Supplementary Material

Can Statistical Evaluation Tools for Chromatographic Method Development Assist in the Natural Products Workflow? A Case Study on Selected Species of the Plant Family Malpighiaceae

Helena Mannochio-Russo,[*,†,‡] Paula Carolina P. Bueno,[§,^] Anelize Bauermeister,[‡,‖] Rafael Felipe de Almeida,[▽] Pieter C. Dorrestein,[‡] Alberto José Cavalheiro,[†] and Vanderlan S. Bolzani[*,†]

[†]*NuBBE, Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, SP, Brazil;*

[‡]*Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA;*

[§]*Faculty of Pharmaceutical Sciences of Ribeirão Preto, Department of Physics and Chemistry, University of São Paulo, Ribeirão Preto, SP, Brazil;*

[^]*Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany;*

[‖]*Biomedical Sciences Institute, University of São Paulo, São Paulo, SP, Brazil;*

[▽]*Lamol Lab, Feira de Santana State University (UEFS), Department of Biological Sciences, Feira de Santana, BA, Brazil.*

*Authors for correspondence: Prof. Dr. Vanderlan da Silva Bolzani; M.Sc. Helena Mannochio-Russo.

Phone number: + 55 16 3301 9660; Fax number: + 55 16 33222308; e-mail: vanderlan.bolzani@unesp.br; helenamrusso@gmail.com

Authors contribution

H.M.R. designed the research, performed the UHPLC-PDA and UHPLC-HRMS/MS analyses, processed the data, wrote the manuscript, and revised the manuscript. P.C.P.B. Assisted in the validation procedure, wrote the manuscript, and revised the manuscript. A.B. Assisted in the MS/MS data analysis, wrote the manuscript, and revised the manuscript. R.F.A. provided the plant material, identified the species, and revised the manuscript. P.C.D. provided consultation on the manuscript elaboration and revised the manuscript. A.J.C. designed the research and revised the manuscript. V.S.B. designed the research, acquired funding, and revised the manuscript. All authors have approved to the final version of the manuscript.

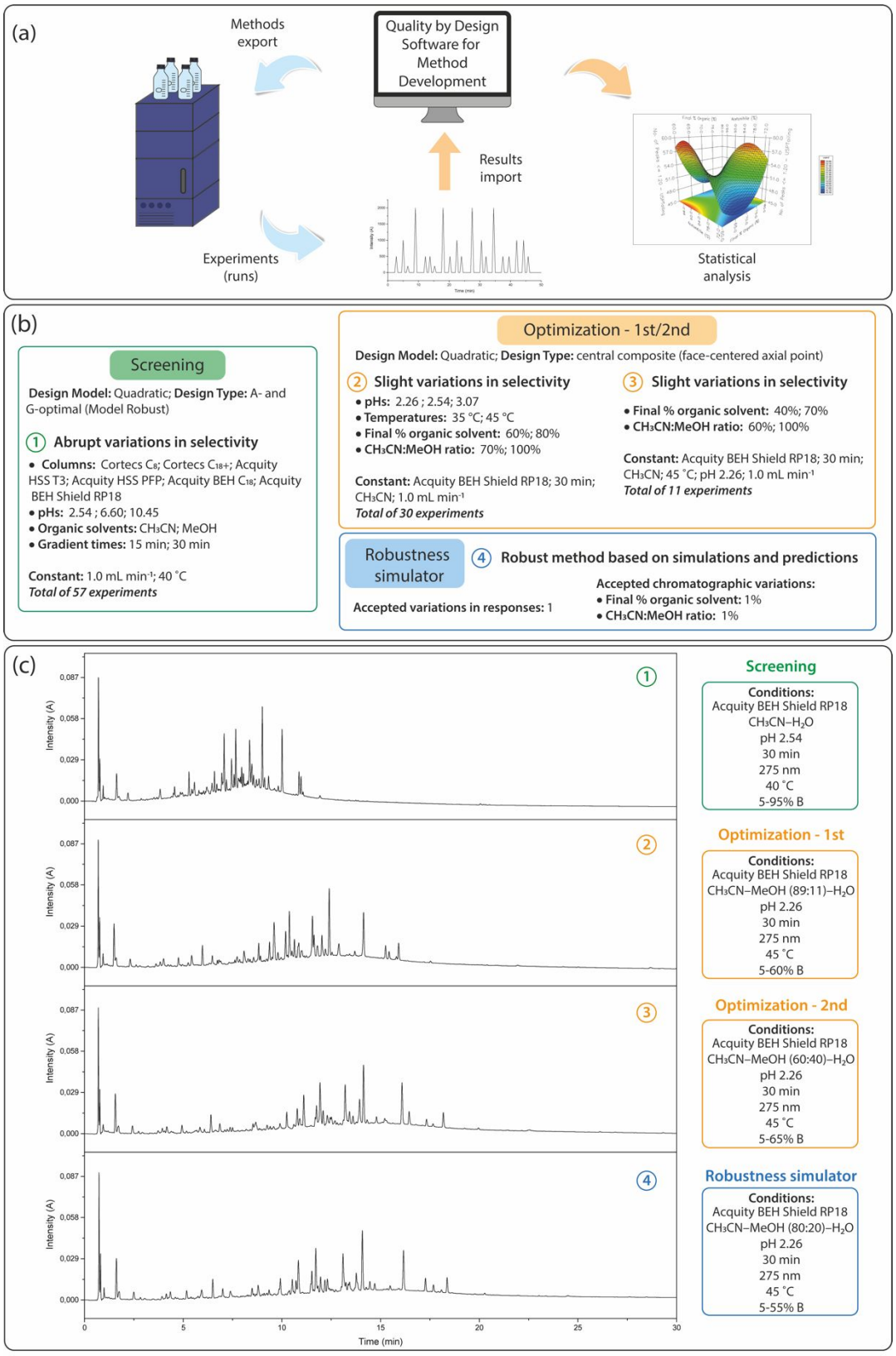


Figure S1. Workflow followed for chromatographic method development. (a) The QbD software created and exported the methods needed for the method development directly to the UHPLC software. The chromatograms are processed and the selected responses (for

instance, number of total peaks, number of peaks with high resolution and low tailing) are imported back to the QbD software, which provides the statistical analyses for the responses selected (such as the Analysis of Variance, mathematical equations and surface response). (b) Three main steps were followed in the chromatographic method development: screening, optimization and robustness simulator. The designs, variables chosen, and fixed parameters are described. (c) Chromatograms with the higher responses obtained for the RSMS sample for each step are shown (275 nm), such as the conditions employed.

Table S2. Description of the columns selected for the chromatographic method development.

