Supplementary Information

A comprehensive toxic plant-phytotoxin (TPPT) database and its application to assess their aquatic micropollution potential

Barbara F. Günthardt[1,2], Juliane Hollender[2,3], Konrad Hungerbühler[4], Martin Scheringer[4,5], Thomas D. Bucheli[1*]

[1] Environmental Analytics, Agroscope, Reckenholzstrasse 191, 8046 Zürich, Switzerland

[2] Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Universitätsstrasse 16, 8092 Zürich, Switzerland

[3] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland

[4] Institute for Chemical and Bioengineering, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zürich, Switzerland

[5] Masaryk University, RECETOX, Kamenice 753/5, 625 00 Brno, Czech Republic

*Corresponding author

Telephone number: 041 44 377 73 42

E-mail address: thomas.bucheli@agroscope.admin.ch.

Table of contents

Supplementary methods

Data compilation

For the start, a set of predefined data sources was used to compile the plant information. Plant species in general and biological information is mainly based on the plant compendium from Lauber et al. that contains all wild growing and several agricultural plant species in Switzerland.[1] Non-domestic ornamental garden plants and additional agricultural plants were compiled from the two specialized books by Roth et al.[2] and Teuscher and Lindequist[3]. Plants toxic to husbandry animals were collected from the Clinical Toxicology (CliniTox) database (http://www.vetpharm.uzh.ch/perldocs/index_x.htm)[4]. Information about the invasive plant species in Switzerland were taken from the National Data and Information Center on the Swiss Flora (info flora) (https://www.infoflora.ch/en/)[5], whereof the prohibited invasive alien plant species according to the Swiss Ordinance on the Handling of Organisms in the Environment are specially labeled.[5] Furthermore plants forbidden in the European Union were added for completion.[6]

The starting point for the collection of the phytotoxin-toxic plant species relationships were again the specialized books by Roth et al.[2] and, in particular, Teuscher and Lindequist.[3] For confirmation of the major toxins, the Clinical Toxicology (CliniTox) database (http://www.vetpharm.uzh.ch/perldocs/index_x.htm)[4] and the compendium from the European Food Safety Authority (EFSA)[7] were considered. Furthermore, the metabolite-species relationship database KNApSAcK (http://kanaya.naist.jp/knapsack_jsp/top.html)[8] was queried. The collected toxin compositions were as far as possible further extended with information from scientific journals. Therefore, a literature search was performed for each SPM class and for each included plant species using the Web of Science.

Chemical toxin information, including structure data and compound identification, was assembled from the publicly available compound databases PubChem[9] and ChemSpider[10].

Database searches were performed in both databases based on the phytotoxin name and, if no entries were found, the chemical structure. The found structures were compared with the structure from the phytotoxin data source, if available including the stereochemistry. In cases where the possible structures or stereochemistry were different the structure from PubChem was preferred since there were overall more structures available in PubChem. The combined use of two databases minimizes the sensitivity for errors in the chemical structure identification.

Measured toxicity data

Toxicity data was collected from three sources: Teuscher and Lindequist[3], Roth et al.[2], and Duke and Williams[11], and for each toxicity endpoint the specific reference was included. Experimental toxicity endpoints were included for all common test animals and exposure routes. However, still only for a limited number of phytotoxins experimental data was available from these data sources making estimations unavoidable. A more in-depth collection of toxicity endpoints is not of primary importance at present and beyond the scope of the TPPT database. Table S1 summarizes the data fields in the experimental toxicity data table.

Table S1. Description of the individual data fields in the experimental toxicity data table of the toxic plants – phytotoxins (TPPT) database used to describe the median lethal dose (LD_{50}) test and exemplified for the phytotoxin ptaquiloside.

Data field	Content description	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
Test_animal	Animal used in the test	*not available*
Exposure_route	Exposure route applied in the test: p.o. peroral, i.v. intravenous, i.p. intraperitoneal, s.c. subcutaneous	*not available*
Dose	Median lethal dose (LD_{50}) value determined through the test (mg/kg)	*not available*
Data_source	Literature reference	*not available*

EPI Suite *in-silico* physicochemical properties estimations

Physicochemical properties of both anthropogenic micropollutants and natural toxins alike are decisive for their environmental exposure, however, for phytotoxins almost no experimental data is available and estimations are necessary. The most often used tool to estimate physicochemical properties is the U.S. EPA's EPI Suite (Estimations Programs Interface), which combines several Quantitative Structure-Activity Relationships (QSARs) to estimate physicochemical properties.[12] Therefore, EPI Suite™ Version 4.11 was used to calculate the distribution and degradation parameters. The database includes output parameters of several QSARs: KOWWIN, HENRYWIN, KOAWIN, WSKOWWIN, WATERNT, KOCWIN, BIOWIN (3 and 4 only), AOPWIN, and HYDROWIN. The applicability domain of EPI Suite covers organic, uncharged, low molecular weight (<1000 g/mol) chemicals, which does not apply for all phytotoxins. Consequently, an applicability specification is given for each phytotoxin. Detailed descriptions of all data fields are found in Table S2. For five phytotoxins no estimations were possible at all.

Table S2. Description of the individual data fields in the EPI Suite table of the toxic plants – phytotoxins (TPPT) database exemplified for the phytotoxin ptaquiloside.

Data field	Content description: parameter (QSAR information, unit)	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
Applicability	Specification about the EPI Suite applicability to the compound. Remarked are phytotoxins that are over 1000 g/mol or charged.	
KOWWIN_log_Kow	Logarithmic octanol-water partition coefficient log K_{ow} (KOWWIN v1.68, -)	-0.95
KOWWIN_log_Kow_experimental	Logarithmic octanol-water partition coefficient log K_{ow} (Epi Suite experimental database, -)	*not available*
HENRYWIN_Kh_bond_method	Henry's law constant K_H (HENRYWIN v3.20 (bond method), atm-m^3/mole)	7.7e-19
HENRYWIN_Kh_group_method	Henry's law constant K_H (HENRYWIN v3.20 (group method), atm-m^3/mole)	*not available*
HENRYWIN_log_Kaw	Logarithmic air-water partition coefficient log K_{aw} (HENRYWIN v3.20 (bond method), -)	-16.5
KOAWIN_log_Koa	Logarithmic octanol-air partition coefficient log K_{oa} (KOAWIN v1.10, -)	15.6

WSKOWWIN_Sw	Water solubility S_w (WSKOW v1.42 (calculated from log K_{ow}), mg/l)	20370
WATERNT_Sw	Water solubility S_w (WatSol v1.01 (calculated from fragments), mg/l)	1e+6
KOCWIN_Koc_MCI_method	Organic carbon-water partition coefficient K_{oc} (KOCWIN v2.00 (MCI method), l/kg)	74.3
KOCWIN_Koc_Kow_method	Organic carbon-water partition coefficient K_{oc} (KOCWIN v2.00 (K_{ow} method), l/kg)	0.39
BIOWIN3	Probability of Rapid Biodegradation (BIOWIN v4.10, Biowin3 (Ultimate Survey Model))	2.28
BIOWIN4	Probability of Rapid Biodegradation (BIOWIN v4.10, Biowin4 (Primary Survey Model))	3.29
BIOWIN_Ready_Biodegradability_Prediction	Statement about the ready biodegradability of a compound, either yes or no	NO
AOPWIN_kOH	Atmospheric oxidation: overall OH rate constant k_{OH} (AopWin v1.92, cm^3/molecule-sec)	1.72e-10
AOPWIN_kOH_cis	Atmospheric oxidation: overall OH rate constant k_{OH} for cis (AopWin v1.92, cm^3/molecule-sec)	*not differentiated*
AOPWIN_kOH_trans	Atmospheric oxidation: overall OH rate constant k_{OH} for trans (AopWin v1.92, cm^3/molecule-sec)	*not differentiated*
AOPWIN_half_life_OH	Atmospheric oxidation: hydroxyl radical reaction half-life $t_{1/2}$ (AopWin v1.92, days)	0.06
AOPWIN_half_life_OH_cis	Atmospheric oxidation: hydroxyl radical reaction half-life $t_{1/2}$ for cis (AopHalfWin v1.92, days)	*not differentiated*
AOPWIN_half_life_OH_trans	Atmospheric oxidation: hydroxyl radical reaction half-life $t_{1/2}$ for trans (AopWin v1.92, days)	*not differentiated*
AOPWIN_kozone	Atmospheric oxidation: overall ozone rate constant k_{ozone} (AopWin v1.92, cm^3/molecule-sec)	1.14e-17
AOPWIN_kozone_cis	Atmospheric oxidation: overall ozone rate constant k_{ozone} for cis (AopWin v1.92, cm^3/molecule-sec)	*not differentiated*
AOPWIN_kozone_trans	Atmospheric oxidation: overall ozone rate constant k_{ozone} for trans (AopWin v1.92, cm^3/molecule-sec)	*not differentiated*
AOPWIN_half_life_ozone	Atmospheric oxidation: ozone reaction half-life $t_{1/2}$ (AopWin v1.92, days)	1.01
AOPWIN_half_life_ozone_cis	Atmospheric oxidation: ozone reaction half-life $t_{1/2}$ for cis (AopWin v1.92, days)	*not differentiated*
AOPWIN_half_life_ozone_trans	Atmospheric oxidation: ozone reaction half-life $t_{1/2}$ for trans (AopWin v1.92, days)	*not differentiated*
HYDROWIN_kb	Aqueous base-catalyzed hydrolysis: total K_b (HYDROWIN v2.00, pH > 8, l/mol-sec)	*not available*
HYDROWIN_ka	Aqueous acid-catalyzed hydrolysis: total K_a (HYDROWIN v2.00, l/mol-sec)	*not available*
HYDROWIN_ka_cis	Aqueous acid-catalyzed hydrolysis: total K_a for cis (HYDROWIN v2.00, l/mol-sec)	*not available*
HYDROWIN_ka_trans	Aqueous acid-catalyzed hydrolysis: total K_a for trans (HYDROWIN v2.00, l/mol-sec)	*not available*
HYDROWIN_half_life_kb	Aqueous base-catalyzed hydrolysis: K_b half-life $t_{1/2}$ (HYDROWIN v2.00, pH 7, days)	*not available*
HYDROWIN_half_life_ka	Aqueous acid-catalyzed hydrolysis: K_a half-life $t_{1/2}$ (HYDROWIN v2.00, pH 7, days)	*not available*
HYDROWIN_half_life_ka_cis	Aqueous acid-catalyzed hydrolysis: K_a half-life $t_{1/2}$ for cis (HYDROWIN v2.00, pH 7, days)	*not available*
HYDROWIN_half_life_ka_trans	Aqueous acid-catalyzed hydrolysis: K_a half-life $t_{1/2}$ for trans (HYDROWIN v2.00, pH 7, days)	*not available*

