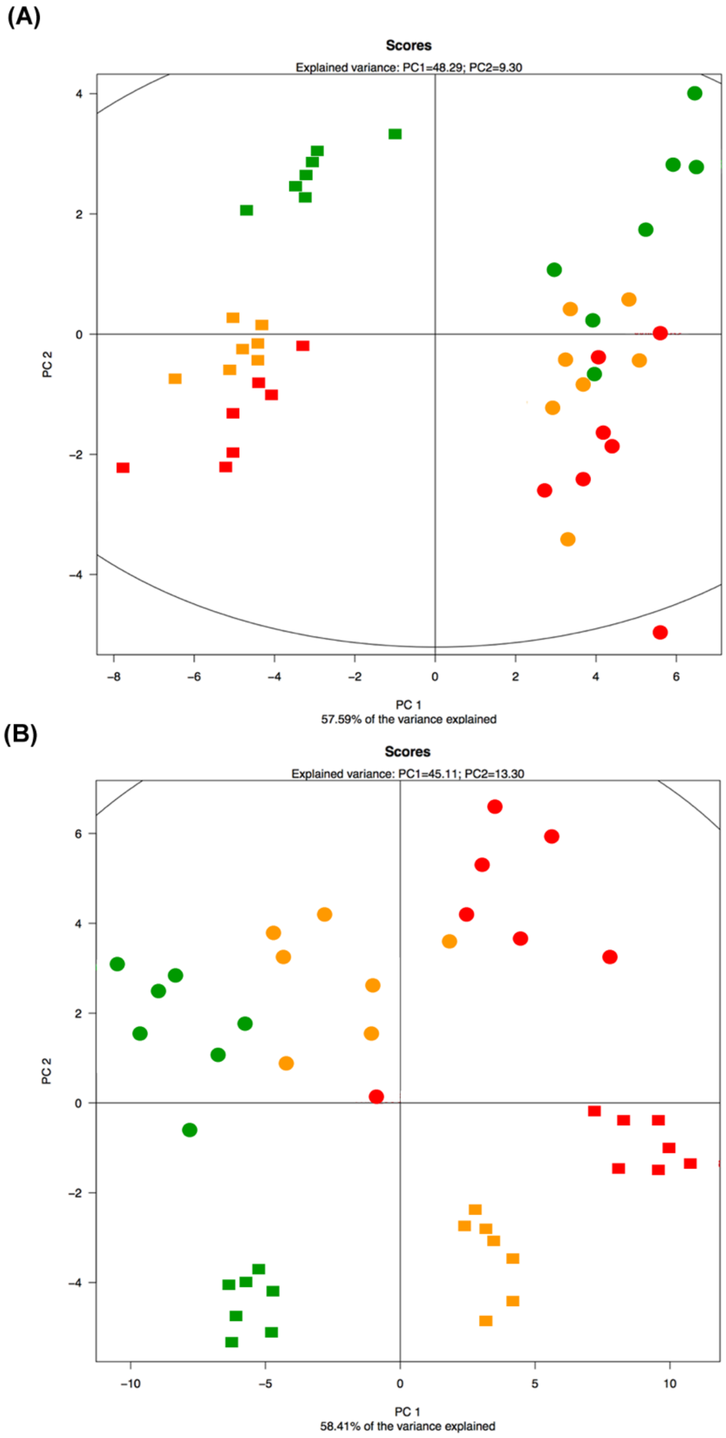
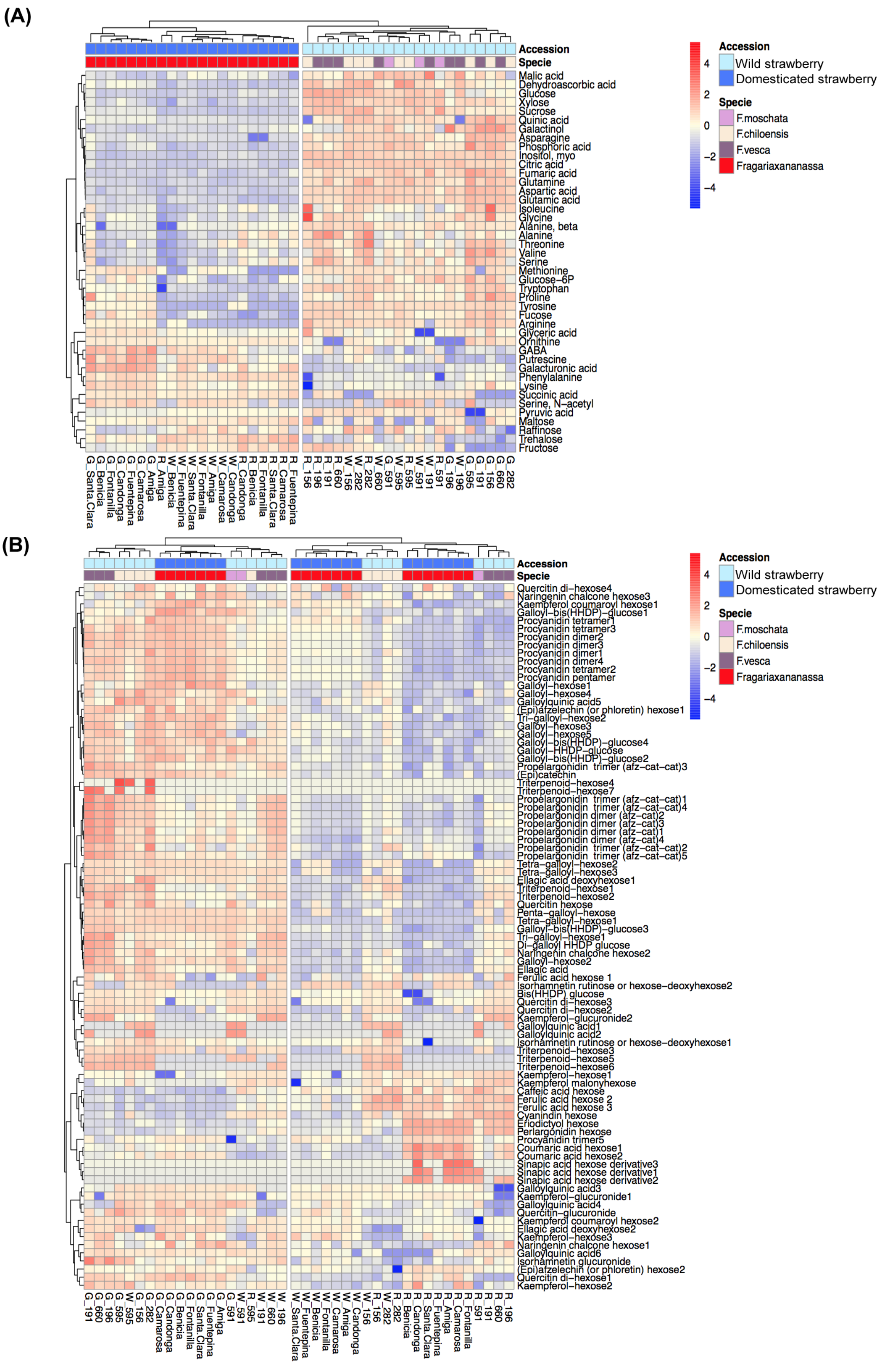
Supplementary information

Genetic diversity of strawberry germplasm using metabolomic biomarkers

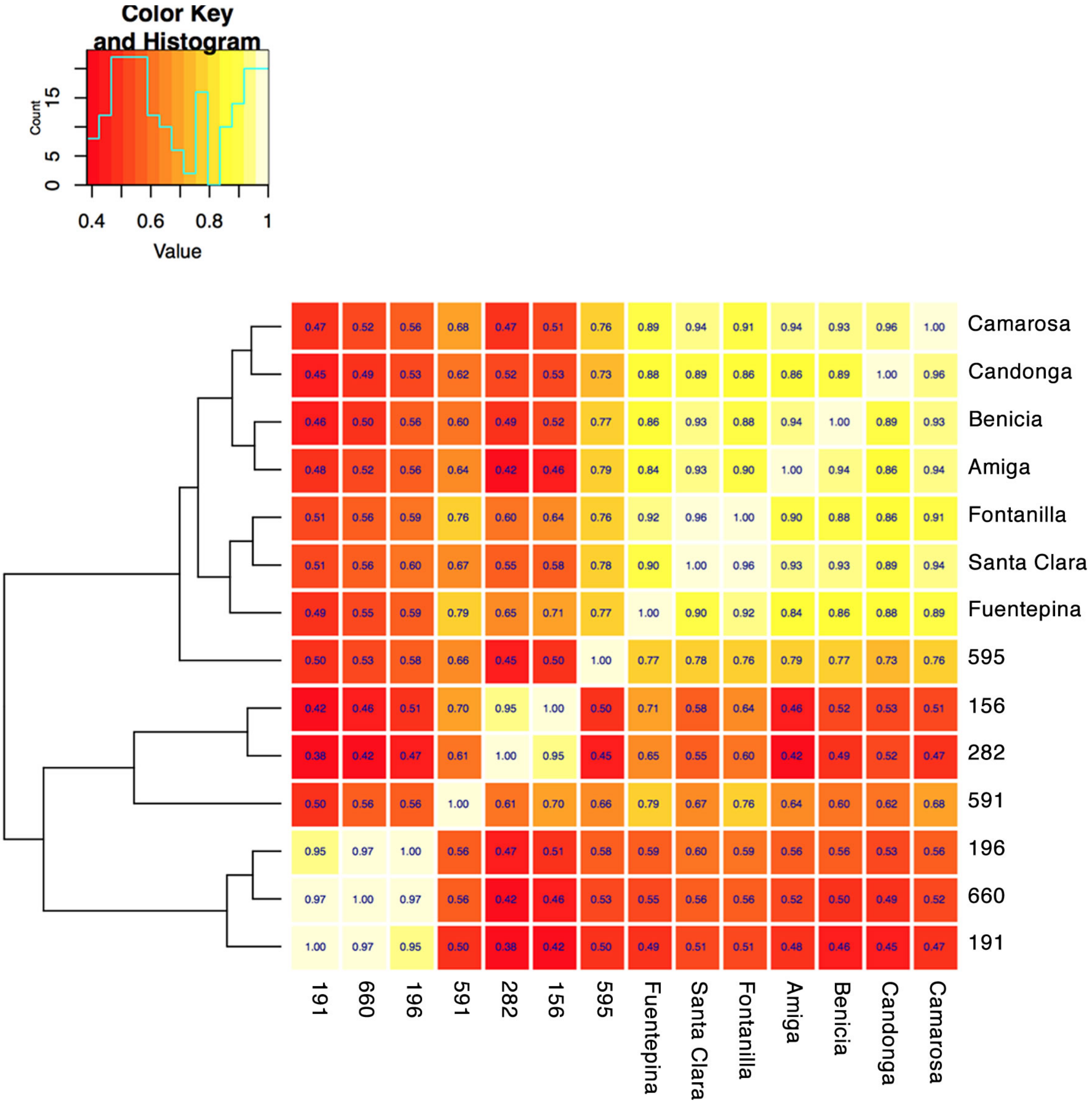
José G Vallarino, Francisco de Abreu e Lima, Carmen Soria, Hao Tong, Delphine M. Pott, Lothar Willmitzer, Alisdair R Fernie, Zoran Nikoloski, Sonia Osorio.



Supplementary Figure S1. Principal component analysis (PCA) of metabolites data of seven domesticated and seven wild strawberry accessions across different developmental and ripening fruit stages. PCAs of (A) primary metabolites and (B) polar secondary metabolites. Score plots of the PCA, domesticated (squares) and wild (circles) strawberry accessions. Green symbols, green stage; orange symbols, white stage; red symbols, red stage.