Column	Particle	Bonded phase	pH range	Max temperature (˚C)	Particle size (µm)
Cortecs	Solid Core	C_8	2–8	45	1.6
Acquity HSS	High Strength Silica	T3 (C_{18})	2–8	45	1.8
Acquity BEH	Ethylene Bridged Hybrid	Shield RP18 (C_{18})	2–11	High pH: 45; Low pH: 50	1.7
Acquity HSS	High Strength Silica	PFP (pentafluorophenyl)	2–8	45	1.8
Cortecs	Solid Core	C_{18}^+	2–8	45	1.6
Acquity BEH	Ethylene Bridged Hybrid	C_{18}	2–12	High pH: 60; Low pH: 80	1.7

Table S3. Factors levels, coded values and results of the experimental design for the **screening** step. Constant parameters: temperature (40 ˚C), injection volume (2 µL), flow rate (0.35 µL min^{-1}).

Run	Variables				Responses			
	Organic Solvent	**Gradient time (min)**	**pH**	**Column**	**Total number of peaks**	**Number of peaks with resolution ≥ 1.5**	**Number of peaks with resolution ≥ 2.0**	**Number of peaks with tailing ≤ 1.2**
1	CH_3CN	18.8	2.54	Acquity HSS T3	58	32	18	35
2	CH_3CN	26.6	2.54	Cortecs C_{18}^+	64	40	26	43
3	CH_3CN	30.0	2.54	Acquity BEH Shield RP18	69	42	26	53
4	CH_3CN	15.0	2.54	Acquity BEH Shield RP18	61	39	25	47
5	CH_3CN	30.0	2.54	Acquity BEH C_{18}	64	33	20	49
6	CH_3CN	15.0	2.54	Acquity BEH C_{18}	59	27	18	45
7	CH_3CN	22.5	2.54	Acquity HSS PFP	47	30	22	28
8	CH_3CN	15.0	2.54	Acquity BEH Shield RP18	61	37	23	42
9	CH_3CN	22.5	6.60	Cortecs C_8	47	28	18	29
10	CH_3CN	22.5	6.60	Acquity HSS T3	54	31	18	37
11	CH_3CN	22.5	6.60	Acquity BEH Shield RP18	48	30	20	30
12	CH_3CN	22.5	6.60	Acquity HSS PFP	35	24	14	23
13	CH_3CN	22.5	6.60	Cortecs C_{18}^+	46	27	17	28
14	CH_3CN	22.5	6.60	Acquity BEH C_{18}	45	31	16	32
15	CH_3CN	30.0	6.60	Cortecs C_8	49	31	23	34
16	CH_3CN	15.0	6.60	Cortecs C_8	39	23	14	27
17	CH_3CN	30.0	6.60	Acquity HSS T3	58	32	21	40
18	CH_3CN	15.0	6.60	Acquity HSS T3	49	25	13	36
19	CH_3CN	30.0	6.60	Acquity HSS PFP	40	25	17	28
20	CH_3CN	15.0	6.60	Acquity HSS PFP	41	20	14	26
21	CH_3CN	30.0	6.60	Cortecs C_{18}^+	45	33	25	26
22	CH_3CN	15.0	6.60	Cortecs C_{18}^+	43	20	17	29

23	CH_3CN	22.5	6.60	Cortecs C_8	45	27	20	30
24	CH_3CN	22.5	6.60	Acquity HSS T3	57	34	18	36
25	CH_3CN	22.5	6.60	Acquity BEH Shield RP18	43	27	20	29
26	CH_3CN	22.5	6.60	Acquity HSS PFP	38	19	17	20
27	CH_3CN	22.5	6.60	Cortecs C_{18}^+	44	28	16	23
28	CH_3CN	22.5	6.60	Acquity BEH C_{18}	49	30	18	36
29	CH_3CN	18.8	10.45	Acquity BEH Shield RP18	47	32	25	36
30	CH_3CN	30.0	10.45	Acquity BEH Shield RP18	43	26	24	37
31	CH_3CN	22.5	10.45	Acquity BEH C_{18}	43	29	22	33
32	MeOH	26.3	2.54	Acquity BEH Shield RP18	57	33	22	45
33	MeOH	30.0	2.54	Acquity HSS T3	55	34	24	45
34	MeOH	15.0	2.54	Acquity HSS T3	56	27	15	42
35	MeOH	30.0	2.54	Acquity HSS PFP	45	31	24	35
36	MeOH	15.0	2.54	Acquity HSS PFP	51	24	15	39
37	MeOH	30.0	2.54	Cortecs C_{18}^+	53	40	31	36
38	MeOH	15.0	2.54	Cortecs C_{18}^+	56	30	17	37
39	MeOH	22.5	2.54	Cortecs C_8	57	40	24	45
40	MeOH	30.0	2.54	Acquity HSS T3	55	37	25	45
41	MeOH	15.0	2.54	Cortecs C_{18}^+	58	34	21	36
42	MeOH	22.5	6.60	Cortecs C_8	44	30	16	32
43	MeOH	22.5	6.60	Acquity HSS T3	47	24	18	33
44	MeOH	22.5	6.60	Acquity BEH Shield RP18	37	24	16	26
45	MeOH	22.5	6.60	Acquity HSS PFP	39	24	19	30
46	MeOH	22.5	6.60	Cortecs C_{18}^+	43	28	22	33
47	MeOH	22.5	6.60	Acquity BEH C_{18}	42	22	15	32
48	MeOH	22.5	6.60	Cortecs C_8	40	27	18	30
49	MeOH	22.5	6.60	Acquity HSS T3	50	28	19	37
50	MeOH	22.5	6.60	Acquity BEH Shield RP18	32	22	16	21

51	MeOH	22.5	6.60	Acquity HSS PFP	35	20	16	25
52	MeOH	22.5	6.60	Cortecs C_{18}^{+}	39	28	19	29
53	MeOH	22.5	6.60	Acquity BEH C_{18}	40	24	17	32
54	MeOH	30.0	10.45	Acquity BEH Shield RP18	40	29	21	32
55	MeOH	15.0	10.45	Acquity BEH Shield RP18	43	27	19	29
56	MeOH	30.0	10.45	Acquity BEH C_{18}	46	29	20	39
57	MeOH	15.0	10.45	Acquity BEH C_{18}	50	26	19	42

Table S4. Regression ANOVA statistics obtained for the **screening step** of the chromatographic method development for the responses selected.