ACD/Percepta *in-silico* physicochemical properties estimations

Additional physicochemical parameters not available through EPI Suite were predicted with ACD/Percepta, the ACD/Labs percepta predictor software (Advanced Chemistry Development Inc., 2016 (ACD/Lab)) that predicts properties based on statistical algorithms from measured data.[13] In earlier studies it was shown that ACD/Percepta predicts accurate pK_a values within 1 pK_a unit.[14] Here, pK_a values for acids and bases estimated from both the Classic pK_a module and the GALAS module were included in the database together with the confidence interval. Several other physicochemical properties were also included: number of hydrogen bond donors and acceptors, ring number, freely rotatable bond number, carbon ratio, nitrogen and oxygen ratio, density, molar volume, surface tension and polarizability (detailed descriptions of all data fields are found in Table S3). Similar to EPI Suite, the ACD/Percepta software treats all compounds as uncharged species, and following critical species are specified in the applicability data field.

Table S3. Description of the individual data fields in the ACD/Percepta table of the toxic plants – phytotoxins (TPPT) database exemplified for the phytotoxin ptaquiloside.

Data field	Content description	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
Applicability	Specification of the EPI Suite applicability to the compound. Remarked are phytotoxins that are over 1000 g/mol or charged.	
pKa_acid_classic	Acidic pK_a values calculated with the classic algorithm	12.85,13.57,13.77,14.48,14.84
pKa_acid_conf_limits_classic	Confidence limits for the acidic pK_a values calculated with the classic algorithm	0.70,0.70,0.70,0.10,0.70
pKa_acid_GALAS	Acidic pK_a values calculated with the GALAS algorithm	12.60,14.01,14.81,15.52
pKa_base_conf_limits_GALAS	Confidence limits for the acidic pK_a values calculated with the GALAS algorithm	1.00,0.90,0.90,0.90
pKa_base_classic	Basic pK_a values calculated with the classic algorithm	*not basic ionizable*
pKa_base_conf_limits_classic	Confidence limits for the basic pK_a values calculated with the classic algorithm	*not basic ionizable*
pKa_base_GALAS	Basic pK_a values calculated with the GALAS algorithm	*not basic ionizable*
pKa_base_conf_limits_GALAS	Confidence limits for the basic pK_a values calculated with the GALAS algorithm	*not basic ionizable*
Number_hydrogen_bond_donors	Number of hydrogen bond donors	5

Number_hydrogen_bond_acceptors	Number of hydrogen bond acceptors	8
Number_rotatable_bonds	Number of freely rotating bonds	3
Number_rings	Number of rings	4
C_ratio	Ratio of carbon to all other elements	0.71
NO_ratio	Ration of nitrogen and oxygen to all other elements	0.29
Surface_tension	Surface tension (dyne/cm)	70.6
Density	Density (g/cm^3)	1.44
Polarizability	Polarizability (*10^{-24} cm^3)	38.6
Molar_volume	Molar volume (cm^3)	276.9

ProTox *in-silico* rodent toxicity prediction

A hazard assessment, including toxicity, is one of the major parts in an environmental and human risk assessment. To characterize the toxicity, the plant toxicity is complemented by a more specific phytotoxin toxicity, as predicted by ProTox. ProTox is a free online tool (http://tox.charite.de/tox/) that predicts the rodent oral toxicity based on compound similarity and returns the median lethal dose (LD_{50}) and the toxicity class ranging from I to VI according the globally harmonized system of classification and labelling of chemicals (GHS).[15] The database includes for each phytotoxin the predicted LD_{50} in mg/kg, the predicted toxicity class as well as the average similarity and prediction accuracy as indicated on the ProTox homepage (detailed descriptions of all data fields are found in Table S4).

Table S4. Description of the individual data fields in the ProTox table of the toxic plants – phytotoxins (TPPT) database exemplified for the phytotoxin ptaquiloside.

Data field	Content description	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
Predicted_LD50	The predicted rodent oral toxicity from ProTox (mg/kg body weight) defined by the median lethal dose (LD_{50})	100
Predicted_toxicity_class	The toxicity class from the globally harmonized system of classification and labelling of chemicals (GHS) for the predicted toxicity: - Class I: fatal if swallowed ($LD_{50} \leq 5$ mg/kg) - Class II: fatal if swallowed ($5 < LD_{50} \leq 50$ mg/kg) - Class III: toxic if swallowed ($50 < LD_{50} \leq 300$ mg/kg)	3

	- Class IV: harmful if swallowed (300 < LD_{50} ≤ 2000 mg/kg) - Class V: may be harmful if swallowed (2000 < LD_{50} ≤ 5000 mg/kg) - Class VI: non-toxic (LD_{50} > 5000 mg/kg)	
Quality_average_similarity	ProTox calculates an average similarity comparing the similarity of the input compounds to compounds which have been shown to bind at a target.	73.29
Quality_prediction_accuracy	ProTox calculates a prediction accuracy depending on the similarity of the input compound to the compounds with known LD_{50} values.	69.26

ECOSAR *in-silico* aquatic toxicity prediction

For environmental risk assessments not only the mammalian toxicity but also the eco-toxicity should be considered. Therefore, the U.S. EPA's ecological structure activity relationships (ECOSAR) predictive model Version 2.0 was used, which predicts acute toxicity for three model organisms: fish, daphnia, and green algae.[16] The tool comprises regression models for more than 50 chemical classes using structural similarity and measured aquatic toxicity and has shown to be useful in screening chemicals.[17] The database includes estimations for fish 96 h median lethal concentrations (LC_{50}), daphnid 48 h LC_{50}, green algae half maximal effective concentration (EC50) (combined for 72 and 96 h), fish chronic value (ChV) (28–20 days), daphna ChV (14–21 days), and green algae ChV (combined for 72 h and 96 h). For each toxicity endpoint, only the lowest estimation for the most toxic compound class was included following a conservative approach. Also, the applicability domain of ECOSAR covers only organic, low molecular weight (MW<1000 g/mol) chemicals, which does not apply for all phytotoxins, and consequently an applicability specification is given for each phytotoxin. Detailed descriptions of all data fields are found in Table S5. For four phytotoxins no estimation for the exact structure was possible.

Table S5. Description of the individual data fields in the ECOSAR table of the toxic plants – phytotoxins (TPPT) database exemplified for the phytotoxin ptaquiloside.

Data field	Content description: parameter (QSAR information, unit)	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
Applicability	Specification of the EPI Suite applicability to the compound. Remarked are phytotoxins that are over 1000 g/mol or charged.	-
ECOSAR_class	Specification of the compound classes that are used in the predictions.	Ketone Alcohols
Fish_96h_LC50	Estimated median lethal concentration (LC_{50}) for fish and a duration of 96 h (ECOSAR v2.0, mg/l)	2.14E+03
Daphnia_48h_LC50	Estimated median lethal concentration (LC_{50}) for daphnia and a duration of 48h (ECOSAR v2.0, mg/l)	8.82E+02
GreenAlgae_96h_EC50	Estimated half maximal effective concentration (EC_{50}) for green algae and a combined duration of 72 h and 96 h (ECOSAR v2.0, mg/l)	1.72E+02
Fish_ChV	Estimated chronic value (ChV) for fish (ECOSAR v2.0, mg/l)	1.93E+02
Daphnia_ChV	Estimated chronic value (ChV) for daphnia (ECOSAR v2.0, mg/l)	7.84E+01
GreenAlgae_ChV	Estimated chronic value (ChV) for green algae (ECOSAR v2.0, mg/l)	7.04E+02

Phytotoxin occurrence analysis

As basis to approximate the phytotoxin occurrence, the Swiss frequency parameter from the database was used that is the percentage of 10 km^2 squares in Switzerland in which a plant species is found. First, a factor for the phytotoxin occurrence was calculated by summing up the frequencies of all plant species i (total number k) that contain the particular phytotoxin:

$$Factor\ for\ phytotoxin\ occurrence\ =\ \sum_{i=1}^{k} Swiss\ frequency_{plant\ species\ i} \qquad (1)$$

For the calculations, the Swiss frequency of rarely or occasionally introduced plants was set to 0.1 to also account for their low importance. This calculated factor is however only an approximation for the actual occurrence, because the frequency does not contain information about the plant population density. Therefore, the occurrence factor was transformed into more general occurrence categories: no occurrence (occurrence factor = 0), very low (0 < occurrence factor < 1), low (1 ≤ occurrence factor < 10), medium (10 ≤ occurrence factor < 100), and high (occurrence factor ≥ 100). The factor for the phytotoxin occurrence and the occurrence category

are both given individually for each phytotoxin in the analysis table of the TPPT database (Table S6). In addition, an asterisk was appended in the TPPT database to phytotoxins that are also produced by agricultural plants, which applies for 224 compounds.

PM analysis following Arp et al.

