Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary tables

Supplementary Table 1. Avocado cultivars analyzed in this study.

Cultivar	Genotype[1]	Predicted Genotype[2]	Individual acetogenin concentration[3] (mg/g fresh weight)						
			Pulp					Peel	
			Persenone C	UPA	Persenone B	AcO-avocadene	Persenone A	Persenone C	Persenone B
Comcar 1	Guatemalan	Guatemalan	0.02±0	N.D.	0.09±0.01	0.01±0	0.3±0.02	0.02±0	0.11±0.01
Larrainzar 1	Guatemalan	Guatemalan	0.06±0	N.D.	0.11±0	0.09±0	0.24±0.01	0.03±0	0.05±0
Olanca 2 clon	Guatemalan	Guatemalan	0.16±0.02	N.D.	0.47±0.01	0.67±0.05	2.22±0.18	0.06±0.02	0.2±0.03
Reed	Guatemalan	Guatemalan	0.1±0.02	0.04±0.01	0.12±0.03	0.19±0.04	0.66±0.18	0.02±0	0.03±0.01
257PTB	Hybrid	Hybrid	0.09±0.01	0.03±0	0.48±0.02	0.19±0.01	1.74±0.1	0.06±0.02	0.3±0.07
264PTB	Hybrid	Hybrid (G)	0.18±0.05	0.02±0	0.58±0.13	0.92±0.27	6.1±2.09	0.13±0.01	0.47±0.05
Encinos	Hybrid	Hybrid	0.26±0.06	0.1±0.01	0.68±0.12	0.67±0.19	2.83±0.87	0.03±0	0.05±0.01
Hass	Hybrid	Hybrid	0.26±0.05	0.24±0.04	0.34±0.05	0.51±0.06	1.63±0.37	0.05±0	0.13±0.03
Jimenez	Hybrid	Hybrid	0.24±0.01	0.25±0.01	0.36±0.02	0.3±0.04	1.26±0.16	0.01±0	0.05±0.01
Ag.Negro	Mexican	Mexican	0.17±0.02	0.02±0	0.37±0.05	0.21±0.03	1.03±0.15	0.32±0.05	0.62±0.09
Almoloya clon	Mexican	Mexican	0.07±0.01	0.04±0	0.3±0.03	0.13±0.01	1.01±0.09	0.07±0.02	0.23±0.05
Aquijic S2	Mexican	Mexican (G)	0.09±0.01	0.02±0	0.09±0.01	0.12±0.01	0.39±0.03	0.07±0	0.07±0
Mantequilla 1	Mexican	Mexican	0.03±0	N.D.	0.18±0.02	0.03±0	0.92±0.08	0.03±0	0.13±0.01
Pintle 2	Mexican	Mexican	0.12±0	0.01±0	0.34±0.01	0.11±0.01	1.21±0.07	0.04±0.01	0.11±0.03
Vargas	Mexican	Mexican	0.34±0.05	0.05±0.01	0.86±0.08	0.33±0.06	1.93±0.33	0.05±0.01	0.12±0.02
Ariete	Unknown	Guatemalan (M)	0.24±0	0.04±0	0.61±0.02	0.44±0.02	1.57±0.08	0.07±0.01	0.18±0.03
Pionero	Unknown	Hybrid	0.34±0.02	0.1±0.01	0.98±0.03	0.83±0.04	4.31±0.24	0.01±0	0.05±0.01
Aguilar	Unknown	Hybrid (M)	0.12±0	0.08±0.01	0.23±0.01	0.31±0.01	2.02±0.07	0.05±0	0.12±0.01
L14NE	Unknown	Mexican	0.29±0.07	0.05±0.01	0.71±0.22	0.4±0.09	2.14±0.52	0.3±0.06	0.62±0.1
Fundación 2	Unknown	Mexican (G)	0.11±0.01	0.07±0.01	0.24±0.01	0.23±0.04	1.1±0.1	0.01±0	0.03±0.01
L14Ch	Unknown	Mexican (G)	0.03±0	N.D.	0.12±0.03	0.07±0.01	0.33±0.12	0.09±0.03	0.18±0.03
Aries	Unknown	Mexican (H)	0.26±0.05	0.08±0.01	0.71±0.11	0.42±0.11	1.24±0.32	0.05±0.01	0.12±0.03

Values are means of three independent fruit samples, ± indicate SE. *N.D. Not detected

1 As reported by CICTAMEX (L. López-López, A. F. Barrientos Priego and A. D. Ben-Ya'acov, Revista Chapingo, Serie Horticultura, 1999, 5, 19-23.)