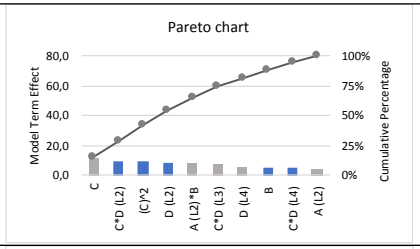
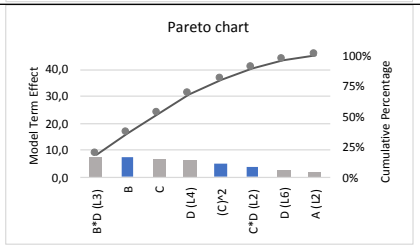
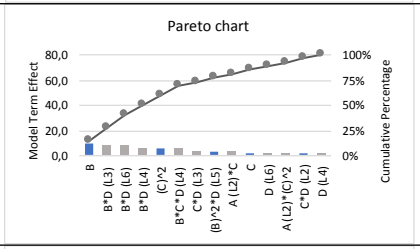
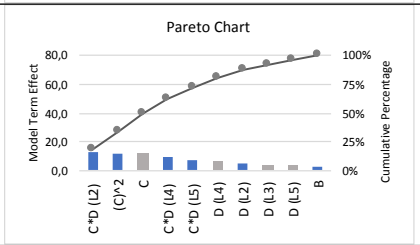
Response (Screening)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	3,504.9995	10	350.4999	41.8169	<0.0001
	Residual	385.5619	46	8.3818		
*MSR: 0.9009; $MSR_{threshold}$: 0.0440	*Lack-of-fit*	307.0619	31	9.9052	1.8927	0.0956
	Pure error	78.5000	15	5.2333		
**MS-LOF: 0.0255; MS-$LOF_{threshold}$: 0.0301	*Total*	3,890.5614	56			
Number of peaks with resolution ≥ 1.5	*Regression*	1,281.7256	8	160.2157	21.4063	<0.0001
	Residual	359.2568	48	7.4845		
	Lack-of-fit	297.2568	33	9.0078	2.1793	0.0550
*MSR: 0.7811; $MSR_{threshold}$: 0.0780	*Pure error*	62.000	15	4.1333		
**MS-LOF: 0.0439; MS-$LOF_{threshold}$: 0.0450	*Total*	1,640.9825	56			
Number of peaks with resolution ≥ 2.0	*Regression*	693.8530	14	49.5609	19.5713	<0.0001
	Residual	106.3575	42	2.5323		
	Lack-of-fit	73.3575	27	2.7169	1.2350	0.3412
*MSR: 0.8671; $MSR_{threshold}$: 0.0857	*Pure error*	33.0000	15	2.2000		
**MS-LOF: 0.0475; MS-$LOF_{threshold}$: 0.0872	*Total*	800.2105	56			
Number of peaks with tailing ≤ 1.2	*Regression*	2,462.2860	10	246.2286	20.1102	<0.0001
	Residual	563.2227	46	12.2440		
	Lack-of-fit	480.7227	31	15.5072	2.8195	0.0184
*MSR: 0.8138; $MSR_{threshold}$: 0.0827	*Pure error*	82.5000	15	5.5000		
**MS-LOF: 0.0513; MS-$LOF_{threshold}$: 0.0407$^{\phi}$	*Total*	3,025.5088	56			

*Modeling goal: MSR ≥ $MSR_{threshold}$.

**Modeling goal: MS-LOF ≤ MS-$LOF_{threshold}$.

$^{\phi}$LOF is statistically significant (P-value < 0.0500).

Box S5. Equations and Pareto chart obtained for the responses selected for the **screening step** during the chromatographic method development.

Response (Screening)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks		
Number of peaks with resolution ≥ 1.5		
Number of peaks with resolution ≥ 2.0		
Number of peaks with tailing ≤ 1.2		

*A: strong solvent type; A (L2): methanol; B: gradient time; C: pH; D: column type; D (L2): Acquity HSS T3; D (L3): Acquity BEH Shield RP18; D (L4): Acquity HSS PFP; D (L5): Cortecs C_{18}^{+}; D (L6): Acquity BEH C_{18}.

**Blue: positive effects; Grey: negative effects.

Table S6. Factor levels, coded values and results of the experimental design for the **first optimization** step. Constant parameters: column (Acquity BEH Shield RP18 – Ethylene Bridged Hybrid C_{18}, 1.7 µm, 100 x 2.1 mm), gradient time (30 min), injection volume (2 µL), flow rate (0.35 µL min^{-1}).

Run	Variables				Responses			
	Final % organic solvent	CH_3CN % organic solvent	Oven temperature (˚C)	pH	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
1	80	100	35	2.26	64	39	21	51
2	80	70	35	2.26	63	42	31	49
3	60	100	35	2.26	71	49	32	57
4	60	70	35	2.26	58	43	32	44
5	60	70	35	2.26	56	41	31	43
6	80	70	35	2.26	59	42	29	49
7	60	100	35	2.26	74	46	28	61
8	70	85	35	2.57	67	46	29	55
9	60	70	35	3.07	51	38	30	42
10	80	70	35	3.07	60	38	30	46
11	60	100	35	3.07	64	44	30	48
12	80	100	35	3.07	68	40	26	48
13	70	85	40	2.26	68	41	29	50
14	60	85	40	2.57	70	40	28	58
15	80	85	40	2.57	66	41	26	50
16	70	70	40	2.57	60	44	29	48
17	70	100	40	2.57	66	38	27	53
18	70	85	40	2.57	68	44	27	59
19	70	85	40	2.57	67	39	30	56
20	70	85	40	2.57	67	39	29	56
21	70	85	40	3.07	63	41	27	51
22	60	70	45	2.26	63	43	33	48
23	80	70	45	2.26	71	42	30	57
24	60	100	45	2.26	71	44	27	57
25	80	100	45	2.26	74	40	26	51
26	70	85	45	2.57	66	46	29	56
27	60	70	45	3.07	53	37	31	45
28	80	70	45	3.07	55	38	27	44
29	60	100	45	3.07	61	37	23	53
30	80	100	45	3.07	63	33	21	48

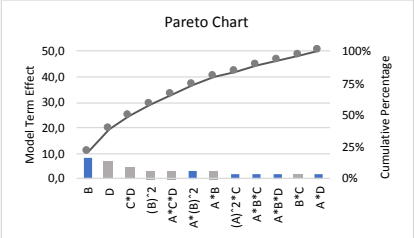
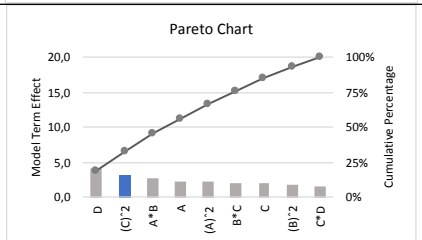
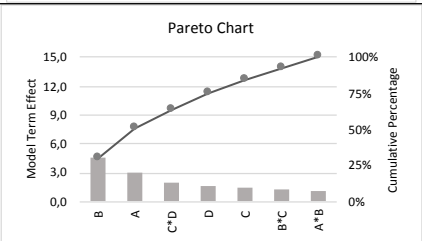
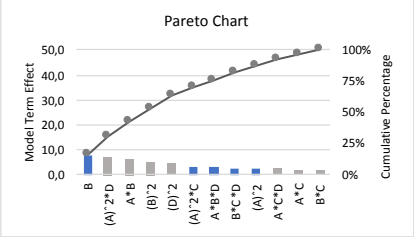
Table S7. Regression ANOVA statistics obtained for the **first optimization step** of the chromatographic method development for the responses selected.