For the PM analysis of the phytotoxins in the database, the classification procedure from Arp et al. was adapted in a simplified manner.[18] First, the phytotoxins were classified by charge and ionizability following the procedure from Arp et al.[18], additionally we noted the number of ionizable functional groups. The pH dependence was evaluated using the estimated pK_a values from ACD/percepta (classic pK_a calculation algorithm).[13] Hence, all the phytotoxins were classified in one of the following categories: neutral (595), acidic ionizable ($pK_a < 12$, monoprotic: 224, diprotic: 67, triprotic: 32, polyprotic: 32) basic ionizable ($pK_a > 2$, monoprotic: 410, diprotic: 62), and amphoteric ionizable (120). The phytotoxin classification is included in the analysis table of the TPPT database (see Table S6).

For the persistence criteria, the same four persistency (P) scores were used as proposed by Arp et al.: P1 (freshwater half-life < 20 d), P2 (20 days < freshwater half-life < 40 days), P3 (40 days < freshwater half-life < 60 days), and P4 (60 days < freshwater half-life).[18] The classification was conducted based on two processes: aerobic biotransformation in water and hydrolysis, and the lowest half-life was taken for the categorization. Biotransformation half-lives were calculated from the EPI Suite's BIOWIN3 output using the correlation from Arnot et al.[19] Hydrolysis half-lives were only considered for the 385 phytotoxins with EPI Suite estimations. For phytotoxins with different hydrolysis half-life estimations for cis and trans forms or acidic and basic reactions the lower half-life was included to be on the safe side. For each phytotoxin, the relevant degradation process, the half-life, and the P score are included in the analysis table of the TPPT database (see Table S6).

Regarding the mobility criteria, again the same five mobility (M) scores were applied as in Arp et al.: M1 (log K_{oc} / log $D_{oc} \geq 4.5$), M2 ($3 \leq$ log K_{oc} / log $D_{oc} < 4.5$), M3 ($2 \leq$ log K_{oc} / log $D_{oc} < 3$), M4 ($1 \leq$ log K_{oc} / log $D_{oc} < 2$) and M5 (log K_{oc} / log $D_{oc} < 1$).[18] For neutral phytotoxins the classification relied on the K_{oc} estimated from EPI Suite (K_{ow} method). For ionizable phytotoxins, the pH dependent D_{oc} was calculated from the estimated K_{oc} and pK_a and used for the classification. Since phytotoxins often have several ionizable functional groups, we differentiated in the calculations between monoprotic, diprotic and triprotic acids and bases, as well as amphoteric phytotoxins. Formulas are given below:

Monoprotic acids: $D_{oc} = (1/(1+10^{(pH-pKa)})) * K_{oc}$ (2)

Diprotic acids: $D_{oc} = (1/(1+10^{(pH-pKa1)} * (1+10^{(pH-pKa2)}))) * K_{oc}$ (3)

Triprotic acids: $D_{oc} = (1/(1+10^{(pH-pKa1)} * (1+10^{(pH-pKa2)} * (1+10^{(pH-pKa3)})))) * K_{oc}$ (4)

(also used for polyprotic acids)

Monoprotic bases: $D_{oc} = (1 - 1/(1 + 10^{(pH-pKa)})) * K_{oc}$ (5)

$= (1/(1+10^{(pKa-pH)})) * K_{oc}$ (6)

Amphoteric : $D_{oc} = (1/(1 + 10^{(pH-pKa,acid)} + 10^{(pKa,base-pH)})) * K_{oc}$ (7)

Polyprotic phytotoxins were treated as triprotic and for amphoteric phytotoxins only the strongest acid and base were included. A first analysis showed that the inclusion of further possible ionizations would not change the classification in over 95% and the Doc calculation was therefore simplified. The environmentally possible pH range was assumed to be between 4 and 10 and included the lowest calculated Doc, which is for bases at pH 4 and for acids at pH 10. For each phytotoxin, the log K_{oc} or log D_{oc} and the M score are included in the analysis table of the TPPT database (see Table S6).

In an overall scoring the P score and M score were combined analogous to Arp et al.[18] Transient phytotoxins are neither persistent nor mobile (P1/M1-M4, and P2/M1-M2), and as such of no concern for the water quality. This also holds for immobile P (P3-P4/M1) and unstable M

(P1/M5). Compounds of possible relevance for the water quality are ranked from 1 to 5 with increasing importance (PM1: P2/M3, PM2: P2/M4, PM3: P2/M5 and P3/M2, PM4: P3/M3 and P4/M2-M3, PM4.5: P3-P4/M5, PM5: P3-P4/M5). For each phytotoxin, the final PM score is included in the analysis table of the TPPT database (see Table S6).

Toxicity analysis

The toxicity analysis takes into account two different acute toxicities: the rodent toxicity to include mammalians and the aquatic toxicity based on fish, daphnia, and green algae to approximate aquatic toxicity. For the mammalian toxicity, experimental oral toxicity data was taken where available and if not available complemented with ProTox LD_{50} estimations. If several experimental oral values were available the smallest (i.e. the most toxic) was chosen. The values were then transformed into the GHS classes for acute human toxicity: 1 ($LD_{50} \leq 5$), 2 ($5 < LD_{50} \leq 50$), 3 ($50 < LD_{50} \leq 300$), 4 ($300 < LD_{50} \leq 2000$), 5 ($2000 < LD_{50} \leq 5000$), and nontoxic ($LD_{50} > 5000$). For the aquatic toxicity, only the ECOSAR estimations for toxicity were considered since no experimental data was available. The estimations for fish LC_{50}, daphnia LC_{50}, and green algae EC_{50} were compared and the smallest (i.e. the most toxic) was further used. The values were then transformed into the GHS classes for substances acute hazardous to the aquatic environment: 1 ($LC_{50}/EC_{50} \leq 1$), 2 ($1 < LC_{50}/EC_{50} \leq 10$), 3 ($10 < LC_{50}/EC_{50} \leq 100$), and nontoxic ($LC_{50}/EC_{50} > 100$). Here we included only the acute toxicity due to simplicity, but for a more in depth risk assessment also estrogenic and mutagenic effects on humans and chronic aquatic toxicity need to be considered. However, those are much more difficult to predict. For each phytotoxin, the LD_{50} source, the LD_{50}, the human toxicity GHS classes, the relevant aquatic toxicity test, the LC_{50} or EC_{50}, and the aquatic toxicity GHS classes are included in the analysis table of the TPPT database (see Table S6).

Sensitivity and uncertainty analysis

First, a sensitivity analysis was completed to evaluate the influence of the different parameters, because many unknowns and estimations were included in the whole procedure. The common approach also applied by Strempel et al was used.[20] Specifically, the phytotoxin occurrence, persistence, mobility and both toxicities were varied by a factor of 2, and the changes in the prioritization was assessed.

For the uncertainty analysis the uncertainty factors from Strempel et al. were adapted and the minimal and maximal number of prioritized phytotoxins determined.[20] Strempel et al. derived uncertainty factors by plotting estimated against measured data and deriving the scatter in those plots. This resulted in the following factors: 4 for half-life, 3.5 for K_{ow}, 100 for chronic aquatic toxicity, and 45 for acute aquatic toxicity. For half-lives and acute aquatic toxicities, the same factors were applied since both rely on the same estimation methods (EPI Suite and ECOSAR, respectively). The log K_{oc} and log D_{oc} were calculated based on the pK_a values, which were shown to be within one order of magnitude, [14] and the K_{oc}, which is often assumed to be proportional to the K_{ow}. Accounting for these additional uncertainties, the factors for the K_{ow} and the D_{ow} were set to 7 corresponding to twice the K_{ow} factor. For the acute rodent toxicity also a factor of 45 was applied analogous to the acute aquatic toxicity. Regarding the phytotoxin occurrence no information about the uncertainty was available, and therefore the same factor 2 as in the sensitivity analysis was used. These derived uncertainty factors were applied in the prioritization procedure, and the maximal and minimal numbers of prioritized phytotoxins determined.

Final aquatic micropollution potential analysis table

Table S6. Description of the individual data fields in the "Aquatic micropollution potential analysis" table of the toxic plants – phytotoxins (TPPT) database exemplified for the phytotoxin ptaquiloside.