2 Based on acetogenin chemotype by the model generated in this work. Parenthesis indicates conflicting assignation.

3 Acetogenins relevant for classification by the model (all acetogenin contents are in Figure 3)

Supplementary Table 2. Summary of Mass fragmentation patterns of acetogenins present in avocado samples. Calculated monoisotopic masses of the molecular positive ions$[M+H]^+$ are shown, along with the fragments obtained in positive mode, as they appear in a sample, and the fragments reported in the available literature. Fragments marked in bold and underlined are fragments that correspond to the fragments detected in the samples.

	Calculated m/z $[M+H]^+$	Detected fragments	Reported fragments	Reference
AcO-avocadenyne **(1)**	325.2379	347.2127, 265.2112, 247.2076	325, 307, **265**, **247**, 147	1
AcO-avocadene **(2)**	329.2692	351.2621	**351**, 311, 259, 251	2
Unknown Putative Acetogenin **(3)**	327.2535	349.2255, 267.2242, 249.2218	**349** **349**, 287, 166	1
Persediene **(4)**	353.2692	375.2556, 293.2517, 275.2435	**375**, 335, **293**, **275**	3, 4
Persenone C **(5)**	353.2692	375.2593, 275.2435	**375**, 335, 293	3, 4
Persenone A **(6)**	379.2848	401.2701, 379.2880, 361.2776, 319.2664, 301.2564, 283.2448	**401**, **361**, **319**, **301**	2
Persenone B **(8)**	355.2848	377.2677, 295.2643, 277.2569	**377**, 337, **295**	3, 4
Persin **(7)**	381.3005	403.2971, 321.2874, 303.2765, 285.2650	**403**, 363, **321**, **303**	2

1. M. d. R. Ramos-Jerz, Dr. rer. nat., Universität Carolo-Wilhelmina, 2007.
2. F. Domergue, G. L. Helms, D. Prusky and J. Browse, *Phytochemistry*, 2000, **54**, 6.
3. D. Rodríguez Sanchez, A. Pacheco, M. I. García-Cruz, J. A. Gutiérrez-Uribe, J. A. Benavides-Lozano and C. Hernández-Brenes, *Journal of Agricultural and Food Chemistry*, 2013, **61**, 7403-7411.
4. D. Rodríguez Sanchez, C. Silva-Platas, R. P. Rojo, N. García, L. Cisneros-Zevallos, G. García-Rivas and C. Hernández-Brenes, *Journal of Chromatography B*, 2013, **942-3**, 37-45.