Response (1st optimization)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	968.3290	12	80.6941	33.4278	<0.0001
	Residual	41.0377	17	2.4140		
*MSR: 0.9593; $MSR_{threshold}$: 0.0683	*Lack-of-fit*	25.8710	12	2.1559	0.7107	0.7102
	Pure error	15.1667	5	3.0333		
**MS-LOF: 0.0256; MS-$LOF_{threshold}$: 0.1687	*Total*	1,009.3667	29			
Number of peaks with resolution ≥ 1.5	*Regression*	247.4675	9	27.4964	6.4927	0.0003
	Residual	84.6991	20	4.2350		
	Lack-of-fit	61.5325	15	4.1022	0.8854	0.6130
*MSR: 0.7450; $MSR_{threshold}$: 0.2746	*Pure error*	23.1667	5	4.6333		
**MS-LOF: 0.1111; MS-$LOF_{threshold}$: 0.5798	*Total*	332.1667	29			
Number of peaks with resolution ≥ 2.0	*Regression*	189.6903	7	27.0986	10.2476	<0.0001
	Residual	58.1764	22	2.6444		
	Lack-of-fit	43.0097	17	2.5300	0.8341	0.6488
*MSR: 0.7653; $MSR_{threshold}$: 0.1840	*Pure error*	15.1667	5	3.0333		
**MS-LOF: 0.0714; MS-$LOF_{threshold}$: 0.3932	*Total*	247.8667	29			
Number of peaks with tailing ≤ 1.2	*Regression*	0.0000	12	0.0000	54.1882	<0.0001
	Residual	<0.0001	16	<0.0001		
	Lack-of-fit	0.0000	11	0.0000	0.7459	0.6826
*MSR: 0.9760; $MSR_{threshold}$: 0.0437	*Pure error*	0.0000	5	0.0000		
**MS-LOF: 0.0163; MS-$LOF_{threshold}$: 0.1027	*Total*	0.0000	28			

*Modeling goal: $MSR \geq MSR_{threshold}$.

**Modeling goal: MS-LOF ≤ $MS\text{-}LOF_{threshold}$.

Box S8. Equations and Pareto chart obtained for the responses selected for the **first optimization step** during the chromatographic method development.

Response (1st optimization)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks		
Number of peaks with resolution ≥ 1.5		
Number of peaks with resolution ≥ 2.0		
Number of peaks with tailing ≤ 1.2		

*A: final percentage of organic solvent; B: CH_3CN/MeOH ratio; C: oven temperature; D: pH.
**Blue: positive effects; Grey: negative effects.

Table S9. Factor levels, coded values and results of the experimental design for the **second optimization** step. Constant parameters: column (Acquity BEH Shield RP18 – Ethylene Bridged Hybrid C_{18}, 1.7 µm, 100 x 2.1 mm), gradient time (30 min), pH (2.26), temperature (45 ˚C), injection volume (2 µL), flow rate (0.35 µL min^{-1}).

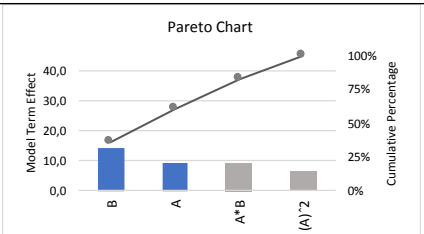
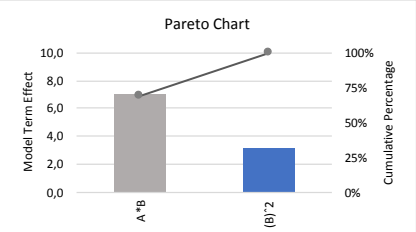
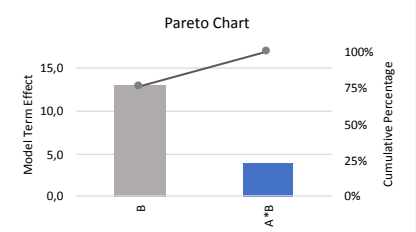
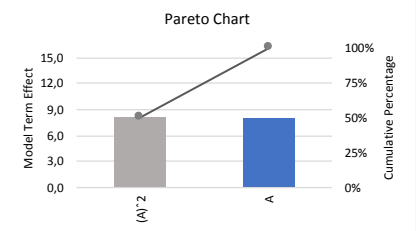
Run	Variables		Responses			
	Final % organic solvent	**CH_3CN % organic solvent**	**Total number of peaks**	**Number of peaks with resolution ≥ 1.5**	**Number of peaks with resolution ≥ 2.0**	**Number of peaks with tailing ≤ 1.2**
1	55	100	92	51	25	69
2	55	80	91	51	31	76
3	40	80	78	54	32	60
4	70	100	90	51	28	69
5	70	80	86	53	32	66
6	40	60	65	50	39	53
7	55	60	80	58	40	69
8	55	80	90	48	30	72
9	55	80	89	49	31	71
10	70	60	84	57	36	67
11	40	100	89	58	34	65

Table S10. Regression ANOVA statistics obtained for the **second optimization step** of the chromatographic method development for the responses selected.

Response (2nd optimization)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	617.3758	4	154.3439	36.2688	0.0002
	Residual	25.5333	6	4.2556		
	Lack-of-fit	23.5333	4	5.8833	5.8833	0.1505
*MSR: 0.9603; $MSR_{threshold}$: 0.1200	*Pure error*	2.0000	2	1.0000		
	Total	642.9091	10			
Number of peaks with resolution ≥ 1.5	*Regression*	76.3485	2	38.1742	5.8918	0.0267
	Residual	51.8333	8	6.4792		
	Lack-of-fit	47.1667	6	7.8611	3.3690	0.2465
*MSR: 0.5956; $MSR_{threshold}$: 0.4508	*Pure error*	4.6667	2	2.333		
	Total	128.1818	10			
Number of peaks with resolution ≥ 2.0	*Regression*	189.1769	2	94.5885	71.7894	<0.0001
	Residual	9.2231	7	1.3176		
	Lack-of-fit	8.5564	5	1.7113	5.1338	0.1710
*MSR: 0.9535; $MSR_{threshold}$: 0.0629	*Pure error*	0.6667	2	0.333		
	Total	198.4000	9			
Number of peaks with tailing ≤ 1.2	*Regression*	273.4667	2	136.7333	9.8963	0.0069
	Residual	110.5333	8	13.8167		
	Lack-of-fit	96.5333	6	16.0889	2.2984	0.3339
*MSR: 0.7122; $MSR_{threshold}$: 0.3209	*Pure error*	14.0000	2	7.0000		
	Total	384.0000	10			

*Modeling goal: $MSR \geq MSR_{threshold}$.

Box S11. Equations and Pareto chart obtained for the responses selected for the **second optimization step** in the chromatographic method development.

Response (2nd optimization)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks		
Number of peaks with resolution ≥ 1.5		
Number of peaks with resolution ≥ 2.0		
Number of peaks with tailing ≤ 1.2		

*A: final percentage of organic solvent; B: CH_3CN/MeOH ratio.
**Blue: positive effects; Grey: negative effects.

