant ecological strategies? *Plant Ecol.* **215**, 1351–1359 (2014).
51. Pierce, S., Cerabolini, B.E.L. *Plant Economics and Size Trait Spectra are Both Explained by One Theory.* (Plant Press, 2018).

52. Pierce, S., Vianelli, A., Cerabolini, B. From ancient genes to modern communities, the cellular stress response and the evolution of plant strategies. *Funct. Ecol.* **19(5)**, 763-776 (2005).
53. Grime, J.P., Pierce, S. *The Evolutionary Strategies that Shape Ecosystems*. (Wiley-Blackwell, 2012).
54. Cerabolini, B., Brusa, G., Ceriani, R.M., De Andreis, R., Luzzaro, A., Pierce, S. Can CSR classification be generally applied outside Britain? *Plant Ecol.* **210(2)**, 253-261 (2010).
55. Sander, H., Hartmann, T. Site of synthesis, metabolism and translocation of senecionine N-oxide in cultured roots of *Senecio erucifolius*. *Plant Cell Tiss. Org.* **18**, 19-31 (1989).
56. Shitan, N., Bazin, I., Dan, K., Obata, K., Kigawa, K., Ueda, K., Sato, F., Forestier, C., Yazaki, K. Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. USA* **100**, 751-756 (2003).
57. Samanani, N., Park, S.U., Facchini, P.J. Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis. *Plant Cell* **17(3)**, 915-926 (2005).
58. Ashihara, H., Monteiro, A.M., Gillies, F.M., Crozier, A. Biosynthesis of caffeine in leaves of coffee. *Plant Physiol.* **111**, 747-753 (1996).
59. Bird, D.A., Franceschi, V.R., Facchini, P.J. A tale of three cell types, alkaloid biosynthesis is localized to sieve elements in opium poppy. *Plant Cell* **15**, 2626-2635 (2003).
60. Smallegange, R.C., van Loon, J.J.A., Blatt, S.E., Harvey, J.A., Agerbirk, N., Dicke, M. Flower vs. leaf feeding by *Pieris brassicae*, glucosinolate-rich flower tissues are preferred and sustain higher growth rate. *J. Chem. Ecol.* **33**, 1831-1844 (2007).
61. Bouacida, S., Koubaier, H.B.H., Snoussi, A., Fauconnier, M.L., Bouzouita, N. Glucosinolate profiles by HPLC-DAD, phenolic compositions and antioxidant activity of *Eruca vesicaria longirostris*, impact of plant part and origin. *Mediterr. J. Chem.* **5(5)**, 528-539 (2016).
62. Nartey, F., Brimer, L., Christensen, S.B. Proacaciberin, a cyanogenic glycoside from *Acacia sieberiana* var. *woodii*. *Phytochemistry* **20(6)**, 1311-1314 (1981).
63. Pierce, S., Negreiros, D., Cerabolini, B.E.L., Kattge, J., Díaz, S., Kleyer, M., Shipley, B., Wright, S.J., Soudzilovskaia, N.A., Onipchenko, V.G., van Bodegom, P.M., Frenette-Dussault, C., Weiher, E., Pinho, B.X., Cornelissen, J.H.C., Grime, J.P., Thompson, K., Hunt, R., Wilson, P.J., Buffa, G., Nyakunga, O.C., Reich, P.B., Caccianiga, M., Mangili, F., Ceriani, R.M., Luzzaro, A., Brusa, G., Siefert, A., Barbosa, N.P.U., Chapin III, F.S., Cornwell, W.K., Fang, J., Fernandes, G.W., Garnier, E., Le Stradic, S., Peñuelas, J., Melo, F.P.L., Slaviero, A., Tabarelli, M., Tampucci, D. A global method for calculating plant CSR ecological strategies applied across biomes worldwide. *Funct. Ecol.* **31(2)**, 444-457 (2017).
64. Shinbo, Y., Nakamura, Y., Altaf-Ul-Amin, M., Asahi, H., Kurokawa, K., Arita, M., Saito, K., Ohta, D., Shibata, D., Kanaya, S. KNApSACk, A Comprehensive Species-Metabolite

- Relationship Database in *Plant Metabolomics. Biotechnology in Agriculture and Forestry*, vol 57 (eds. Saito, K., Dixon, R.A., Willmitzer, L.) 165-181 (Springer, 2006).
65. Kotera, M., Tabei, Y., Yamanishi, Y., Moriya, Y., Tokimatsu, T., Kanehisa, M., Goto, S. KCF-S, KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics. *BMC Syst. Biol.* **7(sup.6)**, S2 (2013).
66. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **40**, D136-D143 (2012).
67. Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R. Terrestrial ecoregions of the world, a new map of life on Earth. *BioScience* **51**, 933-938 (2001).
68. Flores, O., Garnier, E., Wright, I.J., Reich, P.B., et al. An evolutionary perspective on leaf economics in vascular plants, phylogenetic patterns in LMA. *Ecol. Evol.* **4(14)**, 2799-2811 (2014).
69. Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., et al. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89-92 (2014).
70. Qian, H., Jin, Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* **9**, 233-239 (2016).
71. R Core Team. R, A language and environment for statistical computing. Vienna, Foundation for Statistical Computing. Retrieved from <https://www.rproject.org>. (2020).
72. Webb, C.O., Ackerly, D.D., Kembel, S.W. Phylocom, software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**, 2098-2100 (2008).
73. Ho, L.S.T., Ane, C. Package "PhyloLM"- Phylogenetic Linear Regression (2014).
74. Schmelz, E.A., Huffaker, A., Sims, J.W., Christensen, S.A., Lu, X., Okada, K., Peters, R.J. Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* **79**, 659-678 (2014).
75. Zhou, Y., Massonnet, M., Sanjak, J.S., Cantu, D., Gaut, B.S. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl. Acad. Sci. USA* **114(44)**, 11715-11720 (2017).
76. Strange, R.N., Scott, P.R. Plant disease, a threat to global food security. *Annu. Rev. Phytopathol.* **43**, 83-116 (2005).
77. Huot, B., Yao, J., Montgomery, B.L., He, S.Y. Growth-defense tradeoffs in plants, a balancing act to optimize fitness. *Mol. Plant* **7**, 1267-1287 (2014).
78. Turcotte, M.M., Turley, N.E., Johnson, M.T.J. The impact of domestication on resistance to two generalist herbivores across 29 independent domestication events. *New Phytol.* **204**, 671-681 (2014).