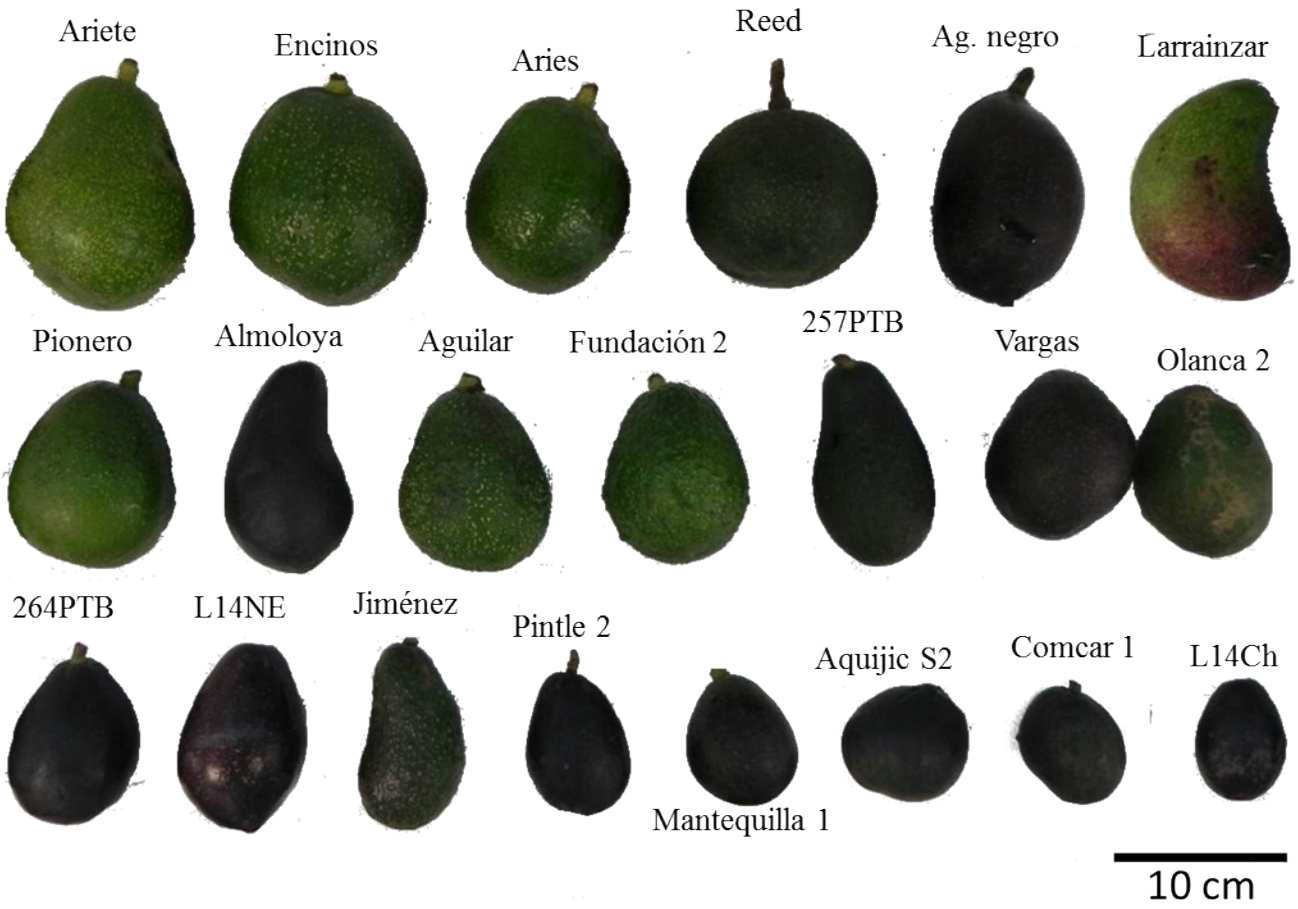
Supplementary Table 3. Coefficients of the reduced model. Multipliers generated from the reduced matrix, when running the pLDA, in order to classify cultivars with unknown origin. Each acetogenin concentration is multiplied by their respective coefficient, and then summed, resulting in three scores: one for each genotype. The individual is then assigned the genotype which has the highest score.

Genotype	Intercept	Pulp					Peel	
		Persenone C	UPA	Persenone B	AcO-avocadene	Persenone A	Persenone C	Persenone B
Guatemalan	-2.52575	4.391288	7.092056	0.720033	8.817316	-1.32418	-31.668	17.9472
Hybrid	-12.299	-77.5148	125.6501	31.67652	-3.21064	2.358588	5.324453	9.490533
Mexican	-4.48492	-3.32605	29.24575	13.56297	-15.4739	1.929643	26.19647	-0.80269