Table S12. Metabolites identified by UHPLC-ESI-MS^2 and Molecular Network in the extracts present in the RSMS sample, in the positive ionization mode.

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	error (ppm)	MS/MS fragments (% abundance)	Metabolite
1	1.81	*M. bahiana*	$C_{14}H_{17}O_{10}$	345.0832	345.0822	2.9	153.0202 (100)	Galloyl quinic acid
2	5.25	*B. intermedia*	$C_{15}H_{15}O_6$	291.0877	291.0869	2.7	207.0675 (20), 165.0560 (15), 147.0459 (20), 139.0403 (100), 123.0459 (50)	Catechin
2	5.25	*A. septentrionalis*	$C_{15}H_{15}O_6$	291.0874	291.0869	1.7	207.0667 (15), 165.0555 (15), 147.0462 (20), 139.0411 (100), 123.0459 (35)	Catechin
3	5.55	*A. septentrionalis*	$C_{30}H_{27}O_{12}$	579.1505	579.1503	0.3	409.0920 (30), 287.0572 (80), 275.0572 (60), 247.0611 (50), 233.0461 (30), 163.0416 (50), 139.0410 (70), 127.0390 (100), 123.0457 (40)	Proanthocyanidin dimer
4	5.77	*B. intermedia*	$C_{30}H_{27}O_{12}$	579.1501	579.1503	-0.3	409.0939 (40), 301.0712 (15), 287.0581 (85), 271.0616 (45), 247.0648 (40), 233.0464 (15), 191.0378 (15), 163.0419 (40), 139.0420 (60), 127.0412 (100), 123.0467 (30)	Proanthocyanidin dimer
5	6.04	*A. septentrionalis*	$C_{30}H_{27}O_{12}$	579.1508	579.1503	0.9	409.0963 (70), 301.0708 (20), 287.0569 (100), 275.0569 (60), 247.0605 (40), 233.0446 (20), 163.0405 (45), 139.0428 (60), 127.0405 (55), 123.0458 (35)	Proanthocyanidin dimer
6	6.65	*B. intermedia*	$C_{15}H_{15}O_6$	291.0877	291.0869	2.7	207.0670 (15), 147.0459 (25), 139.0408 (100), 123.0462 (40)	Epicatechin
6	6.65	*B. laevifolia*	$C_{15}H_{15}O_6$	291.0870	291.0869	0.3	207.0670 (10), 165.0558 (100), 147.0466 (20), 139.0408 (100), 123.0465 (40)	Epicatechin
6	6.66	*N. multiglandulosa*	$C_{15}H_{15}O_6$	291.0882	291.0869	4.5	207.0660 (15), 165.0559 (15), 147.0463 (20), 139.0408 (100), 123.0457 (35)	Epicatechin
7	7.23	*N. multiglandulosa*	$C_{30}H27O_{12}$	579.1499	579.1503	-0.7	409.0945 (45), 301.0731 (35), 287.0569 (90), 275.0583 (70), 247.0613 (35), 233.0454 (20), 163.0408 (70), 139.0403 (70), 127.0413 (100), 123.0779 (35)	Proanthocyanidin dimer
7	7.23	*B. laevifolia*	$C_{30}H_{27}O_{12}$	579.1505	579.1503	0.3	409.0950 (30), 287.0583 (70), 271.0630 (40), 247.0613 (30), 233.0455 (20),	Proanthocyanidin dimer

							163.0404 (40), 139.0422 (60), 127.0408 (100), 123.0458 (30)	
8	7.25	*A. septentrionalis*	$C_{15}H_{15}O_5$	275.0924	275.0919	1.8	201.0483 (10), 191.0726 (15), 149.0616 (20), 139.0405 (100), 107.0509 (25)	Afzelechin
9	7.70	*A. septentrionalis*	$C_{27}H_{31}O_{15}$	595.1657	595.1663	-1.0	541.1246 (20), 523.1274 (20), 481.1144 (40), 457.1113 (45), 427.1031 (65), 409.0929 (75), 379.0817 (90), 337.0736 (55), 325.0726 (100), 307.0597 (45), 295.0616 (30)	Genistein-di-*C*-hexoside
10	8.81	*B. intermedia*	$C_{21}H_{19}O_{13}$	479.0831	479.0826	1.0	309.0630 (15), 153.0200 (100)	Digalloyl shikimic acid
11	8.84	*B. intermedia*	$C_{21}H_{21}O_{14}$	497.0938	497.0931	1.4	309.0636 (10), 153.0199 (100)	Digalloyl quinic acid
12	8.87	*B. maritima*	$C_{39}H_{51}O_{23}$	887.2811	887.2821	-1.1	287.0568 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside-deoxyhexoside
13	8.88	*B. intermedia*	$C_{28}H_{25}O_{18}$	649.1038	649.1041	-0.5	309.0635 (10), 153.0201 (100)	Trigalloyl quinic acid
14	9.11	*B. laevifolia*	$C_{45}H_{39}O_{18}$	867.2119	867.2136	-2.9	545.1077 (10), 527.1068 (15), 419.0790 (30), 409.0927 (60), 407.0764 (55), 393.0641 (30), 301.072 (30), 289.0740 (40), 287.0574 (60), 275.0552 (95), 247.0612 (100), 245.0452 (90), 163.0412 (80), 139.0418 (80), 127.0418 (70), 123.0496 (40)	Proanthocyanidin trimer
14	9.12	*N. multiglandulosa*	$C_{45}H_{39}O_{18}$	867.2115	867.2136	-2.4	545.1047 (15), 527.0963 (25), 409.0902 (65), 407.0797 (40), 393.0627 (20), 301.0681 (25), 289.0714 (55), 287.0561 (70), 247.0615 (100), 245.0453 (90), 163.0425 (75), 127.0416 (70), 123.0459 (45)	Proanthocyanidin trimer
15	9.46	*B. maritima*	$C_{33}H_{41}O_{20}$	757.2178	757.2191	-1.7	303.1518 (100)	Quercetin-*O*-hexoside-deoxyhexoside-deoxyhexoside
16	10.11	*H. restingae*	$C_{27}H_{31}O_{14}$	579.1713	579.1714	-0.2	433.1136 (100), 415.1046 (55), 397.0939 (40), 379.0836 (10), 367.0835 (20), 337.0737 (15), 313.0727 (60), 283.0600 (10)	Apigenin-*C*-hexoside-deoxyhexoside