79. Selmar, D., Kleinwächter, M. Stress enhances the synthesis of secondary plant products, the impact of stress-related over-reduction on the accumulation of natural products. *Plant Cell Physiol.* **54(6)**, 817-826 (2013).
80. Cerabolini, B.E.L., Pierce, S., Verginella, A., Brusa, G., Ceriani, R.M., Armiraglio, S. Why are many anthropogenic agroecosystems particularly species-rich? *Plant Biosyst.* **150(3)**, 550-557 (2016).

Figure 1. Significant phylogenetically corrected correlations (α ; positive and negative) between the extent of C-selection (i.e., a predominantly competitor adaptive strategy) and the number of species producing secondary metabolites belonging to different phytochemical clusters, with the strength of phylogenetic relatedness between species (λ) indicated (1.000=highest relatedness, 0.000=lowest relatedness). Clusters (as detailed in Table S1) are: A). pyrrolizidine alkaloids (cluster 2), B). coumarin and furanocoumarin phenylpropanoids (cluster 25), C). linear monoterpenes (cluster 34), and D). galloyl derivatives and ellagitannins (cluster 81). The size of points is proportional to the number of compounds of each class produced per species.

Figure 2. Significant phylogenetically corrected correlations (α ; positive and negative) between the extent of S-selection and the number of species producing secondary metabolites belonging to different phytochemical clusters, with the strength of phylogenetic relatedness between species (λ) indicated (1.000=highest relatedness, 0.000=lowest relatedness). Clusters (as detailed in Table S1) are: A). terpenoids (cluster 38), B). cadinanes (cluster 39), and C). labdanes (cluster 46). The size of points is proportional to the number of compounds of each class produced per species.

Figure 3. Significant phylogenetically corrected correlations (α ; positive and negative) between the extent of R-selection and the number of species producing secondary metabolites belonging to different phytochemical clusters, with the strength of phylogenetic relatedness between species (λ) indicated (1.000=highest relatedness, 0.000=lowest relatedness). Clusters (as detailed in Table S1) are: A). flavanone flavonoids (cluster 14), and B). cyanogenic glucosides derived from valine or isoleucine (cluster 73). The size of points is proportional to the number of compounds of each class produced per species.

Table 1. Phylogenetic linear regression: significant correlations (α ; both positive and negative) between biomes (as defined by Olson *et al.*, 2001) and phytochemical clusters after phylogenetic correction (A) with an indication of biomes for which no correlations were evident (B).

(A). Biomes exhibiting significant correlations	Phytochemical clusters	α	p
Deserts and Xeric Shrublands	17 Coumestanes,	0.1448	0.003
	41 Terpenoids	0.5297	0.012
Flooded Grasslands Savannas	36 Iridoids	0.3168	0.046
Mangroves	13 Chalcones,	0.5977	0.004
	30 Stilbenoid dimers, stilbenoid oligomers and bis(bibenzyl)s	0.3527	<0.001
Montane Grasslands Shrublands	6 Caffeate derivative alkaloids,	0.2789	0.015
	11 Steroid alkaloids,	0.6325	0.001
	21 Lignans,	0.1699	0.002
	38 Terpenoids,	0.4541	0.027
	59 Carotenoids,	0.1925	0.004
	81 Galloyl derivatives and ellagitannins	0.0808	0.006
Temperate Broadleaf and Mixed Forests	12 Purine alkaloids,	-0.0520	0.009
	59 Carotenoids	-0.1667	0.041
Temperate Coniferous Forests	39 Cadinane terpenoids,	-0.0701	0.011
	41 Aphidicolane terpenoids,	-0.7319	<0.001
	73 Cyanogenic glucosides derived from valine or isoleucine	-0.0486	0.016
Temperate Grasslands Savannas and Shrublands	14 Flavanone flavonoids,	0.1748	0.033
	77 Glucosinolates group B	-0.1226	0.008
Tropical and Subtropical Coniferous Forests	17 Coumestanes	-0.2498	<0.001
Tropical and Subtropical Dry Broadleaf Forests	4 Isoquinoline alkaloids,	-0.3231	0.011
	17 Coumestanes,	0.2311	0.006
	71 Betacyanins and betaxanthins	0.0259	0.016
Tropical and Subtropical Grasslands Savannas and Shrublands	21 Lignans	0.1658	0.037
Tropical and Subtropical Moist Broadleaf Forests	1 Pyrrolidine alkaloids,	-0.1575	0.043
	36 Iridoids,	0.1862	0.032
	71 Betacyanins and betaxanthins	0.0166	0.024
(B). Biomes not correlated with any phytochemical cluster			
Boreal Forests Taiga	None		>0.05
Mediterranean Forests Woodlands Scrub	None		>0.05
Tundra	None		>0.05