Data field	Content description	Example
Phytotoxin_number	Continuous numbering of the phytotoxins labeled with a T	T1433
Phytotoxin_name	Most common English name, important alternative names are given in brackets	Ptaquiloside (Braxin C)
PSM_class	Plant secondary metabolite (PSM) class to which the phytotoxin belongs. Here only the major class is given.	Sesquiterpene
Substance_classification	Classification by charge and ionizability following the procedure from Arp et al.[18]	neutral
Koc/Doc_calculation_method	Definition of the property that is used for the scoring (K_{oc} or D_{oc}), and for the D_{oc} the calculation method is specified.	Koc
log(Koc)/log(Doc)	Log K_{oc} or log D_{oc} value used in the mobility scoring.	-0.41
Mobility_score	Five mobility scores (M-scores) were applied as in Arp et al.: M1 (log K_{oc} / log $D_{oc} \geq 4.5$), M2 ($3 \leq$ log K_{oc} / log $D_{oc} < 4.5$), M3 ($2 \leq$ log K_{oc} / log $D_{oc} < 3$), M4 ($1 \leq$ log K_{oc} / log $D_{oc} < 2$) and M5 (log K_{oc} / log $D_{oc} < 1$).[18]	M5
Degradation_process	Specification of the fastest removal process. Included are biodegradation, hydrolysis and volatilization.	Biodegradation
Half-life	Half-life (days) used in the persistence scoring.	47.64
Persistence_score	Four persistency scores (P-scores) were used as proposed by Arp et al.: P1 (freshwater half-life < 20 days), P2 (20 days < freshwater half-life < 40 days), P3 (40 days < freshwater half-life < 60 days), and P4 (60 days < freshwater half-life).[18]	P3
PM_score	PM-scoring as proposed by Arp et al.: transient (P1/M1-M4, and P2/M1-M2), immobile Ps (P3-P4/M1), unstable Ms (P1/M5), and prioritized PM compounds (1: P2/M3, 2: P2/M4, 3: P2/M5 and P3/M2, 4: P3/M3 and P4/M2-M3, 4.5: P3-P4/M5, 5: P3-P4/M5).	5
Ecotoxicity_test	Exotoxicity test with the smallest half-life from ECOSAR predictions. Included are fish LC_{50}, daphnia LC_{50}, and green algae EC_{50}.	Green algae 96h EC50
LC50/EC50	LC_{50} or EC_{50} concentration used in the GHS eco-toxicity categorization (mg/l).	172.03
GHS_acute_ecotoxicity_class	GHS categories for substances acute hazardous to the aquatic environment: 1 ($LC_{50}/EC_{50} \leq 1$), 2 ($1 < LC_{50}/EC_{50} \leq 10$), 3 ($10 < LC_{50}/EC_{50} \leq 100$), and nontoxic ($LC_{50}/EC_{50} >100$).	not eco-toxic
Rodent_toxicity_test	Specification, if the used LD_{50} value was predicted by ProTox or if the used LD_{50} is a measured value from the TPPT database.	predicted LD50 (ProTox)
LD50	LD_{50} value used in the GHs acute toxicity categorization (mg/kg).	100
GHS_acute_toxicity_class	GHS categories for acute toxicity: 1 ($LD_{50} < 5$), 2 ($5 < LD_{50} < 50$), 3 ($50 < LD_{50} < 300$), 4 ($300 < LD_{50} < 2000$), 5 ($2000 < LD_{50} < 5000$), and nontoxic ($LD_{50} > 5000$).	3
Phytotoxin_occurrence	Factor for phytotoxin occurrence (equation (1)).	52
Occurrence_category	Occurrence category: no occurrence (occurrence factor = 0), very low (0 < occurrence factor < 1), low (1 < occurrence factor < 10), medium (10 < occurrence factor < 100), and high (occurrence factor > 100).	medium occurrence
Priority	Classification in the prioritization procedure	priority

Supplementary results

Distribution of different physicochemical properties among all phytotoxins

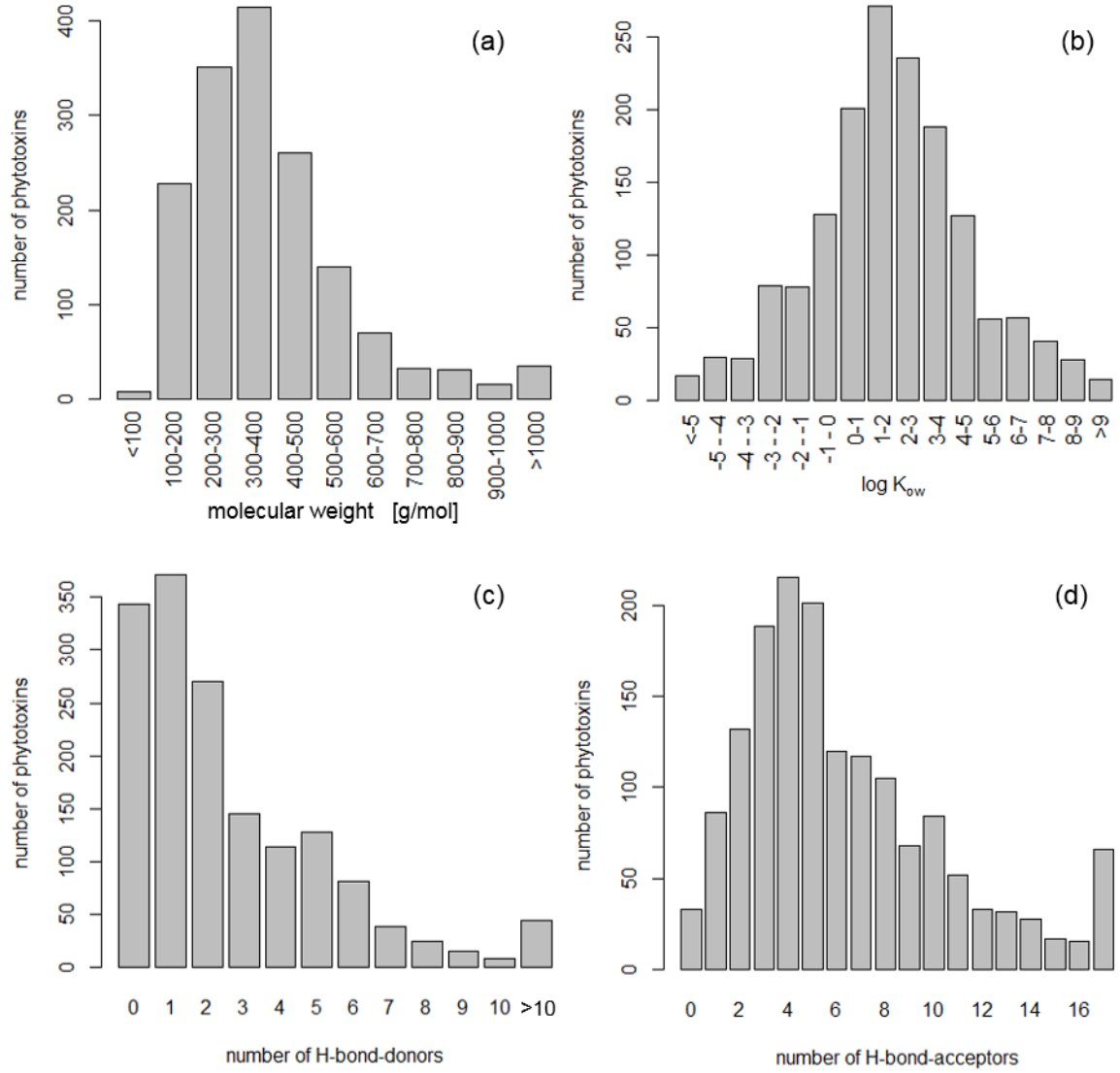


Figure S1. Distribution of different physicochemical properties among all phytotoxins: (a) molecular weight (MW), (b) octanol - water partition coefficient (log K_{ow}), (c) number of hydrogen-bond-donors, and (d) number of hydrogen-bond-acceptors.

Distribution of phytotoxins in the toxicity categories

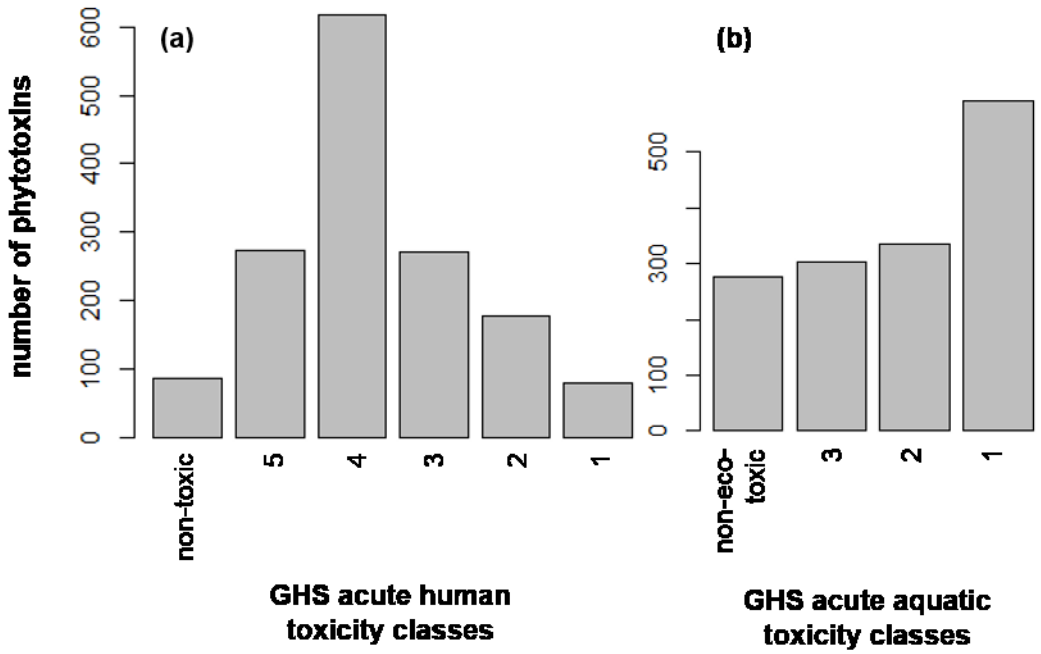


Figure S2. (a) Distribution of all phytotoxins in the globally harmonized system of classification and labelling of chemicals (GHS) classes for acute human toxicity (increasing toxicity from 5 to 1), and (b) distribution of all phytotoxins in the GHS classes for substances acute hazardous to the aquatic environment (increasing toxicity from 3 to 1).