17	10.20	*B. intermedia*	$C_{28}H_{23}O_{17}$	631.0937	631.0935	0.3	153.0201 (100)	Trigalloyl shikimic acid
18	10.31	*B. maritima*	$C_{33}H_{41}O_{19}$	741.2253	741.2242	1.5	287.0567 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
19	10.74	*P. densiflora*	$C_{27}H_{31}O_{15}$	595.1664	595.1663	0.2	287.0567 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
19	10.74	*B. maritima*	$C_{27}H_{31}O_{15}$	595.1670	595.1663	1.2	287.0566 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
20	10.76	*B. maritima*	$C_{33}H_{41}O_{19}$	741.2241	741.2242	-0.1	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
21	10.86	*B. harleyi*	$C_{33}H_{41}O_{20}$	757.2187	757.2191	-0.5	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside-deoxyhexoside
22	10.90	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1614	611.1612	0.3	303.0517 (100)	Quercetin-*O*-hexoside-deoxyhexoside
23	10.93	*N. multiglandulosa*	$C_{27}H_{43}O_{7}$	479.3013	479.3009	0.8	479.3017 (10), 461.2912 (30), 443.2816 (70), 425.2692 (100), 407.2585 (50), 389.2487 (10), 383.2599 (25), 351.1992 (10), 343.2313 (25), 311.2036 (25), 281.566 (30)	5-hydroxypodecdysone B
24	10.96	*N. multiglandulosa*	$C_{27}H_{45}O_{8}$	497.3119	497.3114	1.0	497.3134 (40), 479.3018 (10), 461.2907 (20), 443.2819 (70), 425.2698 (100), 407.2598 (45), 387.219 (20), 369.2086 (45), 351.1957 (65), 343.2288 (20), 311.2013 (25), 309.1983 (20), 281.1558 (30), 269.1534 (20)	Integristerone A
25	11.04	*N. multiglandulosa*	$C_{27}H_{43}O_{6}$	463.3063	463.3060	0.6	463.3047 (10), 445.2959 (100), 427.2848 (40), 409.2748 (30), 391.2643 (10), 371.2236 (15), 353.2130 (10), 329.117 (20), 301.1812 (80), 283.1714 (20), 165.287 (50)	Podecdysone B
26	11.06	*H. restingae*	$C_{27}H_{31}O_{14}$	579.1715	579.1714	0.3	433.1161 (20), 415.1041 (15), 397.0915 (30), 379.0824 (45), 367.0836 (45), 349.0751 (15), 337.0729 (90), 313.0727 (100), 283.0616 (85)	Apigenin-*C*-hexoside-deoxyhexoside

27	11.09	*N. multiglandulosa*	$C_{27}H_{45}O_7$	481.3162	481.3165	-0.6	481.1698 (10), 445.2972 (100), 427.2838 (55), 409.2762 (40), 371.2232 (70), 162.1285 (80)	Ecdysterone
28	11.19	*P. densiflora*	$C_{39}H_{51}O_{24}$	903.2742	903.2770	-3.1	287.0557 (100)	Kaempferol-*O*-hexoside-hexoside-deoxyhexoside-deoxyhexoside
29	11.22	*P. densiflora*	$C_{33}H_{41}O_{20}$	757.2161	757.2191	-4.0	287.0558 (100)	Kaempferol-*O*-hexoside-hexoside-deoxyhexoside
30	11.77	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1612	611.1612	0.0	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside
31	11.92	*P. densiflora*	$C_{33}H_{41}O_{19}$	741.2241	741.2242	-0.1	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
31	11.93	*B. harleyi*	$C_{33}H_{41}O_{19}$	741.2224	741.2242	-2.4	287.0559 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
32	12.01	*B. harleyi*	$C_{21}H_{21}O_{12}$	465.1036	465.1033	0.6	303.0513 (100)	Quercetin-*O*-hexoside
33	12.03	*M. bahiana*	$C_{27}H_{31}O_{16}$	611.1608	611.1612	-0.7	303.0517 (100)	Quercetin-*O*-hexoside-deoxyhexoside
34	12.04	*N. multiglandulosa*	$C_{32}H_{39}O_{19}$	727.2078	727.2086	-0.8	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-pentoside
35	12.05	*H. restingae*	$C_{28}H_{25}O_{16}$	617.1136	617.1143	-1.1	303.0513 (90), 297.0614 (15), 171.0312 (10), 153.0197 (100)	Quercetin-*O*-galloyl-hexoside
36	12.06	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1617	611.1612	0.8	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside
37	12.16	*H. restingae*	$C_{28}H_{25}O_{16}$	617.1126	617.1143	-2.8	303.0510 (100), 153,0196 (95)	Quercetin-*O*-galloyl-hexoside
38	12.27	*B. intermedia*	$C_{21}H_{21}O_{12}$	465.1026	465.1033	-1.5	303,0508 (100)	Quercetin-*O*-hexoside
38	12.27	*M. bahiana*	$C_{21}H_{21}O_{12}$	465.1038	465.1033	1.1	303,0514 (100)	Quercetin-*O*-hexoside
39	12.61	*N. multiglandulosa*	$C_{33}H_{41}O_{20}$	757.2170	757.2191	-2.8	317.0662 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside-pentoside
40	12.66	*P. densiflora*	$C_{50}H_{61}O_{29}$	1125.3295	1125.3299	-0.4	287.0564 (50), 207.0675 (100), 175.0408 (20)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-dimethoxyferulic acid

41	13.02	*B. intermedia*	$C_{28}H_{25}O_{16}$	617.1155	617.1143	1.9	303.0524 (30), 233.0470 (25), 205.0526 (20), 153.0203 (100)	Quercetin-*O*-galloyl-hexoside
42	13.41	*N. multiglandulosa*	$C_{28}H_{33}O_{16}$	625.1757	625.1769	-1.9	317.0670 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside
43	13.43	*B. harleyi*	$C_{26}H_{29}O_{15}$	581.1502	581.1506	-0.7	303.0516 (1000	Quercetin-*O*-deoxyhexoside-pentoside
44	13.44	*B. harleyi*	$C_{20}H_{19}O_{11}$	435.0935	435.0927	1.8	303.0506 (100)	Quercetin-*O*-pentoside
45	13.52	*B. harleyi*	$C_{27}H_{31}O_{15}$	595.1666	595.1663	0.5	287.0568 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
46	13.52	*P. densiflora*	$C_{49}H_{59}O_{28}$	1095.3199	1095.3193	0.5	287,0564 (50), 177.0569 (100), 145.0292 (10)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-methoxycaffeic acid
47	13.53	*A. septentrionalis*	$C_{27}H_{31}O_{15}$	595.1670	595.1663	1.2	287.0562 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
48	13.65	*B. intermedia*	$C_{20}H_{19}O_{11}$	435.0933	435.0927	1.4	303.0513 (100)	Quercetin-*O*-pentoside
49	13.70	*B. harleyi*	$C_{28}H_{33}O_{16}$	625.1766	625.1769	-0.5	317.0670 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside
50	14.06	*P. densiflora*	$C_{48}H_{57}O_{27}$	1065.3075	1065.3087	-1.1	287.0571 (80), 147.0458 (100)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-coumaric acid
51	14.10	*B. laevifolia*	$C_{22}H_{23}O_{12}$	479.1192	479.1190	0.4	317.0664 (100)	Methoxy-quercetin-*O*-hexoside
52	14.32	*B. laevifolia*	$C_{21}H_{21}O_{11}$	449.1085	449.1084	0.2	303.0512 (100)	Quercetin-*O*-deoxyhexoside
52	14.32	*M. bahiana*	$C_{21}H_{21}O_{11}$	449.1088	449.1084	0.9	303.0513 (100)	Quercetin-*O*-deoxyhexoside
53	14.54	*B. laevifolia*	$C_{22}H_{21}O_{13}$	493.0980	493.0982	-0.4	317.0673 (100)	Methoxy-quercetin-*O*-glucuronic acid
54	14.96	*B. harleyi*	$C_{20}H_{19}O_{10}$	419.0980	419.0978	0.5	287.0564 (100)	Kaempferol-*O*-pentoside
55	14.99	*B. harleyi*	$C_{26}H_{29}O_{14}$	565.1560	565.1557	0.5	287.0568 (100)	Kaempferol-*O*-deoxyhexoside-pentoside
56	15.40	*B. intermedia*	$C_{27}H_{23}O_{15}$	587.1040	587.1037	0.5	303.0490 (10), 267.0518 (35), 249.0437 (20), 231.0308 (20), 207.0311 (45), 205.0518 (35), 154.0200 (100)	Quercetin-*O*-galloyl-pentoside

