Call it a "Nightshade"—A Hierarchical Classification Approach to Identification of Hallucinogenic Solanaceae spp. Using DART-HRMS-derived Chemical Signatures

Supplementary Material

Samira Beyramysoltan[a], Nana-Hawwa Abdul-Rahman[a] and Rabi A. Musah[a*]

[a]Department of Chemistry, State University of New York at Albany, 1400 Washington Ave, Albany, NY 12222

*Corresponding author: E-mail address: rmusah@albany.edu

Phone: 518-437-3740

Fax: 518-442-3462

Contained within this supporting information document are: a list of the species represented in this study; the input and output properties of the classification model; the performance results of "between genus" and "between species" discrimination for the PLS-DA models; a list of the significance and rank of the variables detected by PLS-DA, along with their corresponding VIP and COSSs scores; representative DART-high-resolution mass spectra for the species represented in the study; the PLS-DA scores and loadings plots associated with the spectra of the species represented in this study; and presentations of the SPCA results for each of the represented genera.

Supplementary Tables:

Table S1. Species represented in this study and the vendor suppliers.

Genus	Species	Vendor
Atropa	*A. beatica*	Ebay
	A. belladonna	Strictly Medicinal Seeds
	A. komarovii	Strictly Medicinal Seeds
Brugmansia	*B. arborea*	Seedman
	B. aurea	
	B. sanguinea	
	B. suaveolens	
	B. versicolor	
Datura	*D. ceratocaula*	World Seed Supply, Georgia Vines
	D. discolor	World Seed Supply
	D. ferox	World Seed Supply
		Georgia Vines
	D.inoxia	World Seed Supply, Georgia Vines
	D. leichhardtii	Hudson
	D. metel	Georgia Vines
	D. quercifolia	Hirt's Gardens
	D. stramonium	World Seed Supply
	D. wrightii	Georgia Vines
Hyoscyamus	*H. albus*	Ebay
	H. aureus	Ebay
	H. niger	Amazon
	H. pusillus	Ebay
	H. muticus	Ebay
Mandragora	*M. autumnalis*	Amazon
	M. officinarum	Ebay

Table S2. Input and output properties of the generated classification model.

Classifier node	Matrix dimensions		Number of LVs	% EV of data	% EV of response	CV error (5 fold)	Accuracy
	Train	Test					
Genus level							
TAP	186×170	32×170	7	85	84	Error: 0.01	99
species level							
Atropa	31×40	7×40	3	97	90	Error: 0	100
Brugmansia	38×50	8×50	4	50	75	Error: 0	100
Datura	70×80	9×80	11	97	82	Error: 0.05	95
Hyoscyamus	31×30	5×30	5	96	87	Error: 0.03	97
Mandragora	15×15	3×15	2	91	98	Error: 0	100

LVs: Latent variables; EV: Explained variance; CV: Cross validation.

Table S3. Performance results of between genus and between species discrimination models, which were created by PLS-DA for each node of the hierarchical classification tree.

Data type	Class	Classification model performance		
		Sensitivity	Specificity	Precision
Tropane alkaloid plants	*Atropa*	1	0.99	0.96
	Brugmansia	0.97	1	1
	Datura	0.99	0.99	0.99
	Hyoscyamus	1	1	1
	Mandragora	1	1	1
Atropa	*A. baetica*	1	1	1
	A. belladonna	1	1	1
	A. komarovii	1	1	1
Brugmansia	*B. arborea*	1	1	1
	B. aurea	1	1	1
	B. sanguinea	1	1	1
	B. suaveolens	1	1	1
	B. versicolor	1	1	1
Datura	*D. ceratocaula*	1	1	0.8
	D. discolor	1	1	1
	D. ferox	1	1	1
	D. inoxia	0.78	1	1
	D. leichhardtii	1	1	1
	D. metel	1	0.98	0.9
	D. quercifolia	0.89	0.97	0.8
	D. stramonium	1	0.98	0.9
	D. wrightii	0.89	1	1
Hyoscyamus	*H. albus*	1	1	1
	H. aureus	1	0.96	0.83
	H. muticus	1	1	1
	H. niger	1	1	1
	H. pusillus	0.83	1	1
Mandragora	*M. autumnalis*	1	1	1
	M. officinarum	1	1	1

Table S4. Significance and rank of the variables detected by PLS-DA for the tropane alkaloid plant (TAP), *Atropa*, *Brugmansia*, *Datura*, *Hyoscyamus*, and *Mandragora* feature sets, with VIP scores of >1. The calculated VIP and COSS scores are presented.

Data type	Variable	Variable score		Data type	Variable	Variable score	
		VIP*	COSS**			VIP*	COSS**
Tropane alkaloid plants (TAP)	61.0306	1.89	2.01	Brugmansia	85.0294	1.01	0.17
	90.0919	1.55	1.23		96.0835	1.33	0.17
	96.0454	1.85	1.63		124.0430	1.36	0.66
	110.0692	1.06	1.12		130.1241	1.19	0.17
	124.1093	1.76	1.68		140.0793	1.05	0.23
	127.0407	2.20	4.40		141.0888	1.32	0.14
	130.0521	1.53	0.64		144.2680	1.43	0.20
	142.1210	3.18	0.97		151.0790	1.31	0.11
	143.0878	1.33	0.74		166.1236	1.34	0.61
	144.1015	2.43	3.50		174.7375	1.32	0.13
	145.0506	1.57	0.90		175.6778	1.80	1.29
	146.1635	1.13	5.88		193.0501	1.47	0.09
	153.1258	1.11	2.56		219.1118	1.12	0.67
	158.1165	2.67	1.96		238.1287	1.05	0.22
	163.0682	1.21	0.56		260.1597	1.08	0.20
	174.1148	5.36	4.85		317.2113	1.05	0.20
	193.0501	2.76	3.12		445.2109	1.24	0.10
	263.2349	1.26	1.95		618.3181	1.72	0.95
	272.1601	1.37	5.71	Datura	124.1093	1.21	1.74
	281.2442	3.06	1.26		138.0929	1.14	3.18
	290.1726	4.78	1.52		140.1099	1.33	0.45
	291.1783	1.56	2.86		142.1210	2.75	7.06
	298.2717	2.52	0.81		153.0637	1.99	0.70
	304.1571	2.82	1.43		156.1013	1.35	0.82
	381.3498	1.14	3.80		158.1165	1.70	2.57
Atropa	84.0455	1.22	5.85		163.0682	1.84	2.41
	127.0407	1.37	3.47		174.1148	2.79	0.42
	130.0521	1.11	2.64		183.0908	1.29	2.09
	145.0506	1.30	1.38		193.0501	2.03	0.54
	153.1258	1.05	1.25		209.0448	1.33	4.10
	193.0501	3.09	0.99		290.1726	2.71	2.15
	272.1601	2.08	0.76		291.1783	1.31	3.10
	281.2442	2.60	0.67		304.1571	2.79	4.67
	298.2717	1.66	0.52		305.1599	1.01	2.22
Hyoscyamus	90.0919	1.65	1.37		306.1674	1.01	0.94
	127.0407	1.13	0.82		423.3601	1.23	0.59
	130.0521	1.51	2.85		441.3984	1.05	0.59
	142.1210	1.18	2.25		124.1093	1.21	1.74
	281.2442	2.04	0.60	Mandragora	61.0306	1.42	1.25
	290.1726	2.51	6.80		130.0521	1.04	0.47
	304.1571	2.14	7.75		145.0506	1.48	0.39
					290.1726	2.20	15.09
					381.3498	1.11	5.78