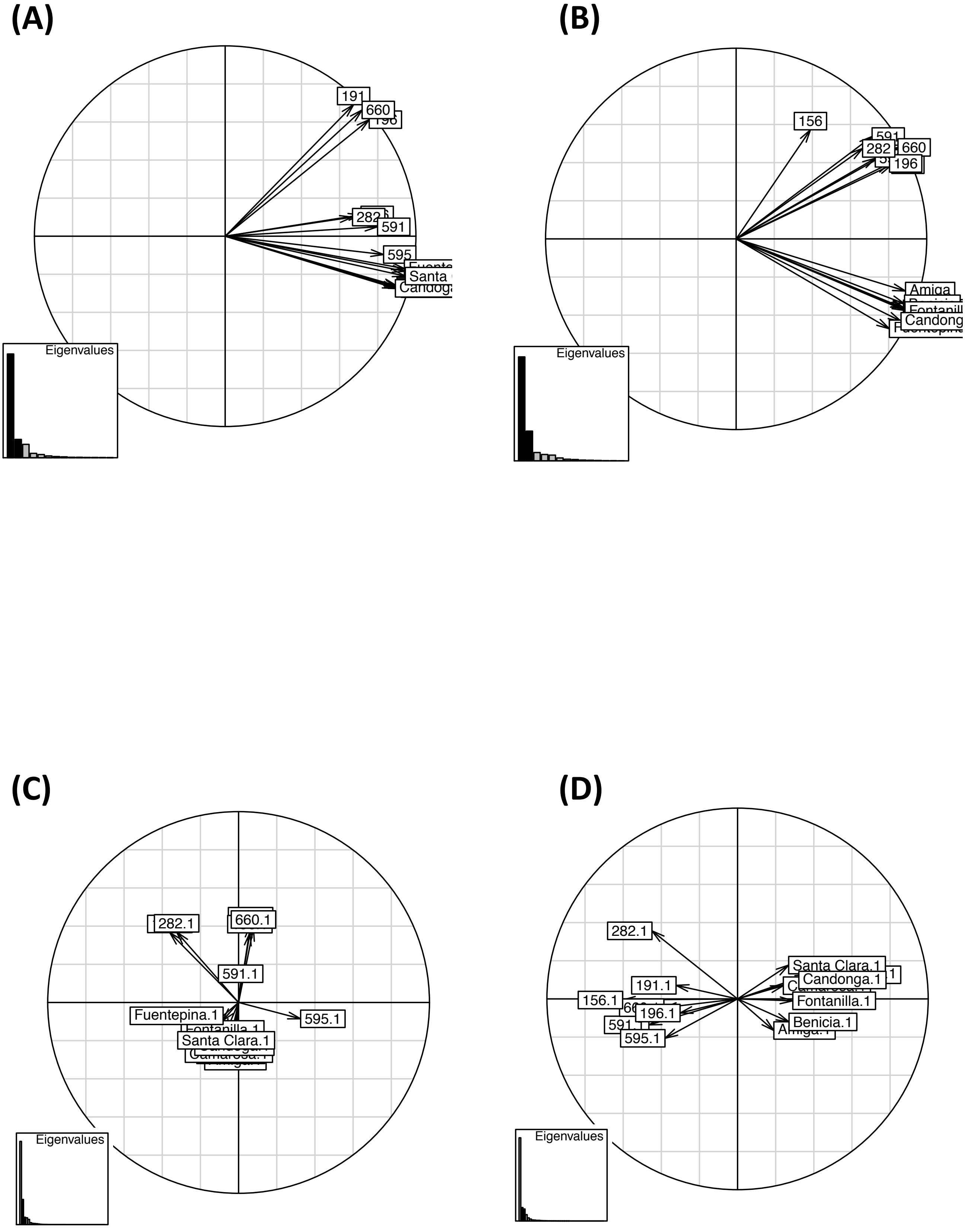


Supplementary Figure S2. (A) Heatmap of 43 primary metabolites, and (B) 87 polar secondary metabolites in seven domesticated and seven wild strawberry accessions across different developmental and ripening fruit stages. A color-coded matrix represents the mean values of the metabolite intensity in three biological replicates of strawberry accessions, which has been log2 transformed and mean-centered.