57	15.60	*B. laevifolia*	$C_{25}H_{25}O_{15}$	565.1213	565.1193	3.5	317.0667 (100)	Methoxy-quercetin-*O*-malonyl-hexoside
58	16.33	*B. laevifolia*	$C_{15}H_{11}O_{17}$	463.1250	462.9996	0.9	317.0670 (100)	Methoxy-quercetin-*O*-hexoside
59	17.88	*M. bahiana*	$C_{28}H_{25}O_{15}$	601.1186	601.1193	-1.2	303.0513 (10), 153.0201 (100)	Quercetin-*O*-galloyl-deoxyhexoside
60	18.30	*A. septentrionalis*	$C_{24}H_{25}O_{13}$	521.1296	521.1295	0.2	273.0774 (100)	Naringenin-*O*-malonyl-hexoside
61	18.58	*M. bahiana*	$C_{28}H_{25}O_{15}$	601.1190	601.1193	-0.5	303.0508 (10), 153.0191 (100)	Quercetin-*O*-galloyl-deoxyhexoside
62	18.58	*M. bahiana*	$C_{13}H_{15}O_{8}$	299.0771	299.0767	1.3	153.0200 (100)	Not identified
63	18.58	*M. bahiana*	$C_{13}H_{13}O_{7}$	281.0666	281.0661	1.8	280.0967 (50), 153.0203 (100)	Not identified

Table S13. List of plant species used for the representative sample of Malpighiaceae species (RSMS) preparation: collection sites, dates and voucher codes, biomes and phylogenetic groups.

Species	Collection sites	Collection dates	Codes	Herbarium*	Biome	Phylogenetic group
Byrsonima intermedia	Mogi Guaçu/SP	Jan/2014	IAC 55281	IAC	Cerrado	Byrsonimoid
Mcvaughia bahiana	Monte Santo/BA	Jan/2006	Guedes 12148	ALCB	Atlantic Forest	Mcvaughioid
Barnebya harleyi	Itatim/BA	Oct/2014	Melo 1518	HUEFS	Caatinga	Barnebyoid
Ptilochaeta densiflora	Corumbá/MS	Apr/2010	Carvalho 290	HUEFS	Pantanal	Ptilochaetoid
Bunchosia maritima	Rio de Janeiro/RJ	Sep/2018	I.R.C. 183	RBv	Atlantic Forest	Bunchosioid
Hiraea restingae	Sooretama/ES	Jan/2012	Almeida 518	SP	Atlantic Forest	Hiraeoid
Niedenzuella multiglandulosa	Campo Grande/MS	Nov/2015	HMS 5206	CGMS	Cerrado	Tetrapteroid
Banisteriopsis laevifolia	Rio de Janeiro/RJ	Sep/2018	Mattos 317	RBv	Atlantic Forest	Stigmaphylloid
Amorimia septentrionalis	Maruim/SE	Nov/2015	Almeida 800	HUEFS	Atlantic Forest	Malpighioid

***IAC:** Agronomic Institute of Campinas; **ALCB:** Herbarium Alexandre Leal Costa – Federal University of Bahia; **HUEFS:** Herbarium of State University of Feira de Santana; **RBv:** Arboretum of Rio de Janeiro Botanical Garden (living collection); **SP:** Institute of Botany of São Paulo; **CGMS**: Herbarium of Federal University of Mato Grosso do Sul.

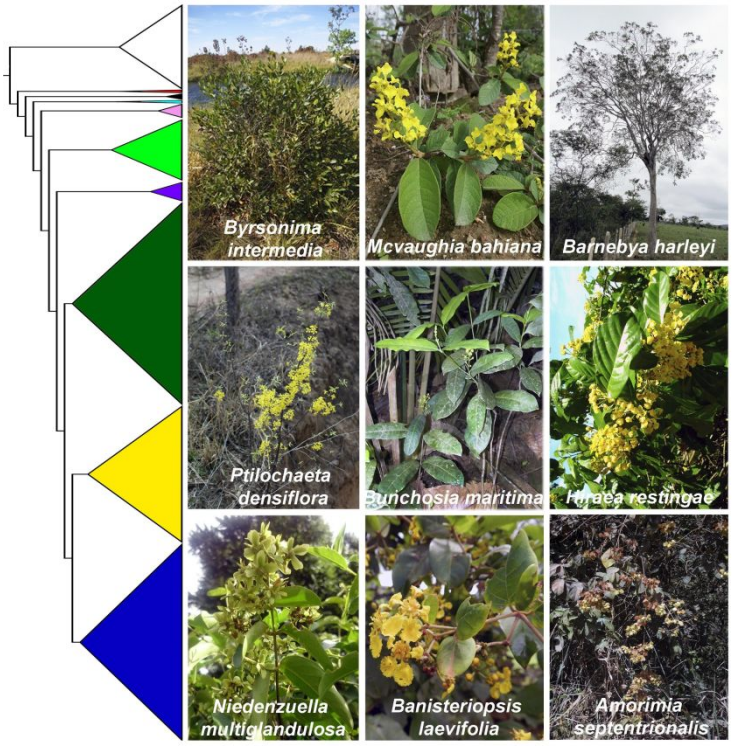


Figure S14. Field photographs of one species from each phylogenetic group form Malpighiaceae family. White - Byrsonimoid clade represented by the species *Byrsonima intermedia* (photograph by R.F. Almeida), Red - Acridocarpoid clade (not sampled in this study), Black - Mcvaughioid clade represented by the species *Mcvaughia bahiana* (photograph by I.R. Guesdon), Light Blue - Barnebyoid clade represented by the species *Barnebya harley*i (photograph by F. Flores), Pink - Ptilochaetoid clade represented by the species *Ptilochaeta densiflora* (photograph by R.F. Almeida), Light Green - Bunchosioid clade represented by the species *Bunchosia maritima* (photograph by J.M.Braga), Purple - Hiraeoid clade represented by the species *Hiraea restingae* (photograph by R.F. Almeida), Dark Green - Tetrapteroid clade represented by the species *Niedenzuella multiglandulosa* (photograph by N. Carvalho, Yellow - Stigmaphylloid clade represented by the species *Banisteriopsis laevifolia* (photograph by C.F. Hall), and Dark Blue - Malpighioid clade represented by the species *Amorimia septentrionais* (photograph by M.O.O. Pellegrini) (phylogenetic classification according to Davis and Anderson 2010).