*VIP refers to variable of importance projection and is derived from the PLS-DA results.
**COSS refers to the calculated conditional synergetic score (COSS) value for variable importance using SPA-PLS-DA.
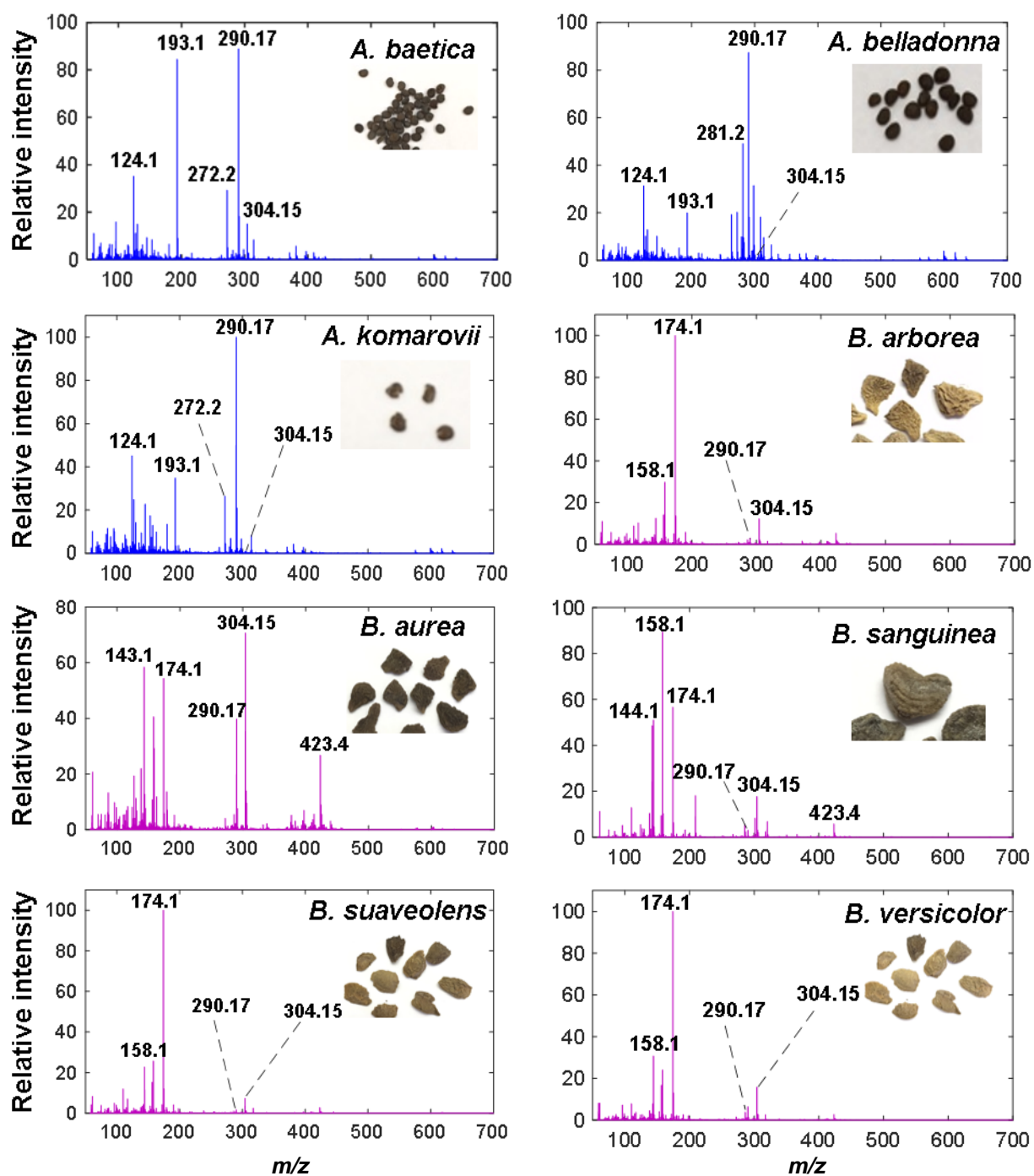
Supplementary Figures:

Figure S1. Representative DART high-resolution mass spectra of the seed species represented in this study.
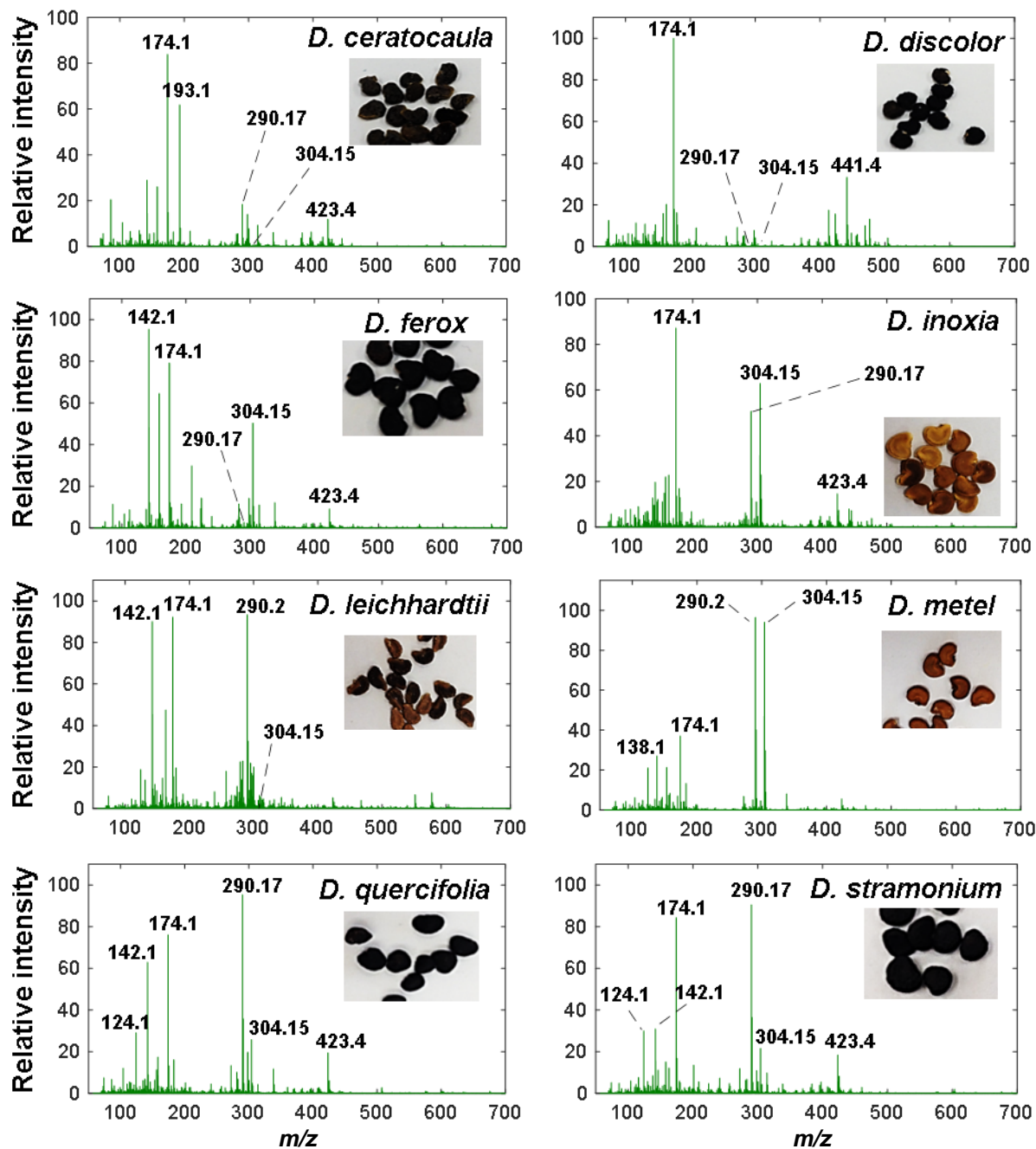
Figure S1 (continued). Representative DART high-resolution mass spectra of the seed species represented in this study.