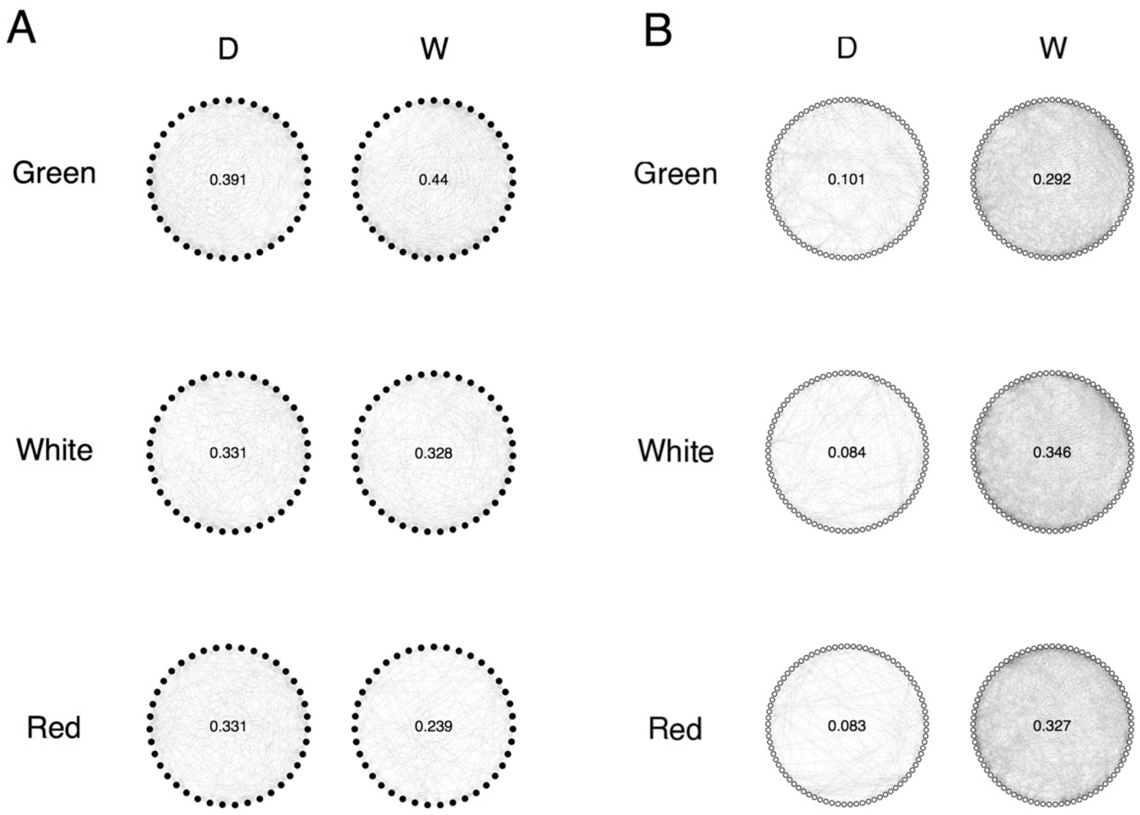


	191	660	196	591	282	156	595	Fuentepina	Santa Clara	Fontanilla	Amiga	Benicia	Candonga	Camarosa
Camarosa	0.47	0.52	0.56	0.68	0.47	0.51	0.76	0.89	0.94	0.91	0.94	0.93	0.96	1.00
Candonga	0.45	0.49	0.53	0.62	0.52	0.53	0.73	0.88	0.89	0.86	0.86	0.89	1.00	0.96
Benicia	0.46	0.50	0.56	0.60	0.49	0.52	0.77	0.86	0.93	0.88	0.94	1.00	0.89	0.93
Amiga	0.48	0.52	0.56	0.64	0.42	0.46	0.79	0.84	0.93	0.90	1.00	0.94	0.86	0.94
Fontanilla	0.51	0.56	0.59	0.76	0.60	0.64	0.76	0.92	0.96	1.00	0.90	0.88	0.86	0.91
Santa Clara	0.51	0.56	0.60	0.67	0.55	0.58	0.78	0.90	1.00	0.96	0.93	0.93	0.89	0.94
Fuentepina	0.49	0.55	0.59	0.79	0.65	0.71	0.77	1.00	0.90	0.92	0.84	0.86	0.88	0.89
595	0.50	0.53	0.58	0.66	0.45	0.50	1.00	0.77	0.78	0.76	0.79	0.77	0.73	0.76