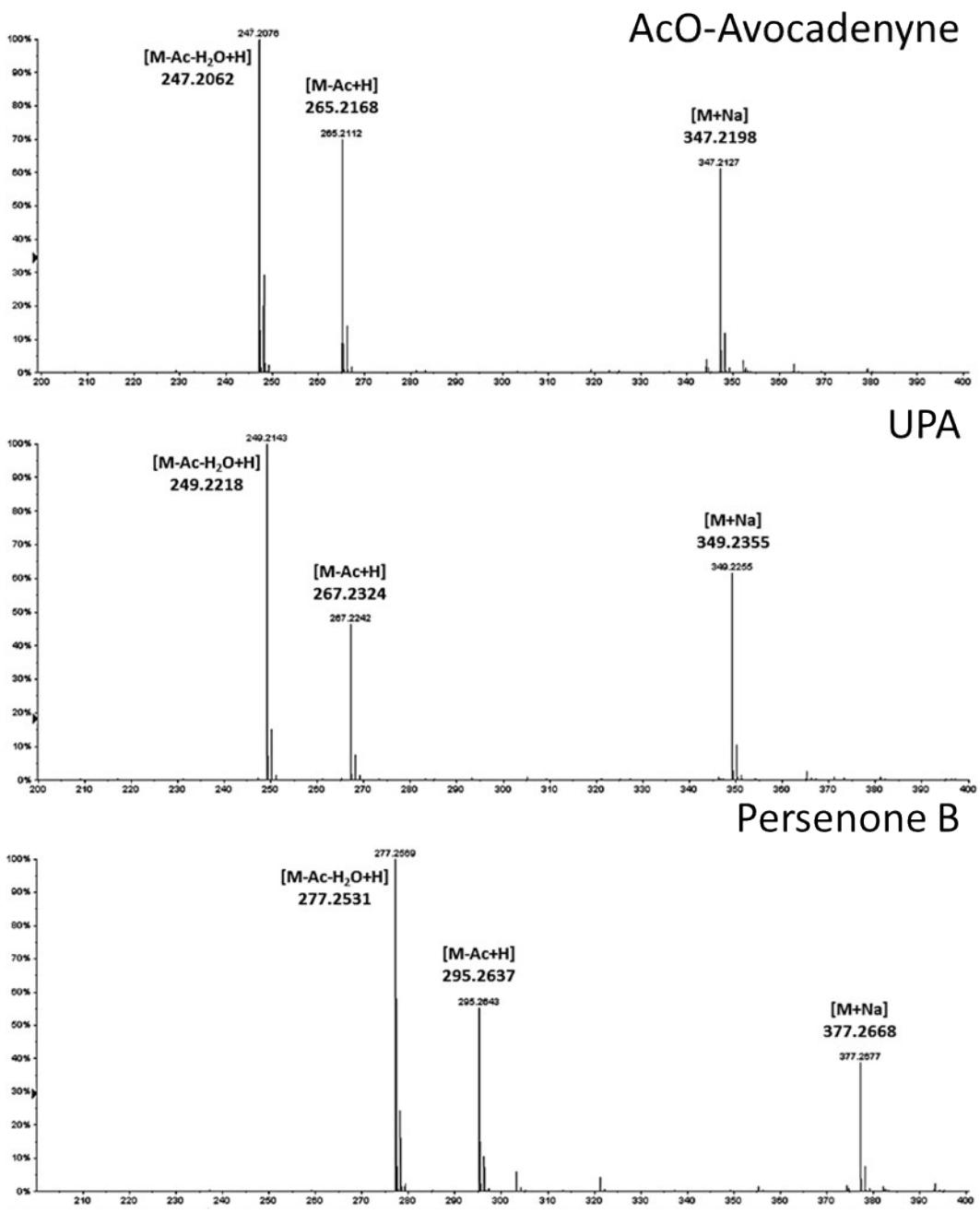
Supplementary Table 4. Confusion matrix of the reduced model. The confusion matrix was generated using a bootstrapped (5,000 iterations) 3-fold cross validation, and normalizing to the original genotypes (i.e. as if there were 100 Guatemalan individuals to be predicted.)

		Predicted		
		Guatemalan	Hybrid	Mexican
Original	Guatemalan	91	5	4
	Hybrid	9	80	11
	Mexican	14	7	79