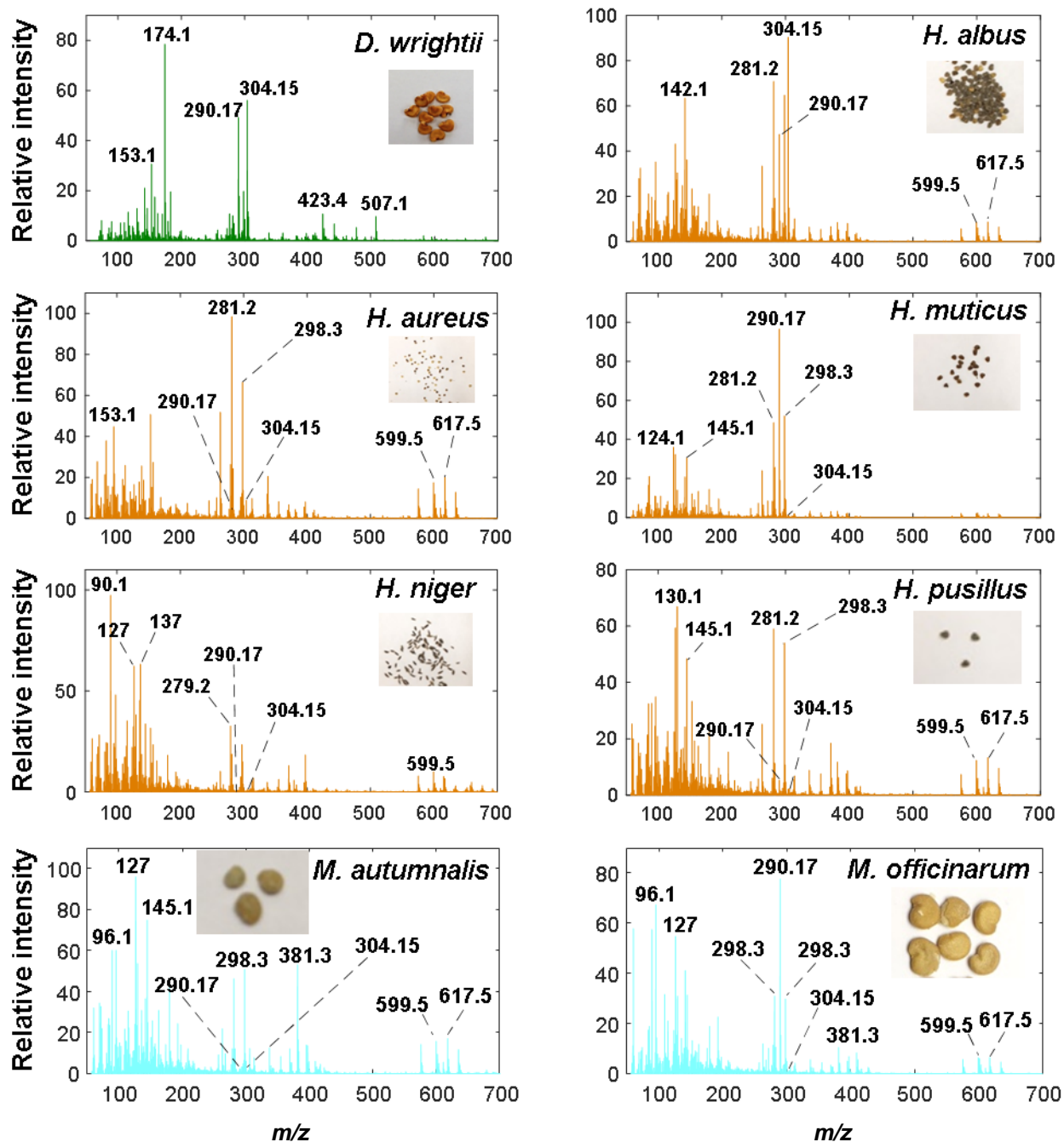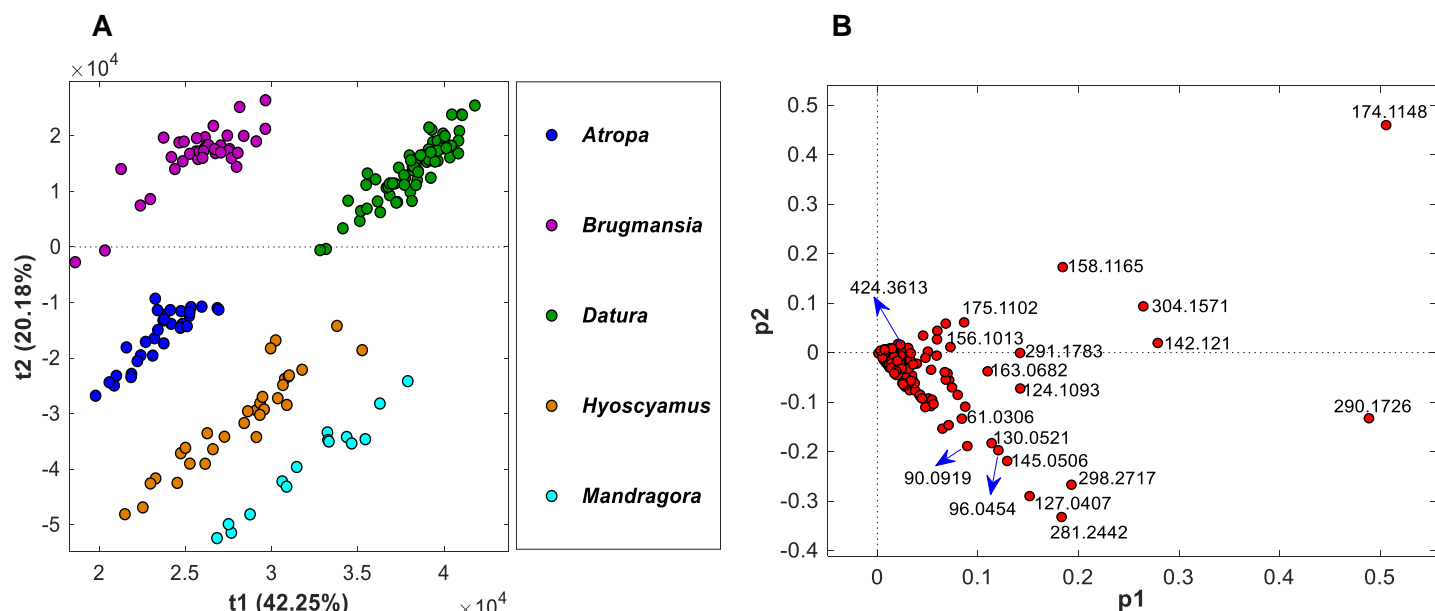
Figure S1 (continued). Representative DART high-resolution mass spectra of the seed species represented in this study.