Comparison of acute rodent toxicity and acute aquatic toxicity

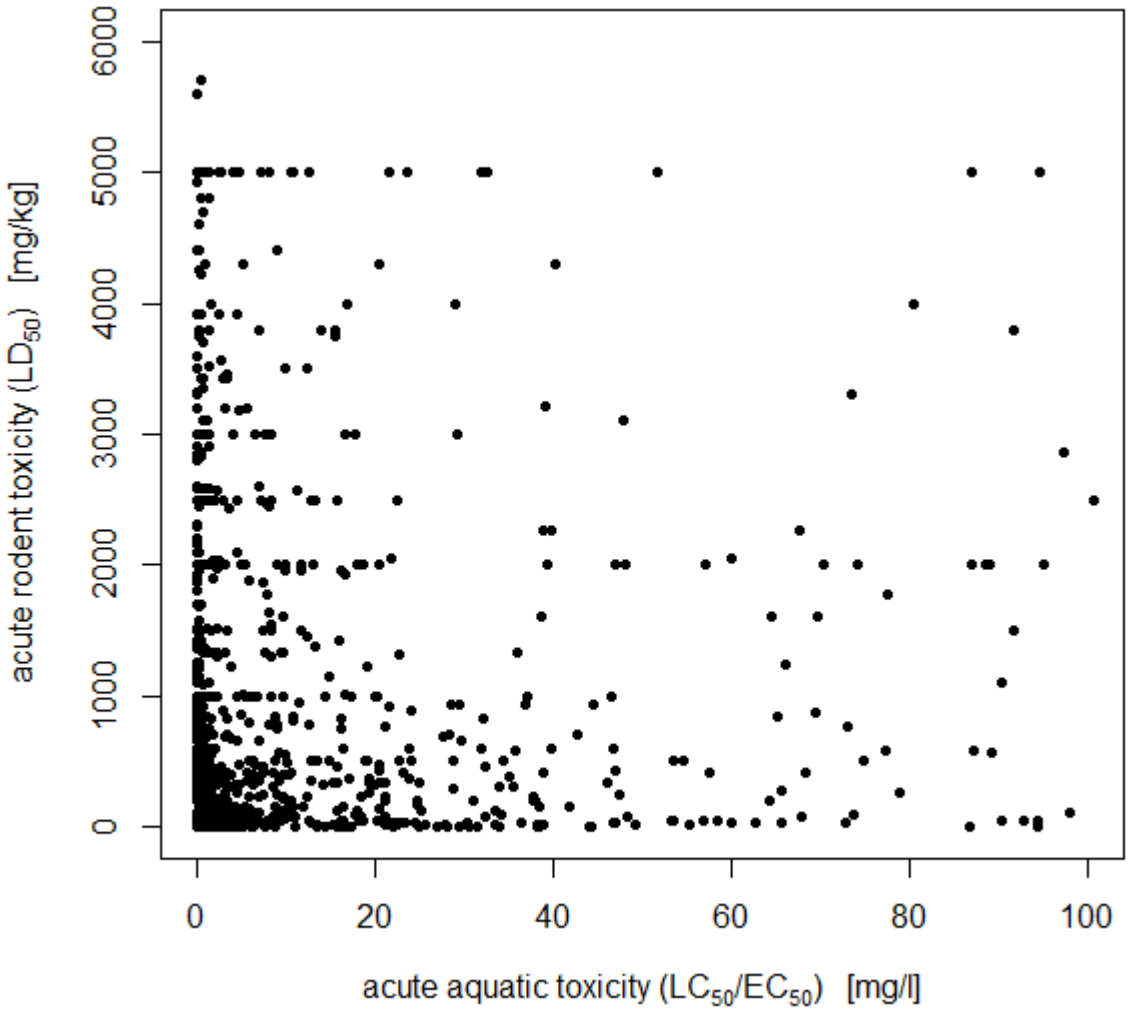


Figure S3. Scatter plot showing the (non-existing) relation between the predicted acute rodent toxicities (median lethal dose (LD_{50})) versus the predicted acute aquatic toxicities (median lethal concentration (LC_{50}) or half maximal effective concentration (EC_{50})).

Phytotoxin occurrence analysis

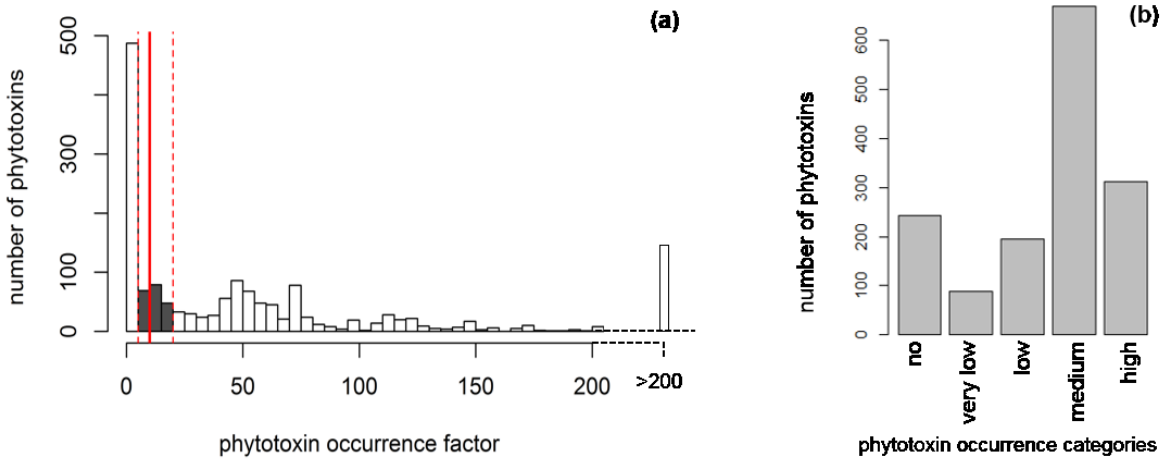


Figure S4. (a) Distribution of all phytotoxins in the primary phytotoxin occurrence factor, and (b) distribution of all phytotoxins in the secondary occurrence categories.

The phytotoxin occurrence was assessed based on the occurrence of their producing plants using an occurrence factor (Figure S4a) and a more general occurrence category (Figure S4b). Of all phytotoxins included, almost 16% are not produced by naturally growing plants in Switzerland, but rather by agricultural crops or garden plants. The other phytotoxins are distributed as follows: 6% have a very low occurrence, 13% a low, 44% a medium, and 21% have a high occurrence (SI Figure S4b). Therefore, at least 65% of all included phytotoxins, namely those from plants with medium and high occurrence, should be evaluated in detail. This high number seems to contrast the often low distribution of the toxic plants themselves, as discussed above. However, the phytotoxin occurrence integrates over several plant species, and the numbers are not directly comparable anymore. Furthermore the phytotoxin occurrence factor is only semi-quantitative and should be interpreted with caution. Therefore, it is not reasonable to put too much weight on the absolute numbers. Here, we only use this factor to set a limit in the preliminary prioritization procedure. In terms of prevailing PSM classes, no clear picture is visible, and almost all PSM classes contain some phytotoxins with medium or high occurrence.

Sensitivity and uncertainty analyses

Table S7. Sensitivity analysis of the prioritization procedure: The phytotoxin occurrence, the mobility, the degradation half-lives, the aquatic and the acute human toxicity were individually divided and multiplied by 2, and with the adapted values the same prioritization was performed. The degradation and the mobility depend on each other as described in the method and have therefore two limits (see method description for details).

Adapted factor	Threshold	Number of prioritized phytotoxins if multiplied by 2	Number of prioritized phytotoxins if divided by 2
Phytotoxin occurrence (occurrence factor)	≥ 10	562 (+46)	443 (-73)
Degradation ($t_{1/2}$ in days)	> 20 or > 40	639 (+123)	404 (-112)
Mobility (K_{oc} or D_{oc})	< 3 or < 4	507 (-9)	526 (+10)
Acute rodent toxicity (LD_{50} in mg/kg body weight)	≤ 300	497 (-19)	546 (+30)
Acute aquatic toxicity (LC_{50}/EC_{50} in mg/l)	≤ 10	489 (-27)	541 (+25)

Table S8. Uncertainty analysis based on Strempel et al.[20] of the persistence and mobility (PM) analysis following Arp et al.[18] The regular distribution results by using the directly estimated half-life and K_{oc}. These two properties are then shifted with the maximal uncertainty to the less critical end and to the most critical end, and then the same PM analysis was performed.

PM categories from Arp et al.	Regular distribution	Distribution with less critical properties	Distribution with more critical properties
transient	218	421	199
immobile P	12	17	19
unstable M	325	548	41
PM1	39	23	3
PM2	22	35	18
PM3	145	154	138
PM4	61	50	46
PM4.5	77	39	18
PM5	607	219	1024

Characterization of PSM classes

Table S9. Distribution of all phytotoxins within a plant secondary metabolite (PSM) class to the phytotoxin occurrence categories (no occurrence, very low occurrence, low occurrence, medium occurrence, and high occurrence), to the mobility (M) scores (from 1 to 5), to the persistence (P) scores (from 1 to 4), to the PM scores (transient, unstable but mobile, immobile but persistent, and PM from 1 to 5), to the GHS (globally harmonized system of classification and labelling of chemicals) acute aquatic toxicity classes (non-eco-toxic and eco-toxic from 3 to 1), and to the GHS acute human toxicity classes (non-toxic and toxic from 5 to 1). Bold marked categories are those prioritized (not indicated for the P and M scores, since only the PM score is relevant for prioritization). Integers in each cell indicate numbers of phytotoxins in individual categories.