156	0.42	0.46	0.51	0.70	0.95	1.00	0.50	0.71	0.58	0.64	0.46	0.52	0.53	0.51
282	0.38	0.42	0.47	0.61	1.00	0.95	0.45	0.65	0.55	0.60	0.42	0.49	0.52	0.47
591	0.50	0.56	0.56	1.00	0.61	0.70	0.66	0.79	0.67	0.76	0.64	0.60	0.62	0.68
196	0.95	0.97	1.00	0.56	0.47	0.51	0.58	0.59	0.60	0.59	0.56	0.56	0.53	0.56
660	0.97	1.00	0.97	0.56	0.42	0.46	0.53	0.55	0.56	0.56	0.52	0.50	0.49	0.52
191	1.00	0.97	0.95	0.50	0.38	0.42	0.50	0.49	0.51	0.51	0.48	0.46	0.45	0.47

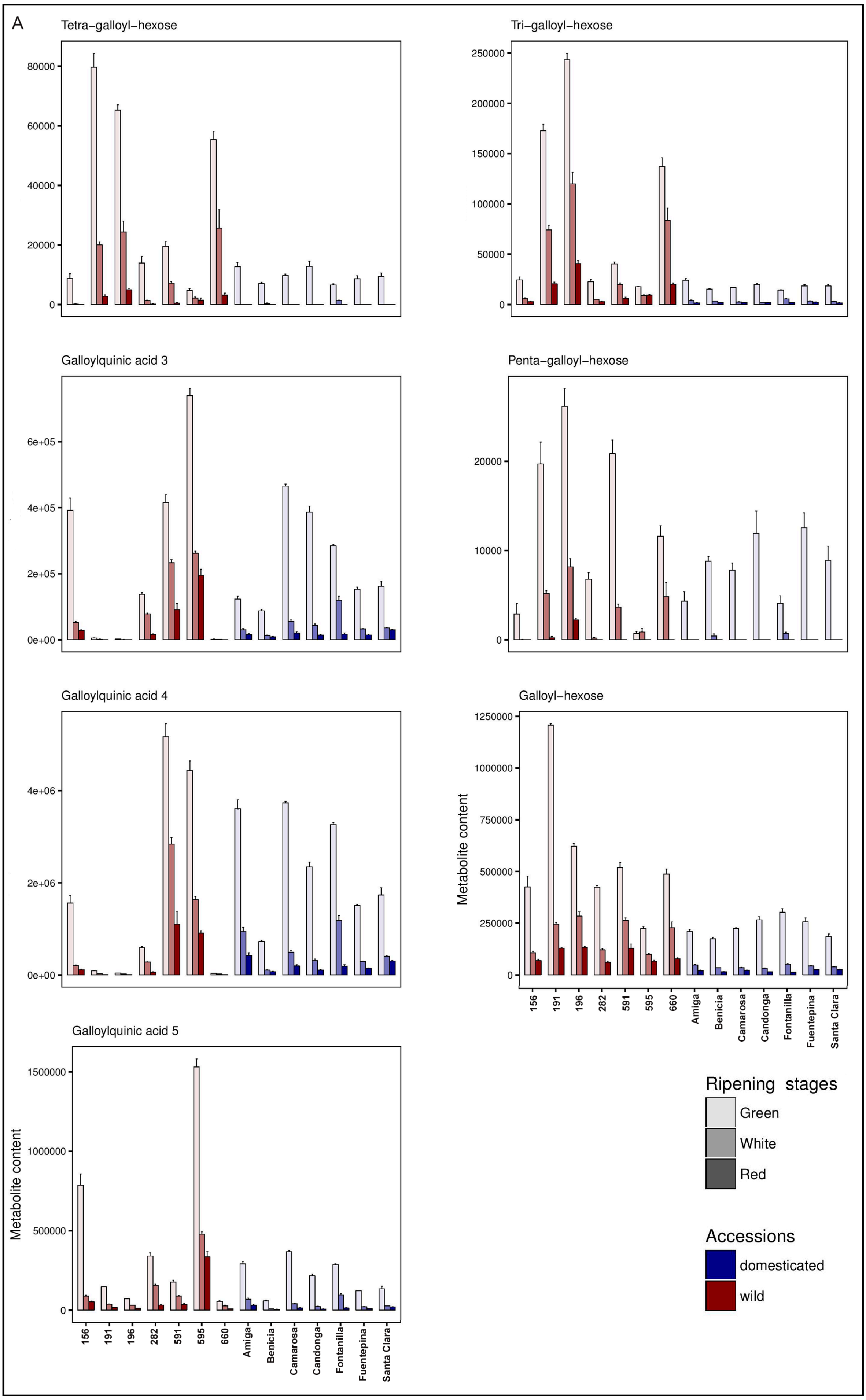
Supplementary Figure S3. Visualization of similarity of secondary metabolite based on hierarchical clustering and heatmap. Similarity between the tables from the different accessions is based on the Rv coefficient. A value of 1 for the Rv coefficient corresponds to equivalence (white), while a value of 0 indicates complete dissimilarity (dark red).