Figure S2. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of the genus node. Two latent variables explain ~62.5 % of the data variance. The figure shows groupings of genera that exhibit similar variables (i. e. *m/z* values) profiles. The *Atropa*, *Mandragora* and *Hyoscyamus* group is separated from *Brugmansia* and *Datura* in different directions along the y-axis (t2). PLS-DA loadings clearly show that *m/z* 281.241, 127.0409 and 298.2709 are important in the clustering of *Atropa*, *Mandragora* and *Hyoscyamus* classes, and 174.1109, 158.11, 304.151 and 175.1109 are significant in the *Brugmansia* and *Datura* grouping.
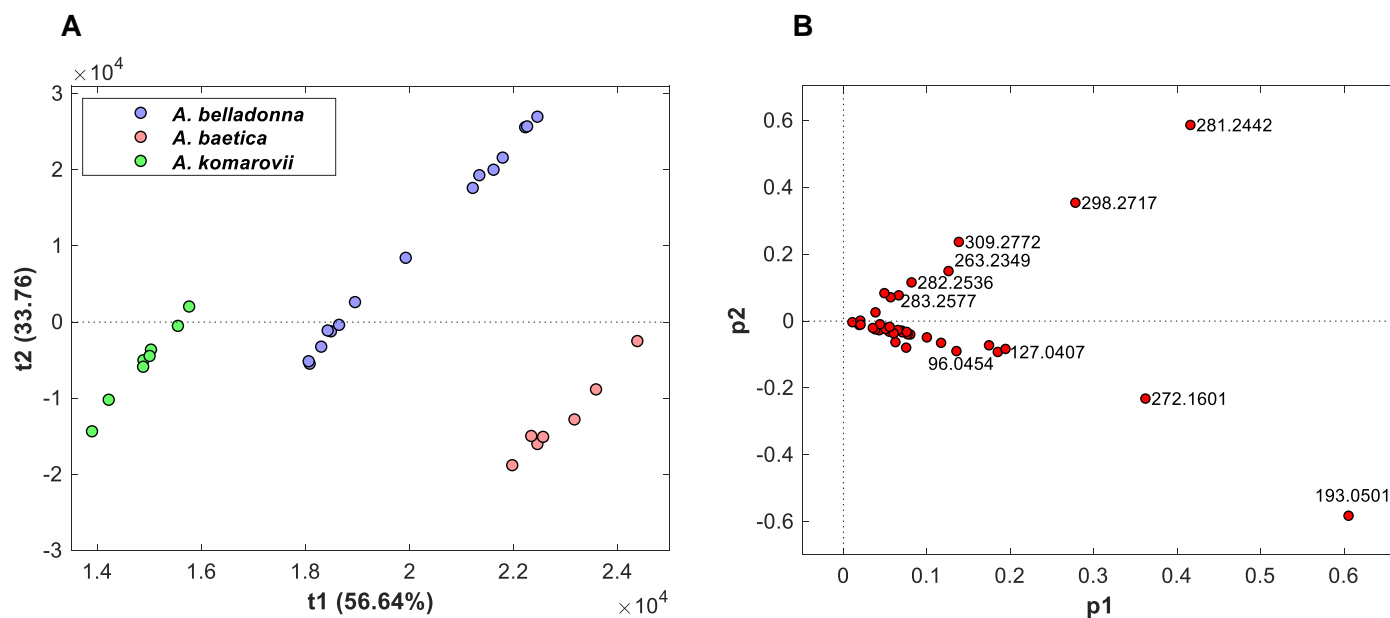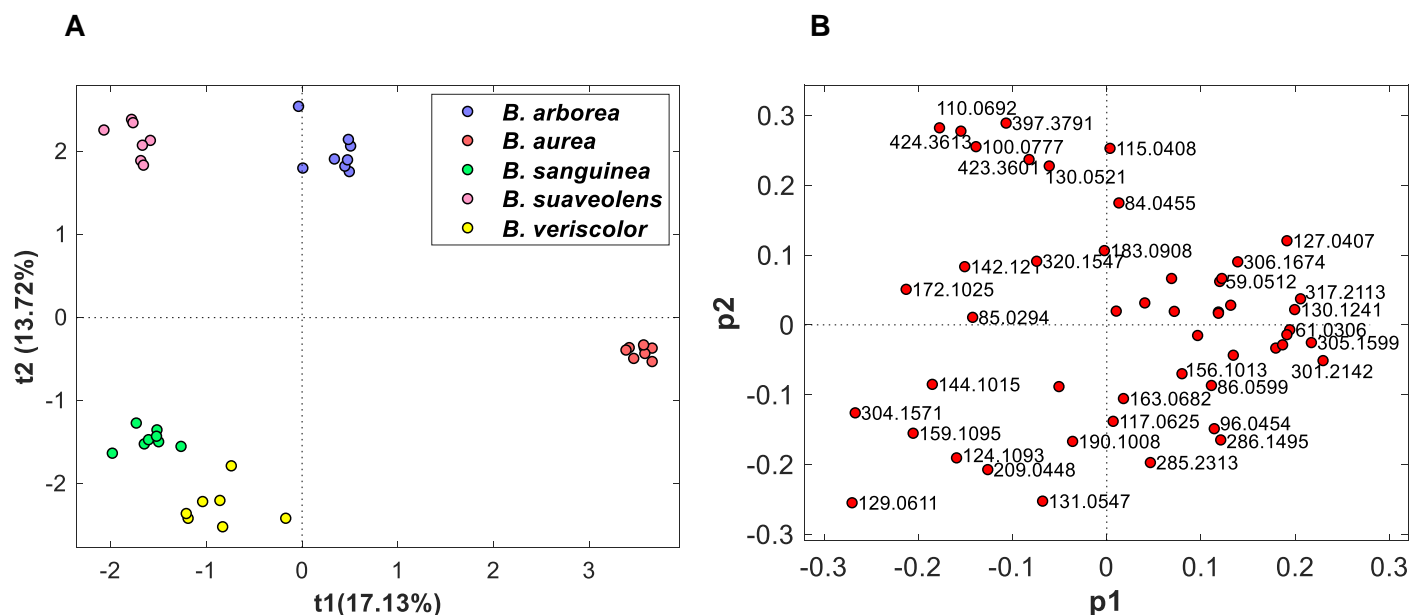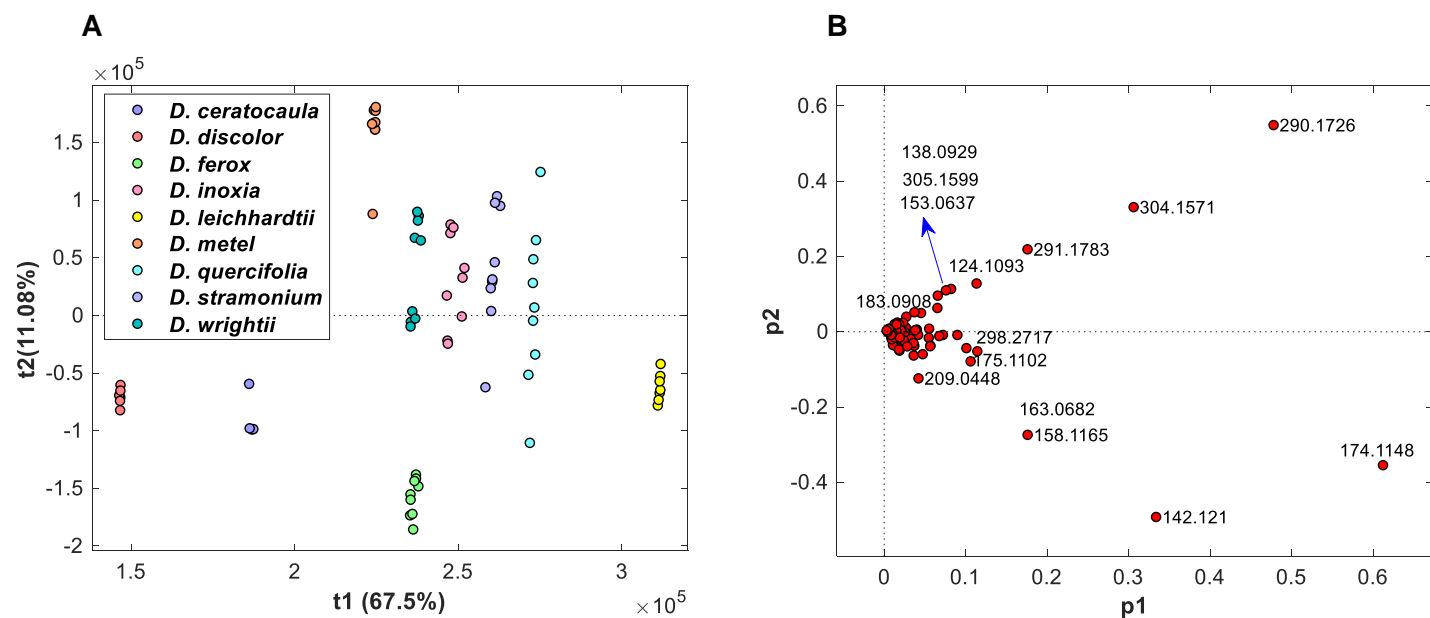
Figure S3. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of *Atropa* seed samples. Two latent variables explained ~91 % of the data variance. The figure shows groupings of species that exhibited similar variables (i. e. *m/z* values) profiles. *A. belladonna* falls within the positive part of the y-axis (t2), while *A. baetica* and *A. komarovii* grouped together and are located in the negative direction. Loadings show that *m/z* 281.2442, 298.2717, 263.2349 and 309.2772 are most important in defining the *A. belladonna* class; 193.0501, 96.0454, 127.0407 and 272.1601 influence the grouping of *A. bactica* and *A. komarovii* samples.