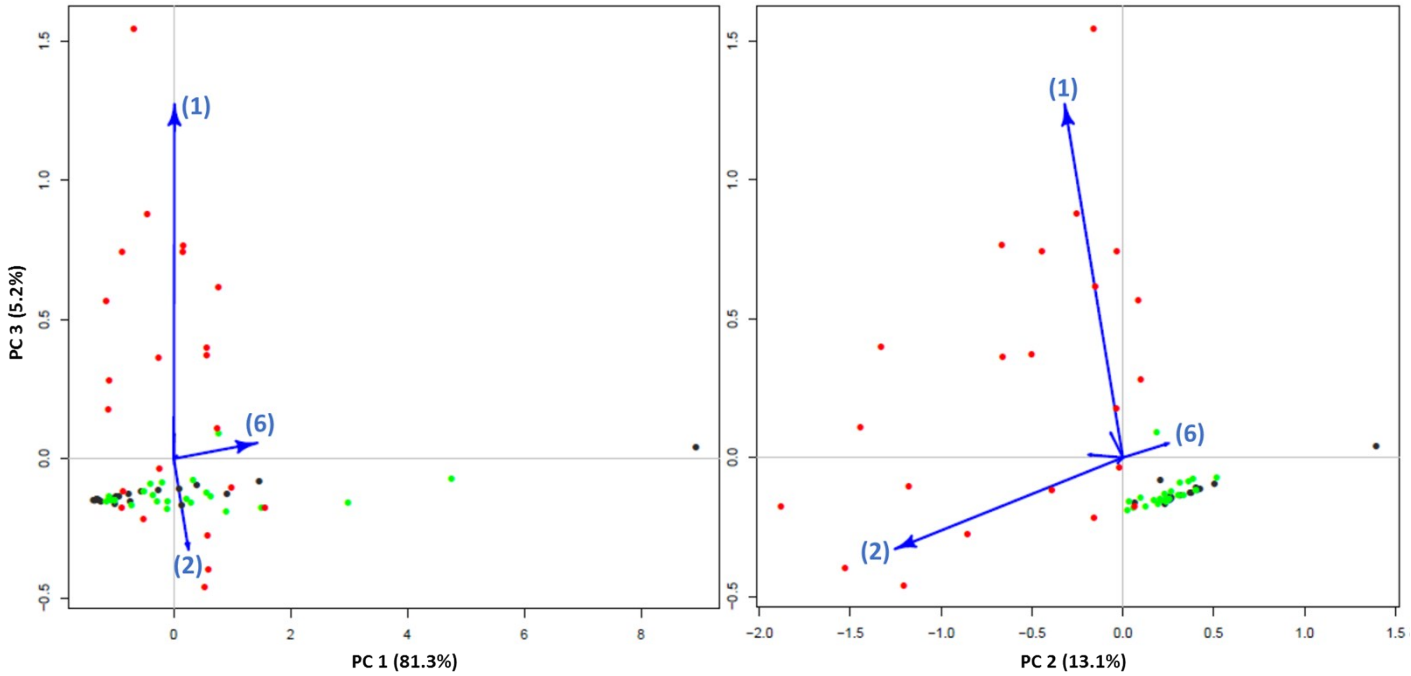
Suplementary Figures



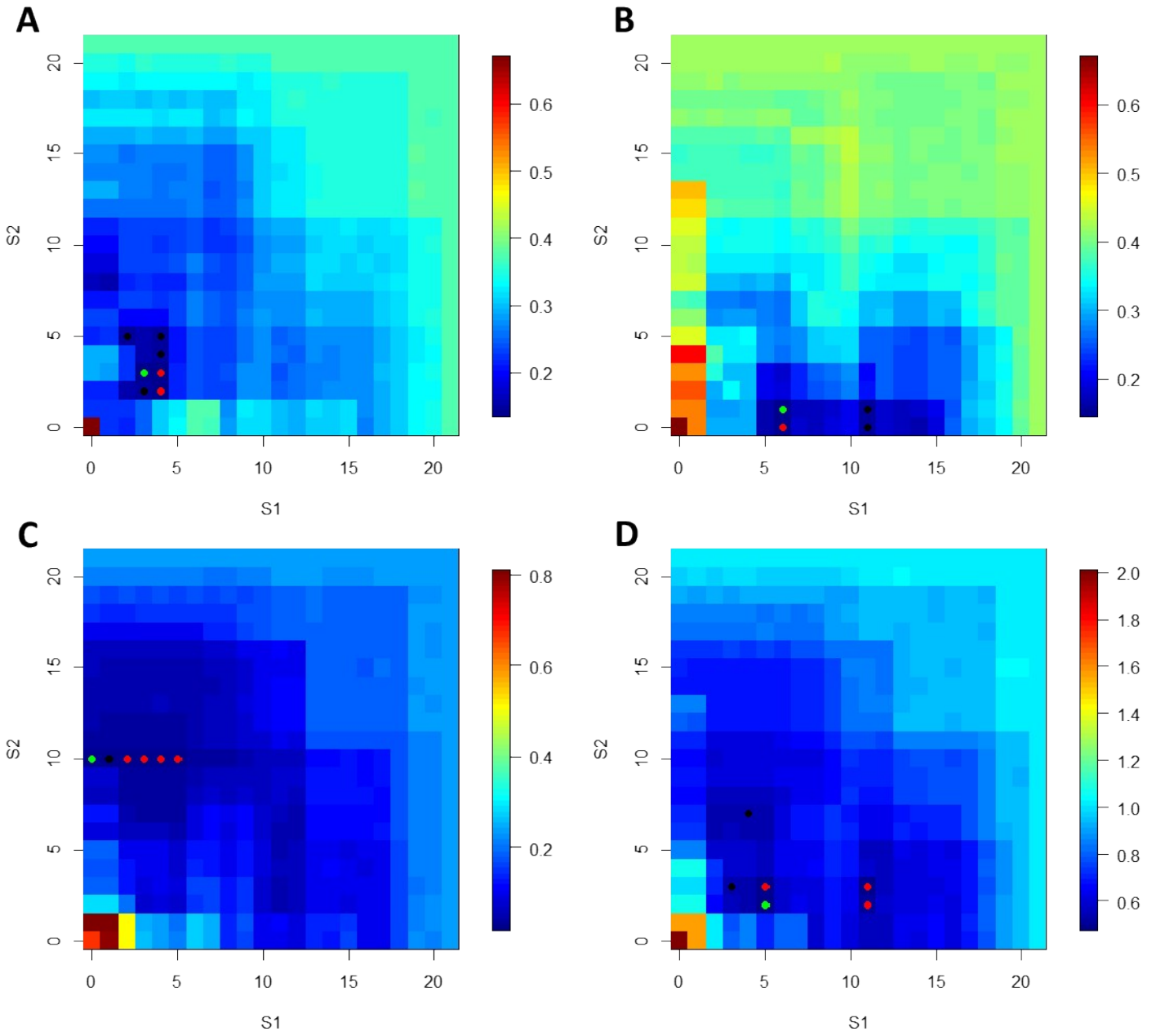
Supplementary Figure 1. Phenotypic diversity of the CICTAMEX collection. Representative images of the avocado fruits used in this study, ordered by size. Photos were taken at the time of harvesting, when the fruit was at mature-green stage.



Supplementary Figure 2. Mass fragmentation patterns of putative acetogenins and a Persenone B standard: MS spectra of a purified standard of Persenone B (bottom,) peak **3** (Unidentified Putative Acetogenin, middle) in a sample, and peak **1** (AcO-avocadenyne, top) as part of a sample. In brackets, the putative adducts and fragments (with indicated neutral losses), with their theoretical masses below. The similarities between fragmentation patterns indicate a strong similarity in structure.



Supplementary Figure 3. Remaining PCA component: Average of PCA scores per cultivar (depicted as points), plotted by tissue (seed: red; pulp: green; and peel: gray) and loadings (blue arrows) projected on the first and third (left) and second and third (right) components (n=3). Size of the arrow is proportional to the magnitude of the loadings; vectors are scaled and therefore, the magnitude does not correspond to the axes; numbers indicate the corresponding acetogenin as depicted in Figure 1, only the main three loadings are shown.



Supplementary Figure 4. Model reduction: Percentages of wrong assignations of Mexican, Guatemalan and Hybrid cultivars. Each panel is the result of 5,000 iterations on the 3-fold cross-validation of the predictive Linear Discriminant Analysis (pLDA) as a function of the number of factors used, ranked by the canonical analysis on the descriprive LDA for the two main dimensions of **(A)**Mexican, **(B)**Hybrid and **(C)**Guatemalan genotypes, and **(D)** the sum of all three.