PSM class (n)	Phytotoxin occurrence	P score	M score	PM score	GHS acute human toxicity classes	GHS acute aquatic toxicity classes
Acridine alkaloids (4)	low: 4	P2: 1 P3: 1 P4: 2	M4: 1 M5: 3	**PM3: 1** **PM4.5: 1** **PM5: 2**	GHS4: 2 **GHS3: 2**	**GHS2: 3** **GHS1: 1**
Aliphatic acids (15)	**middle: 4** **high: 11**	P1: 14 P2: 1	M2: 1 M4: 1 M5: 13	transient: 2 unstable M: 12 **PM3: 1**	non-toxic: 7 GHS5: 3 GHS4: 2 **GHS2: 1** **GHS1: 2**	non-eco-toxic: 7 GHS3: 4 **GHS2: 3** **GHS1: 1**
Amaryllidaceae alkaloids (23)	very low: 1 **middle: 20** **high: 2**	P1: 1 P2: 7 P3: 9 P4: 6	M4: 1 M5: 22	unstable M: 1 **PM2: 1** **PM3: 6** **PM5: 15**	GHS5: 1 GHS4: 4 **GHS3: 13** **GHS2: 5**	GHS3: 3 **GHS2: 5** **GHS1: 15**
Amines (28)	no: 3 very low: 11 low: 9 **middle: 2** **high: 3**	P1: 18 P2: 10	M2: 5 M3: 5 M5: 18	transient: 8 unstable M: 15 **PM1: 2** **PM3: 3**	GHS5: 2 GHS4: 15 **GHS3: 8** **GHS2: 3**	GHS3: 10 **GHS2: 5** **GHS1: 13**
Cyanogenic glycosides (16)	no: 4 **high: 12**	P1: 16	M5: 16	unstable M: 16	non-toxic: 5 GHS 4: 9 **GHS3: 2**	non-eco-toxic: 16

Diterpenes (83)	no: 45 very low: 1 low: 4 **middle: 18** **high: 15**	P1: 3 P2: 4 P3: 8 P4: 68	M1: 4 M2: 8 M3: 10 M4: 11 M5: 50	unstable M: 3 immobile P: 4 **PM1: 2** **PM3: 3** **PM4: 15** **PM4.5: 11** **PM5: 45**	non-toxic: 24 GHS5: 14 GHS4: 21 **GHS3: 8** **GHS2: 9** **GHS1: 7**	non-eco-toxic: 42 GHS3: 3 **GHS2: 18** **GHS1: 20**
Glucosinolates (32)	no: 5 very low: 5 low: 4 **middle: 14** **high: 3**	P1: 31	M5: 31	unstable M: 31	GHS4: 9 **GHS3: 1** **GHS2: 21**	non-eco-toxic: 31
Indole alkaloids (72)	no: 35 very low: 6 low: 1 **middle: 30**	P1: 2 P2: 9 P3: 6 P4: 55	M2: 2 M4: 2 M5: 68	transient: 1 unstable M: 2 **PM3: 8** **PM4: 1** **PM4.5: 2** **PM5: 58**	GHS5: 2 GHS4: 45 **GHS3: 12** **GHS2: 6** **GHS1: 7**	GHS3: 12 **GHS2: 19** **GHS1: 41**
Indolizidine alkaloids (1)	**middle: 1**	P4: 1	M5: 1	**PM5: 1**	**GHS3: 1**	**GHS1: 1**
Isoquinoline alkaloids (89)	no: 14 very low: 3 low: 27 **middle: 37** **high: 18**	P1: 3 P2: 1 P3: 6 P4: 79	M2: 2 M3: 2 M4: 10 M5: 75	unstable M: 3 **PM3: 1** **PM4: 4** **PM4.5: 10** **PM5: 71**	non-toxic: 1 GHS5: 1 GHS4: 73 **GHS3: 13** **GHS2: 1**	GHS3: 6 **GHS2: 9** **GHS1: 74**
Monoterpenes (118)	no: 10 very low: 24 low: 3 **middle: 27** **high: 54**	P1: 85 P2: 23 P3: 4 P4: 6	M1: 3 M2: 36 M3: 40 M4: 5 M5: 34	transient: 68 unstable M: 23 **PM1: 13** **PM3: 5** **PM4.5: 2** **PM5: 7**	non-toxic: 9 GHS5: 44 GHS4: 61 **GHS3: 3** **GHS2: 1**	non-eco-toxic: 21 GHS3: 16 **GHS2: 24** **GHS1: 57**

Naphthalene- and anthracene-derivatives (35)	no: 2 **middle: 9** **high: 24**	P1: 18 P2: 8 P3: 3 P4: 6	M3: 1 M4: 2 M5: 32	unstable M: 18 **PM2: 1** **PM3: 7** **PM4: 1** **PM4.5: 1** **PM5: 7**	non-toxic: 4 GHS5: 27 GHS4: 4	non-eco-toxic: 7 GHS3: 2 **GHS2: 10** **GHS1: 16**
Nitriles (4)	no: 3 **middle: 1**	P1: 4	M5: 4	unstable M: 4	GHS5: 2 GHS4: 1 **GHS3: 1**	non-eco-toxic: 2 GHS3: 2
Non-protein amino acids (21)	no: 3 **middle: 16** **high: 2**	P1: 21	M5: 21	unstable M: 1	non-toxic: 6 GHS5: 4 GHS4: 9 **GHS3: 2**	non-eco-toxic: 18 GHS3: 2 **GHS1: 1**
Phenylpropanoids (74)	no: 6 very low: 3 low: 17 **middle: 30** **high: 18**	P1: 41 P2: 23 P3: 4 P4: 6	M2: 12 M3: 24 M4: 12 M5: 26	transient: 28 unstable M: 19 **PM1: 9** **PM2: 1** **PM3: 9** **PM4: 3** **PM4.5: 5**	non-toxic: 3 GHS5: 24 GHS4: 40 **GHS3: 6** **GHS1: 1**	non-eco-toxic: 8 GHS3: 22 **GHS2: 15** **GHS1: 29**
Piperidine alkaloids (38)	very low: 2 low: 7 **middle: 7** **high: 22**	P1: 22 P2: 7 P3: 4 P4: 5	M5: 38	unstable M: 22 **PM3: 7** **PM5: 9**	GHS5: 3 GHS4: 22 **GHS3: 10** **GHS2: 2** **GHS1: 1**	non-eco-toxic: 1 GHS3: 15 **GHS2: 15** **GHS1: 7**
Polyacetylenes (23)	no: 2 low: 12 **middle: 6** **high: 3**	P1: 23	M1: 1 M2: 8 M3: 14	transient: 23	non-toxic: 2 GHS5: 9 GHS4: 7 **GHS3: 2** **GHS2: 2** **GHS1: 1**	**GHS2: 1** **GHS1: 22**

Polyketides (110)	no: 31 very low: 14 low: 1 **middle: 26** **high: 38**	P1: 59 P2: 30 P3: 9 P4: 12	M1: 16 M2: 6 M3: 2 M4: 5 M5: 81	transient: 19 unstable M: 44 immobile P: 1 **PM1: 1** **PM3: 26** **PM4: 3** **PM4.5: 4** **PM5: 12**	non-toxic: 7 GHS5: 44 GHS4: 41 **GHS3: 14** **GHS1: 4**	non-eco-toxic: 9 GHS3: 17 **GHS2: 17** **GHS1: 67**
Purine alkaloids (6)	no: 3 very low: 1 **middle: 2**	P1: 6	M5: 6	unstable M: 6	GHS4: 5 **GHS3: 1**	GHS3: 1 **GHS1: 5**
Pyridine alkaloids (14)	no: 10 **middle: 2** **high: 2**	P1: 9 P2: 2 P3: 2 P4: 1	M3: 1 M4: 1 M5: 12	transient: 1 unstable M: 8 **PM3: 2** **PM4: 1** **PM5: 2**	non-toxic: 1 GHS5: 1 GHS4: 5 **GHS2: 5** **GHS1: 2**	non-eco-toxic: 3 GHS3: 3 **GHS2: 1** **GHS1: 7**
Pyrimidine alkaloids (4)	high: 4	P1: 4	M5: 4	unstable M: 4	GHS5: 1 GHS4: 3	non-eco-toxic: 2 **GHS1: 2**
Pyrrolidine alkaloids (5)	no: 1 low: 1 **high: 3**	P1: 3 P2: 1 P4: 1	M5: 5	unstable M: 3 **PM3: 1** **PM5: 1**	GHS4: 3 **GHS3: 2**	GHS3: 5
Pyrrolizidine alkaloids (65)	very low: 1 low: 8 **middle: 31** **high: 25**	P1: 8 P2: 20 P3: 18 P4: 19	M5: 65	unstable M: 8 **PM3: 20** **PM5: 37**	GHS5: 3 GHS4: 6 **GHS3: 31** **GHS2: 25**	non-eco-toxic: 4 GHS3: 24 **GHS2: 28** **GHS1: 9**
Quinazoline alkaloids (5)	no: 1 very low: 1 low: 3	P1: 1 P2: 3 P4: 1	M3: 1 M4: 2 M5: 2	unstable M: 1 **PM1: 1** **PM2: 2** **PM5: 1**	GHS4: 3 **GHS3: 2**	GHS3: 2 **GHS2: 2** **GHS1: 1**

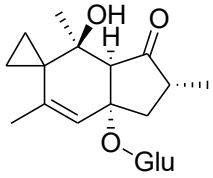
Quinoline alkaloids (13)	no: 2 low: 11	P1: 2 P2: 8 P3: 1 P4: 2	M2: 2 M3: 1 M4: 2 M5: 8	unstable M: 2 **PM1: 1** **PM2: 2** **PM3: 6** **PM4: 1** **PM5: 1**	GHS4: 12 **GHS3: 1** **GHS2: 1**	non-eco-toxic: 1 GHS3: 7 **GHS2: 3** **GHS1: 3**
Quinolizidine alkaloids (52)	no: 2 very low: 2 low: 2 **middle: 40** **high: 6**	P1: 4 P2: 14 P3: 21 P4: 13	M4:2 M5: 50	unstable M: 4 **PM2: 2** **PM3: 12** **PM5: 34**	GHS5: 2 GHS4: 28 **GHS3: 20** **GHS1: 2**	non-eco-toxic: 2 GHS3: 10 **GHS2: 26** **GHS1: 14**
Quinones (4)	**middle: 4**	P1: 3 P2: 1	M4: 1 M5: 3	transient: 1 unstable M: 2 **PM3: 1**	GHS5: 2 GHS4: 1 **GHS3: 1**	non-eco-toxic: 2 **GHS1: 2**
Saponins (42)	very low: 2 low: 19 **middle: 14** **high: 7**	P1: 2 P3: 2 P4: 38	M2: 8 M3: 3 M4: 3 M5: 28	unstable M: 2 **PM4: 11** **PM4.5: 3** **PM5: 26**	non-toxic: 5 GHS5: 12 GHS4: 17 **GHS2: 8**	non-eco-toxic: 12 GHS3: 8 **GHS2: 9** **GHS1: 13**
Sesquiterpenes (122)	no: 18 very low: 8 low: 9 **middle: 71** **high: 16**	P1: 65 P2: 38 P3: 14 P4: 5	M1: 3 M2: 12 M3: 25 M4: 36 M5: 46	transient: 45 unstable M: 28 **PM1: 10** **PM2: 11** **PM3: 10** **PM4: 3** **PM4.5: 6** **PM5: 9**	non-toxic: 5 GHS5: 28 GHS4: 52 **GHS3: 33** **GHS2: 3** **GHS1: 1**	non-eco-toxic: 4 GHS3: 24 **GHS2: 53** **GHS1: 41**
Steroids (115)	no: 16 low: 27 **middle: 61** **high: 11**	P1: 3 P2: 2 P3: 6 P4: 104	M2: 1 M3: 5 M4: 14 M5: 95	transient: 1 unstable M: 2 **PM2: 2** **PM4: 6** **PM4.5: 11** **PM5: 93**	non-toxic: 1 GHS5: 4 GHS4: 18 **GHS3: 6** **GHS2: 46** **GHS1: 40**	non-eco-toxic: 33 GHS3: 39 **GHS2: 34** **GHS1: 9**