Figure S4. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of *Brugmansia* seed samples. Two latent variables explain ~31 % of the data variance. The scores plot shows high similarity between *B. sanguinea* and *B. versicolor*. LV1 in the scores plot displays the differentiation of *B. arborea* and *B. aurea* groups from *B. sanguinea, B. suaveolens* and *B. versicolor*. Masses 159.11, 129.06, 144.10, 124.11 and 304.16 are correlated with the *B. sanguinea* and *B. versicolor* group. Variables most effective for *B. aurea* differentiation are *m/z* 61.03, 301.2142 and 305.16. The *m/z* 110.01, 397.34, 424.36 and 100.0777 are highly correlated with the *B. suaveolens* species. The loadings plot clearly shows the importance of *m/z* values 115.04, 84.0455 and 127.0407 in separation of the *B. arborea* class.

A

B

Figure S5. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of *Datura* seed samples. Two latent variables explain ~79% of the data variance. The loadings show that *m/z* values 158.1165, 142.121, 174.1148, 298.2717, 163.0682, 175.1102, 290.1726, 291.1783, 124.1093, 209.0448 and 304.1571 are heavily weighted in defining all species.
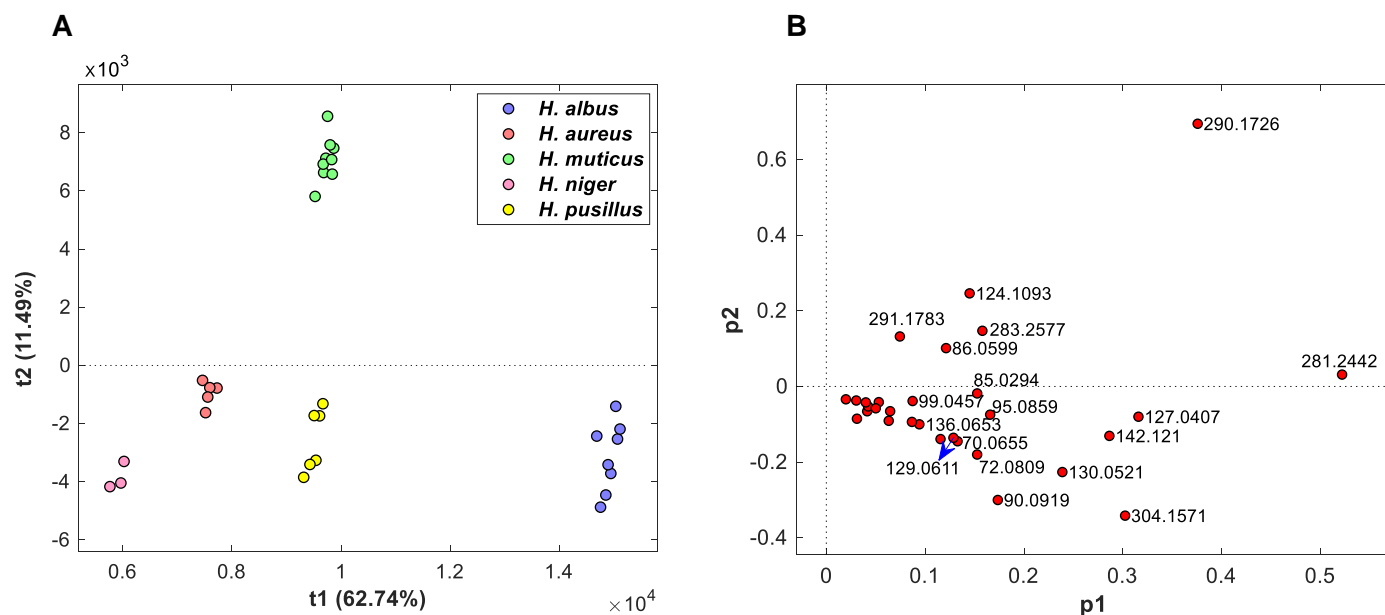
Figure S6. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of *Hyoscyamus* seed samples. Two latent variables account for ~74 % of the data variance. The loadings show that *m/z* values 290. 1726, 124.1093, 283.2577, and 291.1783 are the most significant for the *H. muticus* grouping. The *m/z* values 90.0919, 130.0521, 304.1571, 142.121 and 127.0407 are the most heavily weighted for the *H. aureus*, *H. niger*, *H. pusilus* and *H. albus* group.
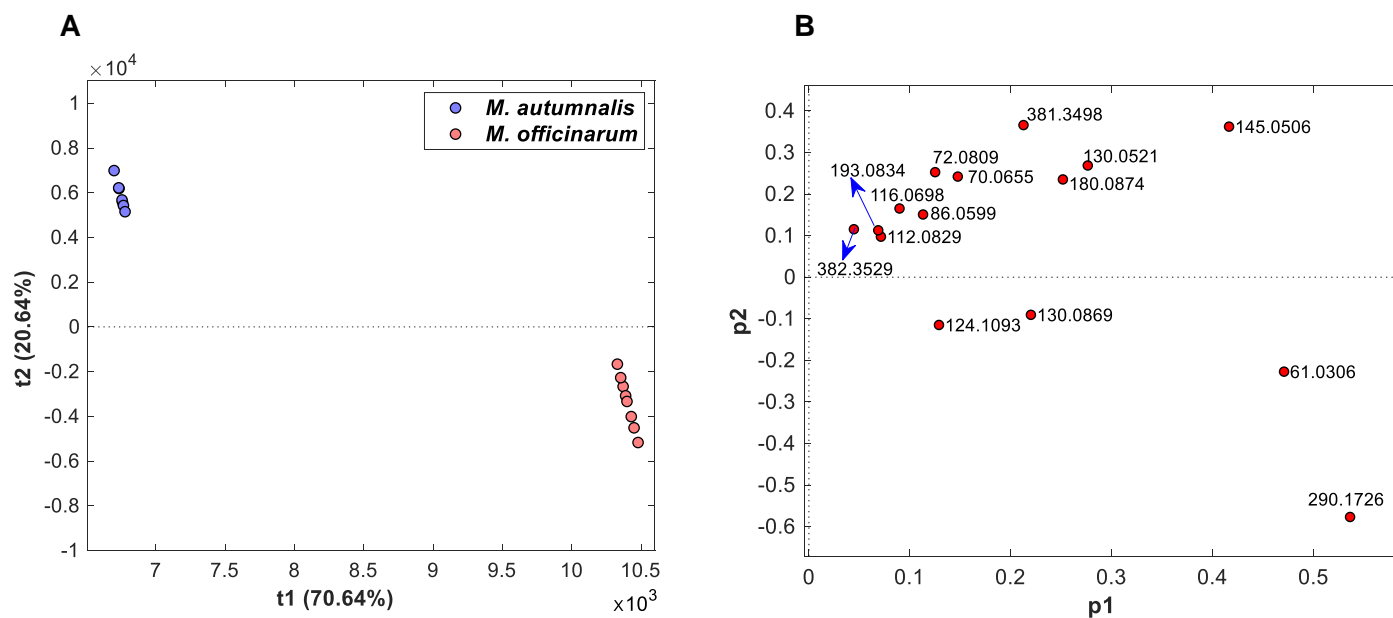