S15. Method validation

Specificity

Specificity is the ability to the method to unequivocally assess and differentiate the analyte signals from potential compounds such as impurities, other matrix components or degradants, among others. In the present study, it was determined through the injection of the solvent solution containing only the internal standard (blank), the standards solution and the RSMS solution (n=3). Peaks retention times and resolution of (+)-catechin, (–)-epicatechin, ecdysterone and rutin present in the standard and sample solutions were used to calculate the relative standard deviation (RSD). In addition, the mass spectra of each compound present in both standard and RSMS solutions were recorded and used to confirm the specificity and identity of the peaks.

The detected peaks corresponded to (+)-catechin (4.88 min), (–)-epicatechin (6.25 min), ecdysterone (10.61 min) and rutin (11.56 min) in the RSMS solution, showed good resolutions (5.38, 1.98, 7.71 and 1.25, respectively) and did not show any interferences when compared to the standard solution. Moreover, the recorded mass spectra allowed the confirmation of each compound's identity.

Precision (repeatability and intermediate precision)

The precision was estimated by the analysis of six RSMS solutions, each one injected once (n=6). Repeatability (or intraday precision) was expressed as the RSD of (+)-catechin, (–)-epicatechin, ecdysterone and rutin amounts (concentration of injection, μg mL^{-1}) measured in two consecutive days. For the first day, the average concentrations measured for (+)-catechin, (–)-epicatechin, ecdysterone and rutin were 10.101, 13.644, 52.377 and 23.672 μg mL^{-1}, respectively. For the second day, the average concentrations measured for (+)-catechin, (–)-epicatechin, ecdysterone and rutin were 10.446, 13.829, 52.545 and

23.605µg mL^{-1}, respectively. For intermediate precision, the results of the two days were compared through F-test.

In both days, the RSDs determined for all compounds were lower than 1% (Table 1). The intermediate precision, calculated by the F-test between the two different days, was also very satisfactory since no significant difference at F= 0.05 (n=6-1) was detected.

Linearity

To determine the linearity of the method at the PDA detector, the calibration curves were prepared in the concentration range expected for each compound in RSMS. It was determined by elaborating calibration curves of each compound ranging from 2.2 to 278% of the working standards concentrations (n=3). Internal standardization was also used to improve method confidence. For that, a stock standard solution containing (+)-catechin (100 µg mL^{-1}), (–)-epicatechin (150 µg mL^{-1}), ecdysterone (776 µg mL^{-1}), and rutin (300 µg mL^{-1}) in MeOH/H_2O 1:1, was prepared for simultaneous acquisition of the analytical curves. From the stock solution, eleven concentration levels for each compound were prepared (in MeOH/H_2O 1:1), ranging from (i) 0.2 to 25 µg mL^{-1} for (+)-catechin, (ii) 0.3 to 37.5 µg mL^{-1} for (–)-epicatechin, (iii) 1.6 to 194 µg mL^{-1} for ecdysterone, and (iv) 0.6 to 75 µg mL^{-1} for rutin. All solutions contained the internal standard sodium diclofenac at 15 µg mL^{-1}. Each calibration concentration sample was injected in three replicates. The linearity was calculated based on the analytical curves built with the nominal concentration of each calibration point and the corresponded average values of the ratios between the area of each calibration point divided by the area of the internal standard ($R=A_{compound}/A_{IS}$). The results were interpreted in function of the correlation coefficients (R) calculated for each compound.

The linear ranges of concentrations obtained were adequate for all compounds, and the obtained values of R coefficients were for 0.9996, 0.9994, 0.9995, 0.9994 for (+)-catechin, (–)-epicatechin, ecdysterone and rutin, respectively (Table 1).

Limits of detection (LOD) and limits of quantification (LOQ)

Limits of detection (LODs) and quantification (LOQs) were estimated from the standard deviation of the y-intercept (Sb) and the slope (a) of three calibration curves prepared in three low concentrations. The mathematical calculations were performed using the following equations: $LOD = 3.3Sb.a^{-1}$ and $LOQ = 10Sb.a^{-1}$, where *a* is the mean of slopes of the calibration curves and *Sb* is the SD of the y-intercept. For that, the same stock solution prepared for the determination of linearity was used. The three concentration levels for each compound were also prepared in MeOH/H_2O 1:1, ranging from (i) 0.2 to 1 μg mL^{-1} for (+)-catechin, (ii) 0.3 to 1.5 μg mL^{-1} for (–)-epicatechin, (iii) 1.6 to 7.8 μg mL^{-1} for ecdysterone, and (iv) 0.6 to 3 μg mL^{-1} for rutin. All solutions also contained the internal standard sodium diclofenac at 15 μg mL^{-1} and were injected in triplicate.

The results obtained for the LOD and LOQ for (+)-catechin, (–)-epicatechin, ecdysterone and rutin were 0.06, 0.02, 0.07 and 0.03 μg mL^{-1} and 0.18, 0.06, 0.22 and 0.11 μg mL^{-1}, respectively (Table 1).

Accuracy

Accuracy was determined by recovery studies, which was performed by standard addition of the analyte in the RSMS solution, in three different concentrations, considering the specified range of the analytical procedure. For that, previously analyzed standard and RSMS solutions were used to prepare three different concentration levels by spiking known amounts of the stock standards solution into the RSMS solution. Three replicates for each

level (low, intermediate and high, or at 66.7, 100.0 and 166.7 % levels, respectively) were prepared to obtain solutions containing 2.99, 8.99 and 14.99 μg mL^{-1} of (+)-catechin, 4.18, 13.18 and 22.18 μg mL^{-1} of (–)-epicatechin, 17.99, 64.55 and 111.11 μg mL^{-1} of ecdysterone and 7.64, 25.64 and 43.64 μg mL^{-1} of rutin. The percentage of recovery of each compound was analyzed using the validated method and the accuracy was calculated by the difference between the nominal and the experimentally measured contents.

Results show an average of recovery corresponded to 86.72% (low level), 88.71% (intermediate level) and 93.73% (high level), taking into account all compounds in each level. Considering each compound separately, the RSD was lower than 5% (Table 1).