Steroidal alkaloids (81)	no: 11 very low: 3 low: 18 **middle: 49**	P3: 2 P4: 79	M3: 1 M4: 2 M5: 78	**PM3: 2** **PM4: 1** **PM5: 78**	non-toxic: 1 GHS5: 5 GHS4: 40 **GHS3: 20** **GHS2: 10** **GHS1: 5**	non-eco-toxic: 4 GHS3: 12 **GHS2: 6** **GHS1: 59**
Sulfur compounds (21)	**middle: 21**	P1: 21	M2: 5 M3: 8 M4: 7 M5: 1	transient: 20 unstable M: 1	GHS5: 3 GHS4: 7 **GHS3: 11**	non-eco-toxic: 8 GHS3: 8 **GHS2: 5**
Terpenoid alkaloids (77)	no: 5 very low: 1 **middle: 56** **high: 15**	P1: 2 P2: 5 P4: 70	M4: 6 M5: 71	unstable M: 2 **PM3: 5** **PM4.5: 6** **PM5: 64**	GHS5: 7 GHS4: 20 **GHS3: 28** **GHS2: 17** **GHS1: 5**	non-eco-toxic: 22 GHS3: 24 **GHS2: 4** **GHS1: 27**
Tetraterpenes (4)	**middle: 4**	P1: 4	M3: 1 M5: 3	transient: 1 unstable M: 3	non-toxic: 1 GHS5: 2 **GHS3: 1**	non-eco-toxic: 1 **GHS1: 3**
Triterpenes (56)	no: 10 low: 4 **middle: 40** **high: 2**	P3: 1 P4: 55	M1: 7 M2: 3 M3: 7 M4: 10 M5: 29	immobile P: 7 **PM4: 10** **PM4.5: 10** **PM5: 29**	non-toxic: 5 GHS5: 13 GHS4: 20 **GHS3: 9** **GHS2: 7** **GHS1: 2**	non-eco-toxic: 6 GHS3: 9 **GHS2: 11** **GHS1: 30**
Tropane alkaloids (27)	low: 4 **middle: 17** **high: 6**	P1: 15 P2: 11 P4: 1	M5: 27	unstable M: 15 **PM3: 11** **PM5: 1**	GHS5: 10 GHS4: 10 **GHS3: 7**	non-eco-toxic: 10 GHS3: 7 **GHS2: 8** **GHS1: 2**
Tropolone alkaloids (8)	**middle: 8**	P3: 1 P4: 7	M3: 1 M4: 3 M5: 4	**PM4: 1** **PM4.5: 3** **PM5: 4**	GHS4: 2 **GHS3: 1** **GHS2: 5**	non-eco-toxic: 1 GHS3: 5 **GHS2: 2**

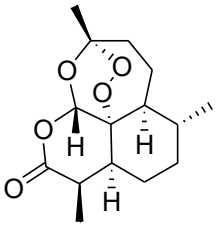
Table S10. Ranking of all plant secondary metabolite (PSM) classes according to the number of priority phytotoxins that were identified by the procedure described in the text. Additionally, the number of phytotoxins removed in each prioritization step (Figure 5, main manuscript) is given.

PSM Class	Total included phytotoxins	Prioritized phytotoxins	Removed in step 1	Removed in step 2	Removed in step 3
Steroids	115	56	43	2	14
Terpenoid alkaloids	77	52	6	6	17
Pyrrolizidine alkaloids	65	48	9	6	2
Steroidal alkaloids	81	45	32	0	4
Isoquinoline alkaloids	89	45	44	0	0
Quinolizidine alkaloids	52	34	6	4	8
Sesquiterpenes	122	31	35	53	3
Polyketides	110	30	46	30	4
Indole alkaloids	72	27	42	0	3
Triterpenes	56	26	14	7	9
Diterpenes	83	26	50	6	1
Amaryllidaceae alkaloids	23	19	1	1	2
Saponins	42	16	21	2	3
Naphthalene- and anthracene-derivatives	35	13	2	18	2
Monoterpenes	118	11	37	64	6
Tropane alkaloids	27	9	4	13	1
Phenylpropanoids	74	9	26	32	7
Piperidine alkaloids	38	8	9	15	6
Tropolone alkaloids	8	6	0	0	2
Pyridine alkaloids	14	2	10	2	0
Indolizidine alkaloids	1	1	0	0	0
Quinones	4	1	0	3	0
Aliphatic acids	15	1	0	14	0
Amines	28	0	23	5	0
Sulfur compounds	21	0	0	21	0
Glucosinolates	31	0	14	17	0
Polyacetylenes	23	0	14	9	0
Non-protein amino acids	21	0	3	18	0
Cyanogenic glycosides	16	0	4	12	0
Quinoline alkaloids	13	0	13	0	0
Purine alkaloids	6	0	4	2	0
Pyrrolidine alkaloids	5	0	2	1	1
Quinazoline alkaloids	5	0	5	0	0
Acridine alkaloids	4	0	4	0	0
Nitriles	4	0	3	1	0
Pyrimidine alkaloids	4	0	0	4	0
Tetraterpenes	4	0	0	4	0

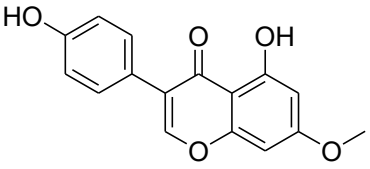
Chemical structures



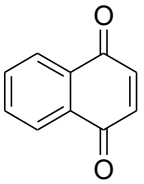
(1)



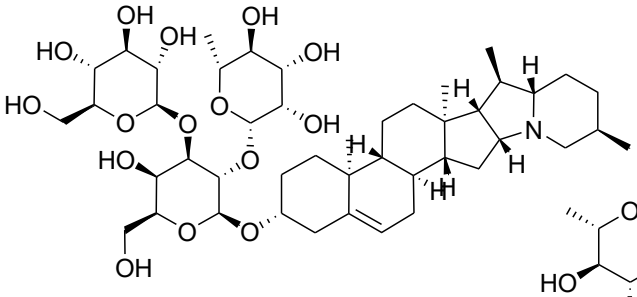
(2)



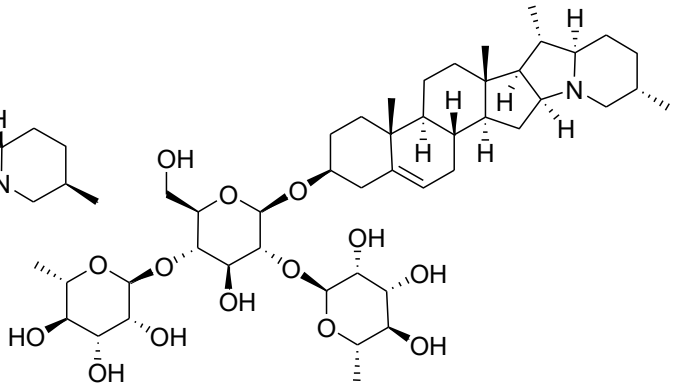
(3)



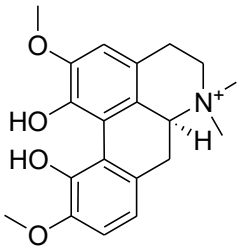
(4)



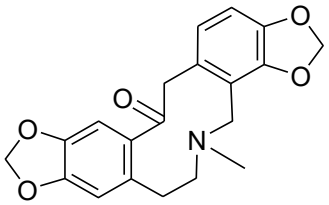
(5)



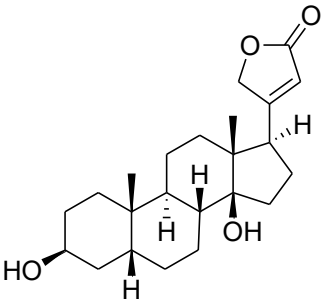
(6)



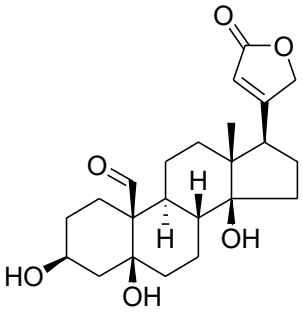
(7)



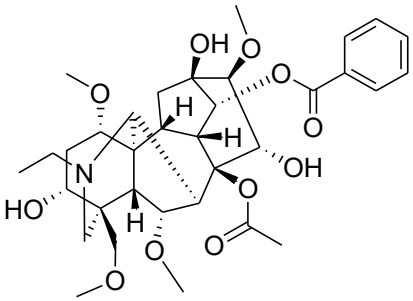
(8)



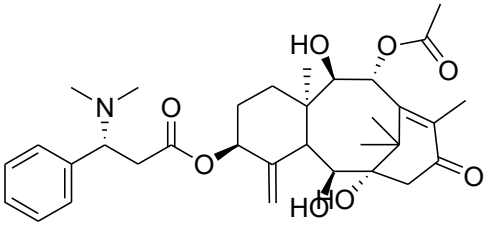
(9)