Figure S7. PLS-DA scores plot (A) and its corresponding loadings plot (B) based on DART-HRMS profiling of *Mandragora* seed samples. Two latent variables explained ~91 % of the data variance. The loadings showed that *m/z* values 290.1726 and 61.0306 are important in defining the *M. officinarum* species, while *M. autumnalis* is mostly influenced by *m/z* 145.0506, 72.0809, 70.0655 and 381.3498.
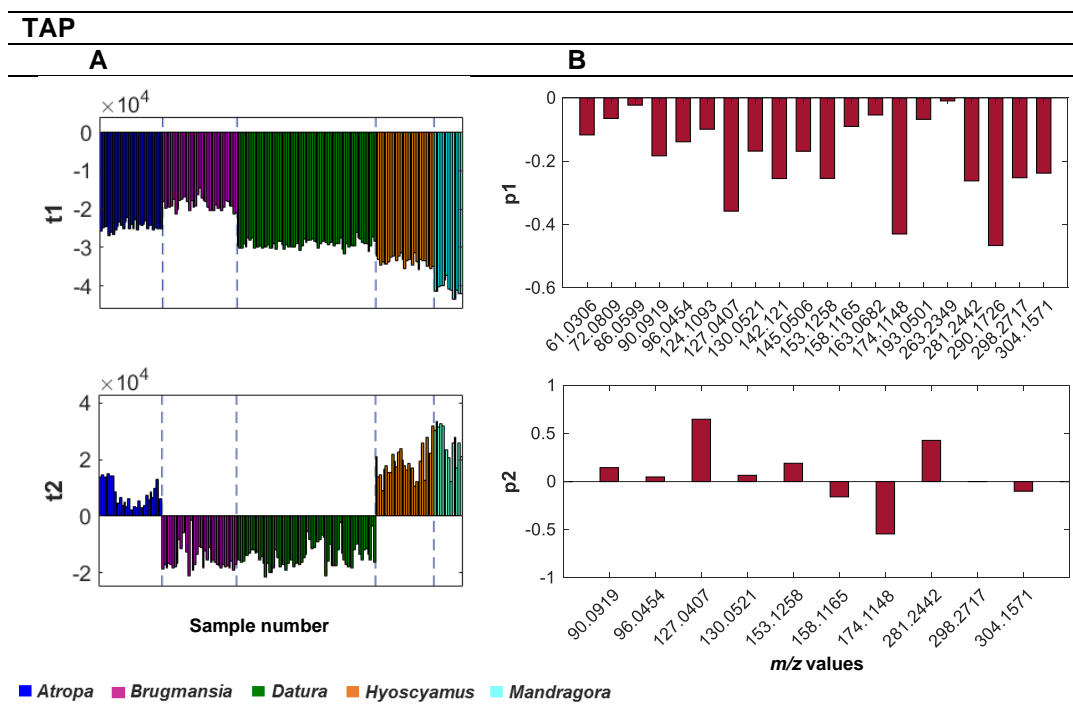
Figure S8. Presentation of the SPCA results for the genera (i.e. tropane alkaloid plant—TAP) node. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The scores bars clearly reveal the sample groupings with intensity thresholds in the first score and intensity and sign thresholds in the second score. The second loading enables visualization of the variables that were effective in grouping *Atropa*, *Mandragora* and *Hyoscyamus*, as distinguished from the *Brugmansia* and *Datura* grouping in the two parts of the plot. The loadings reveal that *m/z* 174.1148 and 127.0407 are the most important variables for differentiation of *Datura* and *Brugmansia* from the others.
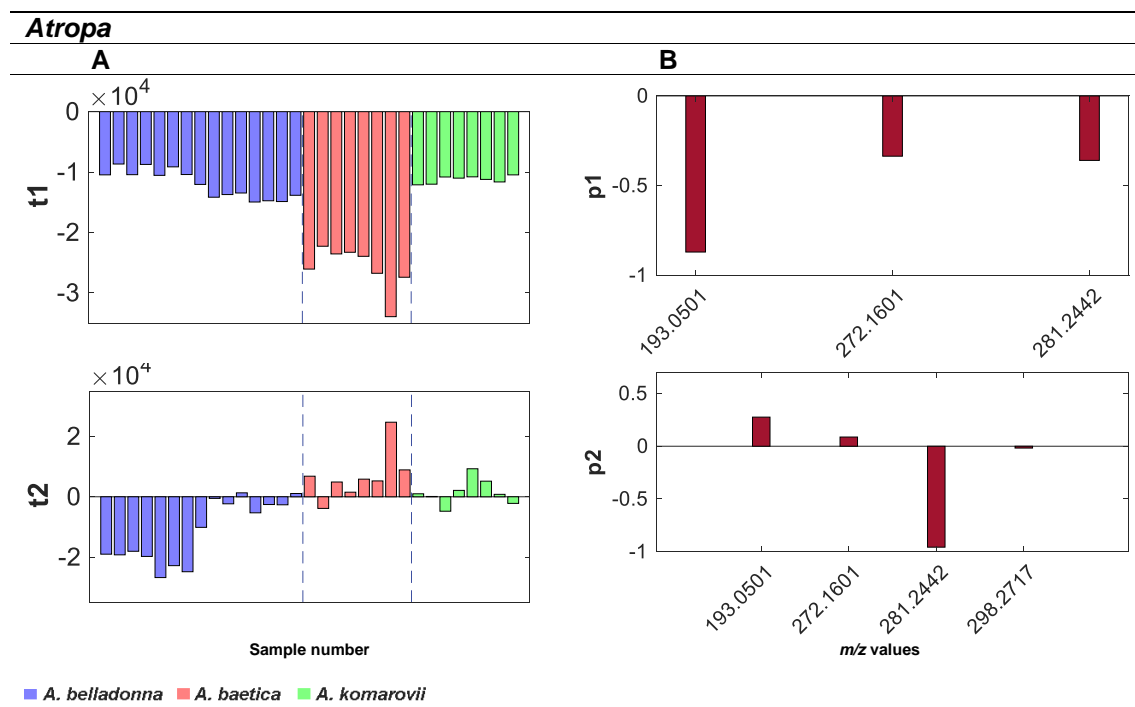
Figure S9. Presentation of the SPCA results for *Atropa* genus features. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The scores bars clearly reveal the sample groupings with intensity thresholds. *A. belladonna*, *A. komarvii* and *A. baetica* are separated by sign along PC2 (negative for *A. belladonna*; positive for *A. komarvii* and *A. baetica*). Based on the loadings, *m/z* 281.2525 is the most effective in enabling *A. belladonna* to be distinguished from the others.
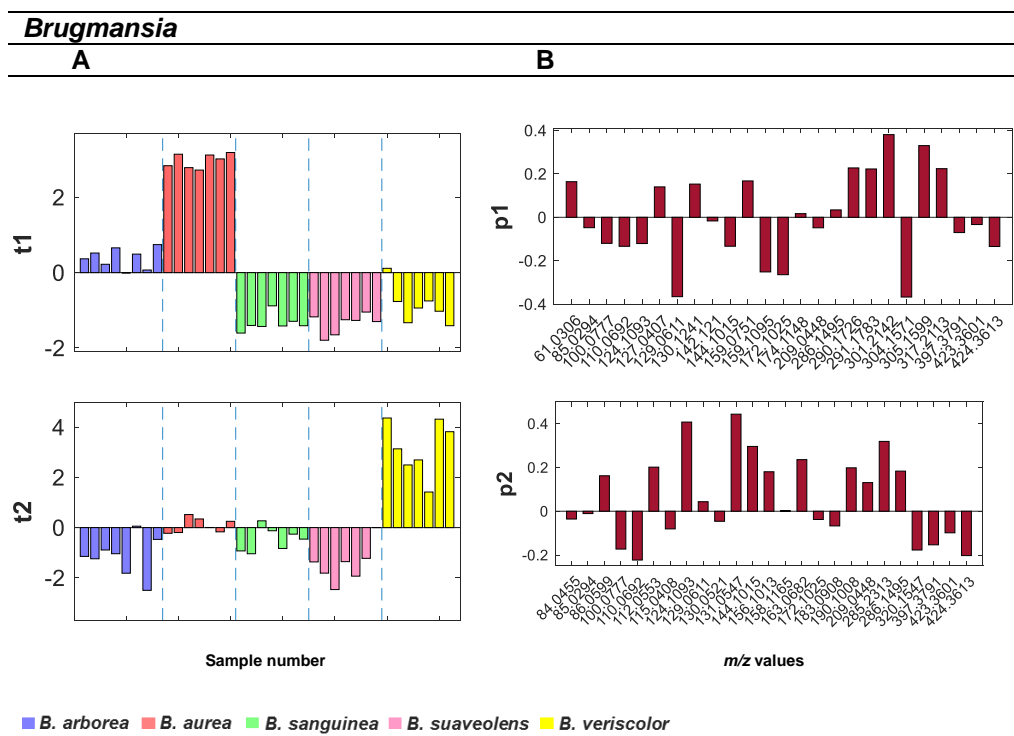
Figure S10. Presentation of the SPCA results for the *Brugmansia* genus features. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The first scores bars clearly reveal the sample groupings with sign thresholds in the first score (for separation of *B. arborea* and *B. aurea* from the others) and sign thresholds in the second score for separation of *B. versicolor* from the others. From the loadings bars, it was apparent that *m/z* 129.0611, 304.1571, 301.2142, 305.1599, 131.0547 and 124.1093 were the most important variables for species discrimination within the *Brugmansia* genus.
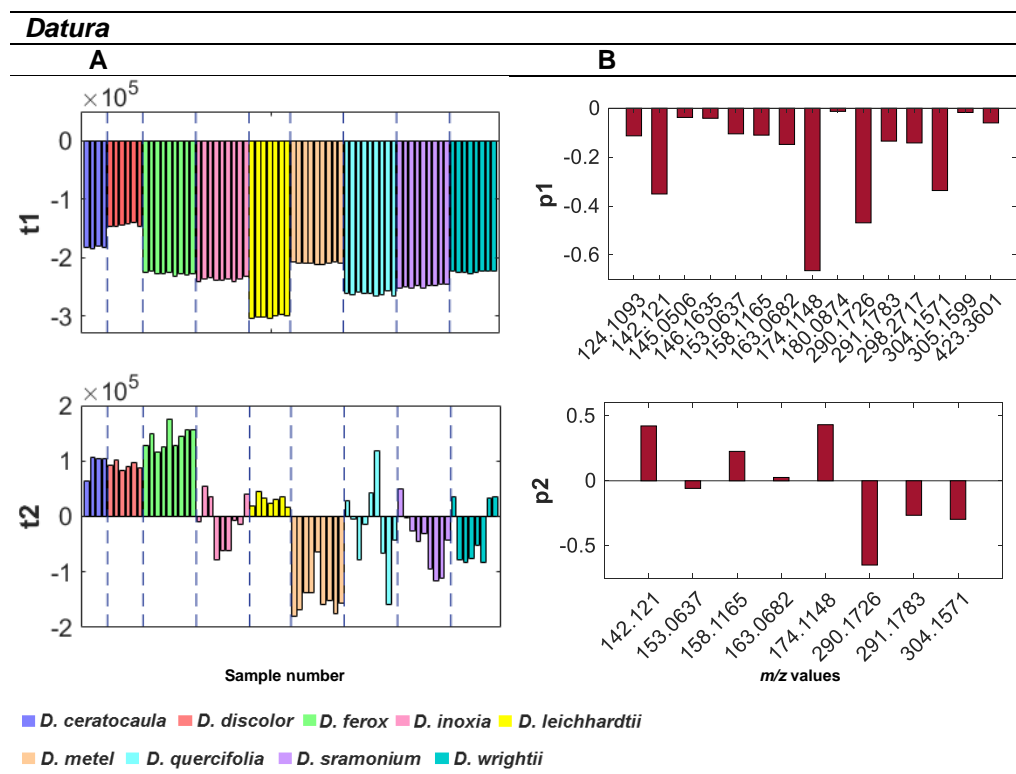
Figure S11. Presentation of the SPCA results for the *Datura* genus dataset. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The scores bars clearly reveal the sample groupings with intensity and sign thresholds in the second score, and with intensity thresholds in the first score.
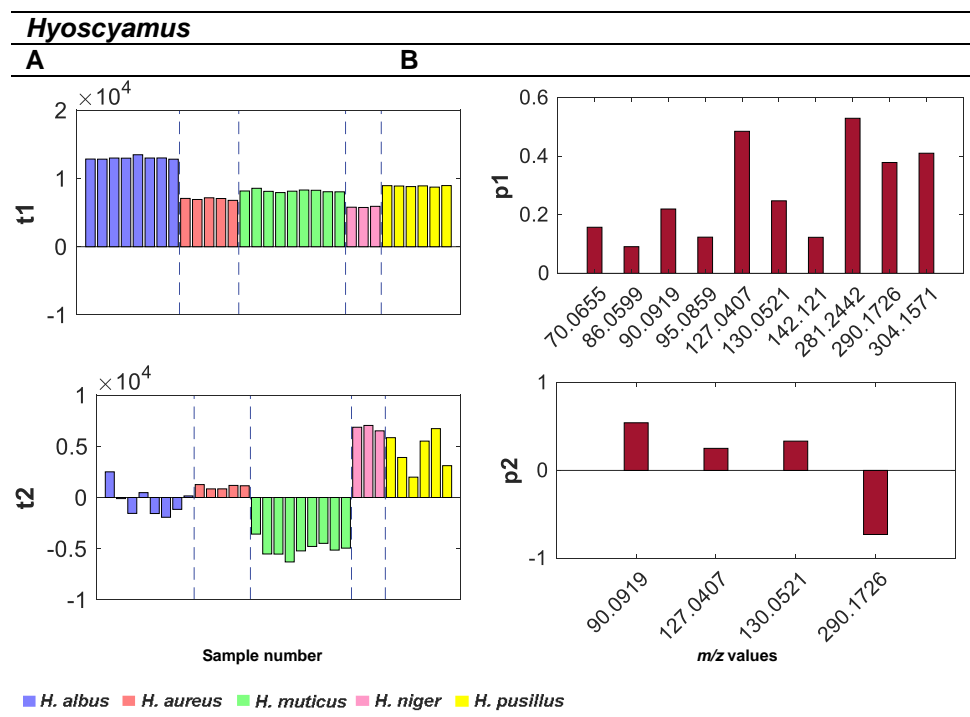
Figure S12. Presentation of the SPCA results for the *Hyoscyamus* genus features. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The scores bars clearly reveal sample groupings with intensity and sign thresholds in the first and second scores respectively. Three species had similar intensity thresholds for discrimination. The second loadings illustrate the impact of *m/z* 90.0919 and 290.1726 in the discrimination of *H. aureus*, *H. niger* and *H. pusillus* from the others.
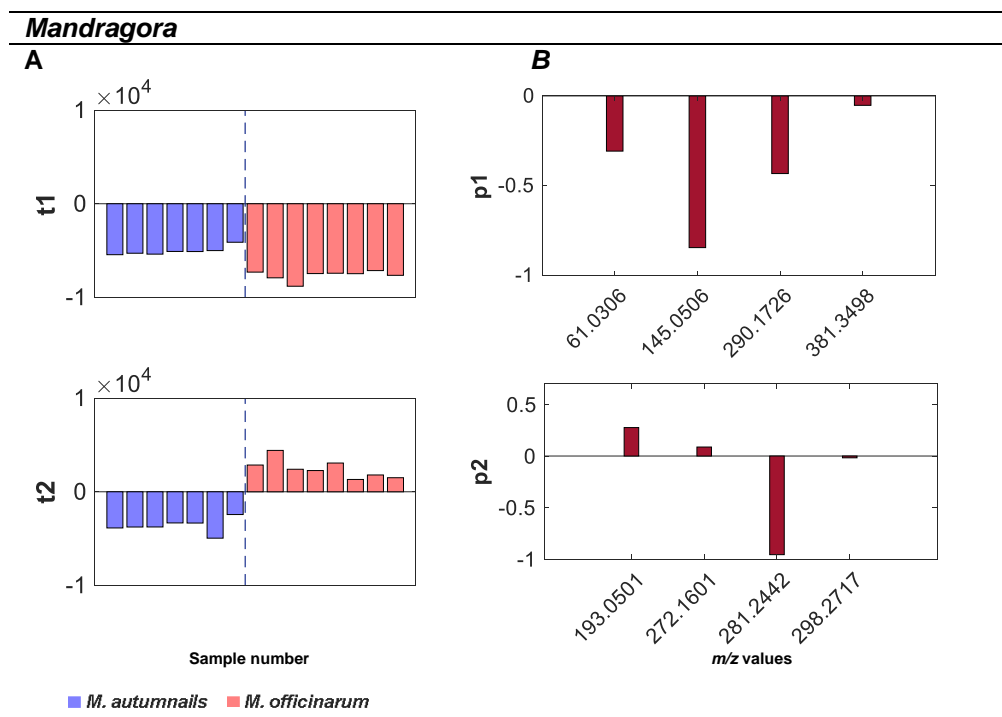
Figure S13. Presentation of the SPCA results for the *Mandragora* feature dataset. The scores (A) and loadings (B) were plotted against sample number and variable index respectively in the bar charts. The scores bars revealed the sample groupings with the intensity thresholds in the first score and the sign thresholds in the second. From the SPCA, the most heavily weighted *m/z* values were 281.2442 and 145.0506.