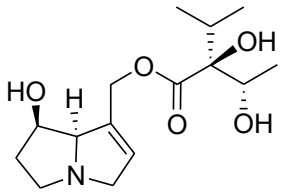
(10)



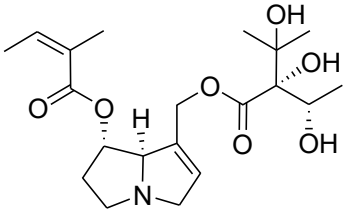
(11)



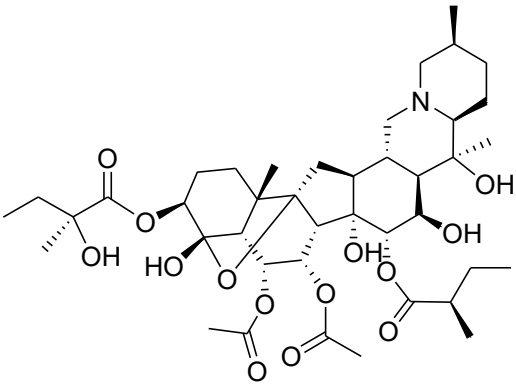
(12)



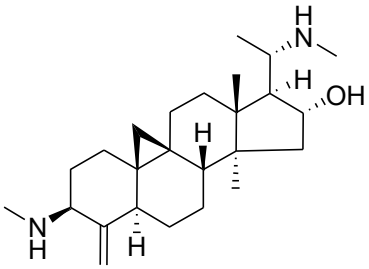
(13)



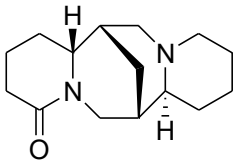
(14)



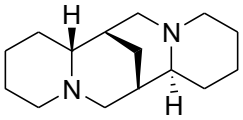
(15)



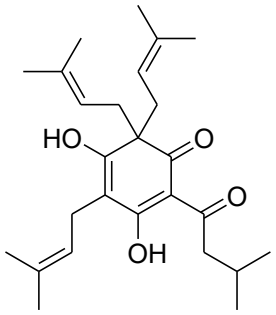
(16)



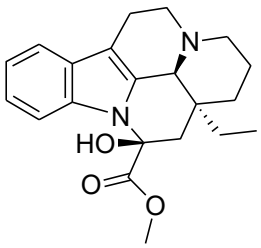
(17)



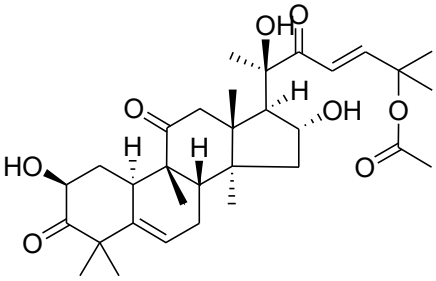
(18)



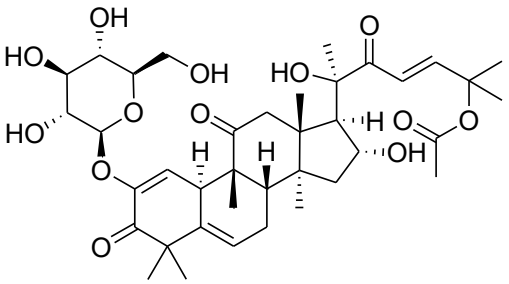
(19)



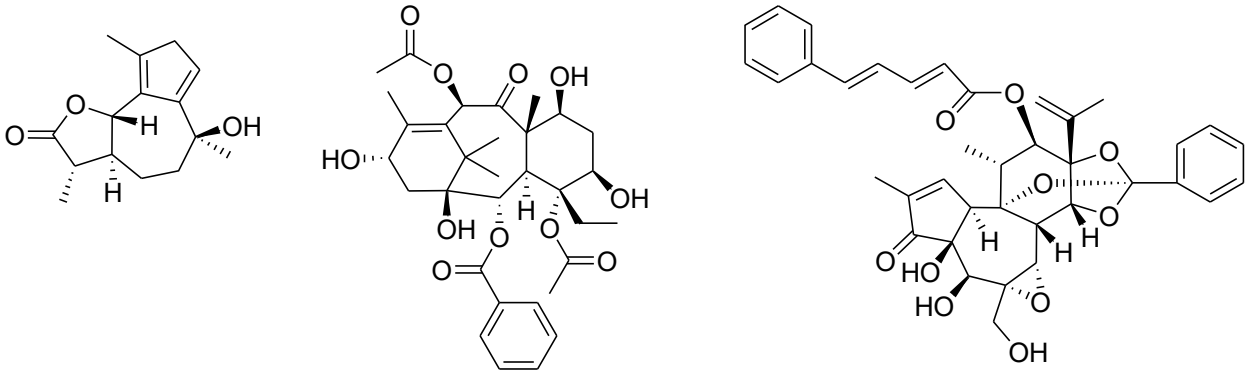
(20)



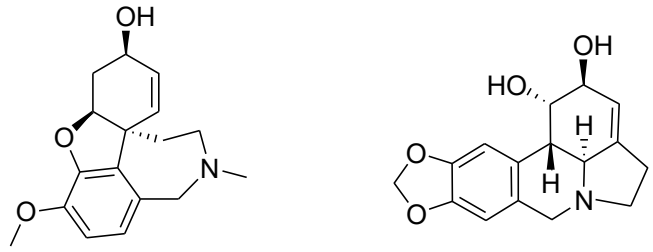
(21)



(22)



(23) (24) (25)



(26) (27)

Figure S5. Chemical structures of phytotoxins discussed in the main text: (1) ptaquiloside, (2) artemisinin, (3) formononetin, (4) juglone, (5) α-solanine and (6) α-chaconine, (7) (+)- magnoflorine, (8) protopine, (9) digitoxigenin, (10) strophanthidin, (11) aconitine, (12) taxine B, (13) lycopsamine, (14) heliosupine, (15) protoveratrine A, (16) cyclobuxine D, (17) (+)-lupanine, (18) (-)-sparteine, (19) lupulone, (20) (+)-vincamine, (21) cucurbitacin B, (22) elaterinide, (23) artabsin, (24) baccatin III, (25) mezerein, (26) galanthamine, and (27) lycorine.

References

1. Lauber, K.; Wagner, G.; Gygax, A. *Flora Helvetica - Illustrierte Flora der Schweiz*, ed. 5; Haupt Verlag: Bern, Switzerland, 2012.
2. Roth, L.; Daunderer, M.; Komann, K. *Giftpflanzen-Pflanzengifte: Vorkommen, Wirkung, Therapie, allergische und phototoxische Reaktionen*, 6th ed.; Nikol: Hamburg, Germany, 2012.
3. Teuscher, E.; Lindequist, U. *Biogene Gifte: Biologie-Chemie-Pharmakologie-Toxikologie*, 3rd ed.; Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, Germany, 2010.
4. Clinical Toxicology (CliniTox). http://www.vetpharm.uzh.ch/clinitox/ (20.10.2017).
5. Info flora: The National Data and Information Center on the Swiss Flora. https://www.infoflora.ch/en/ (20.10.2017).
6. Commission Implementing Regulation (EU): Adopting a list of invasive alien species of Union concern pursuant to Regulation (EU) No 1143/2014 of the European Parliament and of the Council. *Official Journal of the European Union* **2016**, L189/4.
7. European Food Safety, A., Compendium of botanicals reported to contain naturally occuring substances of possible concern for human health when used in food and food supplements. *EFSA Journal* **2012**, *10*, 2663.
8. 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*, Saito, K.; Dixon, R. A.; Willmitzer, L., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 165-181.
9. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound databases. https://pubchem.ncbi.nlm.nih.gov/ (06.12.2016).
10. ChemSpider. http://www.chemspider.com/ (06.12.2017).
11. Duke, J. A.; Williams, M. C., Phytotoxin tables. *CRC Crit. Rev. Toxicol.* **1977**, *5*, 189-237.
12. Estimation Programs Interface Suite for Microsoft Windows, v 4.11; United States Environmental Protection Agency: Washington, DC, 2017.
13. ACD/Percepta, release 2016.2; Advanced Chemistry Development, Inc.: Toronto, Canada, 2016.
14. Manchester, J.; Walkup, G.; Rivin, O.; You, Z., Evaluation of pKa Estimation Methods on 211 Druglike Compounds. *J. Chem. Inf. Model.* **2010**, *50*, 565-571.
15. Drwal, M. N.; Banerjee, P.; Dunkel, M.; Wettig, M. R.; Preissner, R., ProTox: a web server for the in silico prediction of rodent oral toxicity. *Nucleic Acids Res.* **2014**, *42*, W53-W58.
16. Ecological Structure Activity Relationships (ECOSAR) Predictive Model, V 2.0; United States Environmental Protection Agency: Washington, DC, 2017.
17. Sanderson, H.; Johnson, D. J.; Reitsma, T.; Brain, R. A.; Wilson, C. J.; Solomon, K. R., Ranking and prioritization of environmental risks of pharmaceuticals in surface waters. *Regul. Toxicol. Pharm.* **2004**, *39*, 158-183.
18. Arp, H. P. H.; Brown, T. N.; Berger, U.; Hale, S. E., Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility. *Environmental Science: Processes & Impacts* **2017**, *19*, 939-955.
19. Arnot, J.; Gouin, T.; Mackay, D., Practical Methods for Estimating Environmental Biodegradation Rates. *Report to Environment Canada; CEMN Report No. 200503* **2005**.
20. Strempel, S.; Scheringer, M.; Ng, C. A.; Hungerbühler, K., Screening for PBT Chemicals among the “Existing” and “New” Chemicals of the EU. *Environ. Sci. Technol.* **2012**, *46*, 5680-5687.