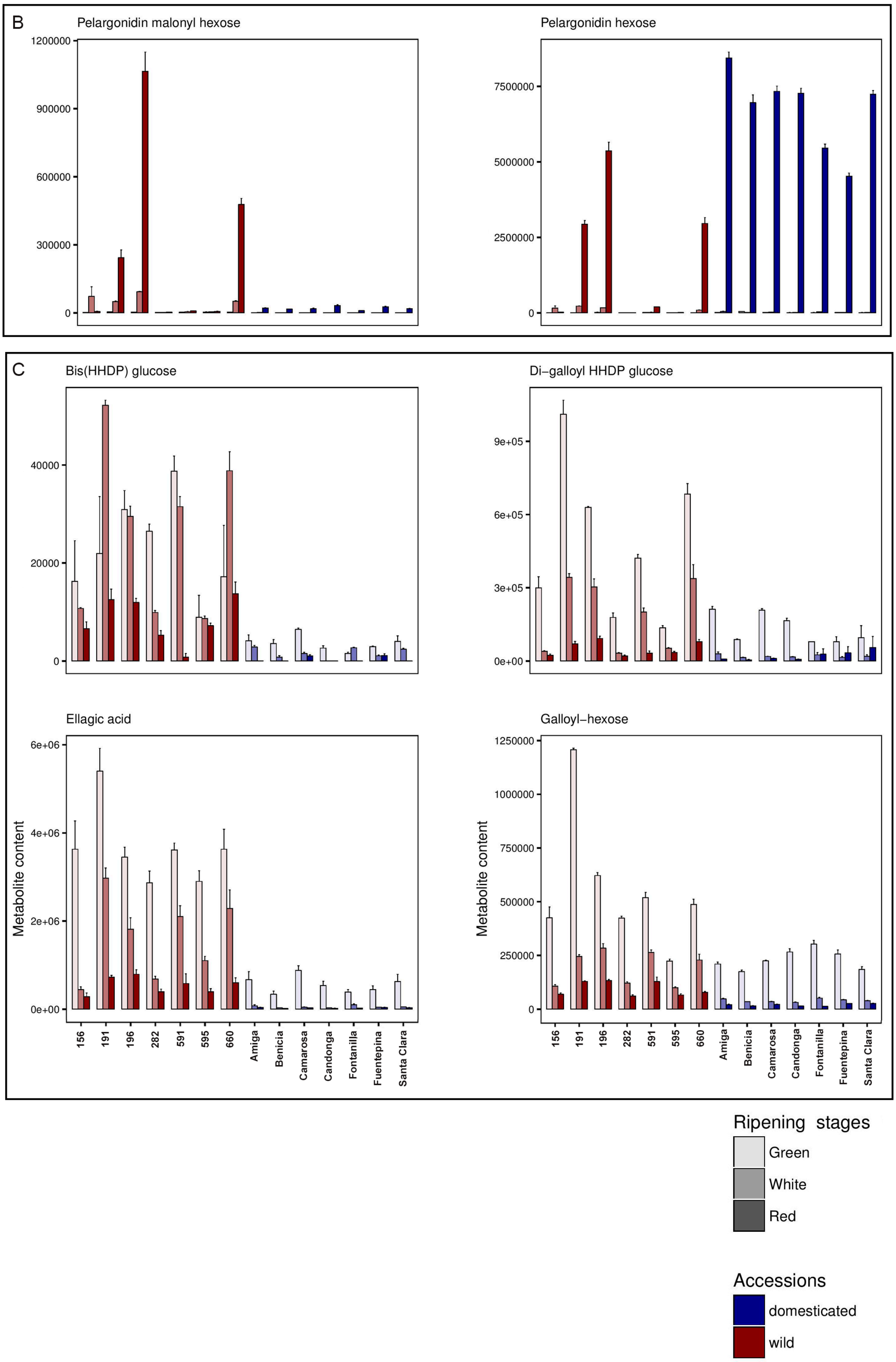
Supplementary Figure S4. Contribution of tables of accessions to PC1 and PC2 (A and B) and PC2 and PC3 (C and D) for secondary metabolites (A and C) and primary metabolites (B and D).



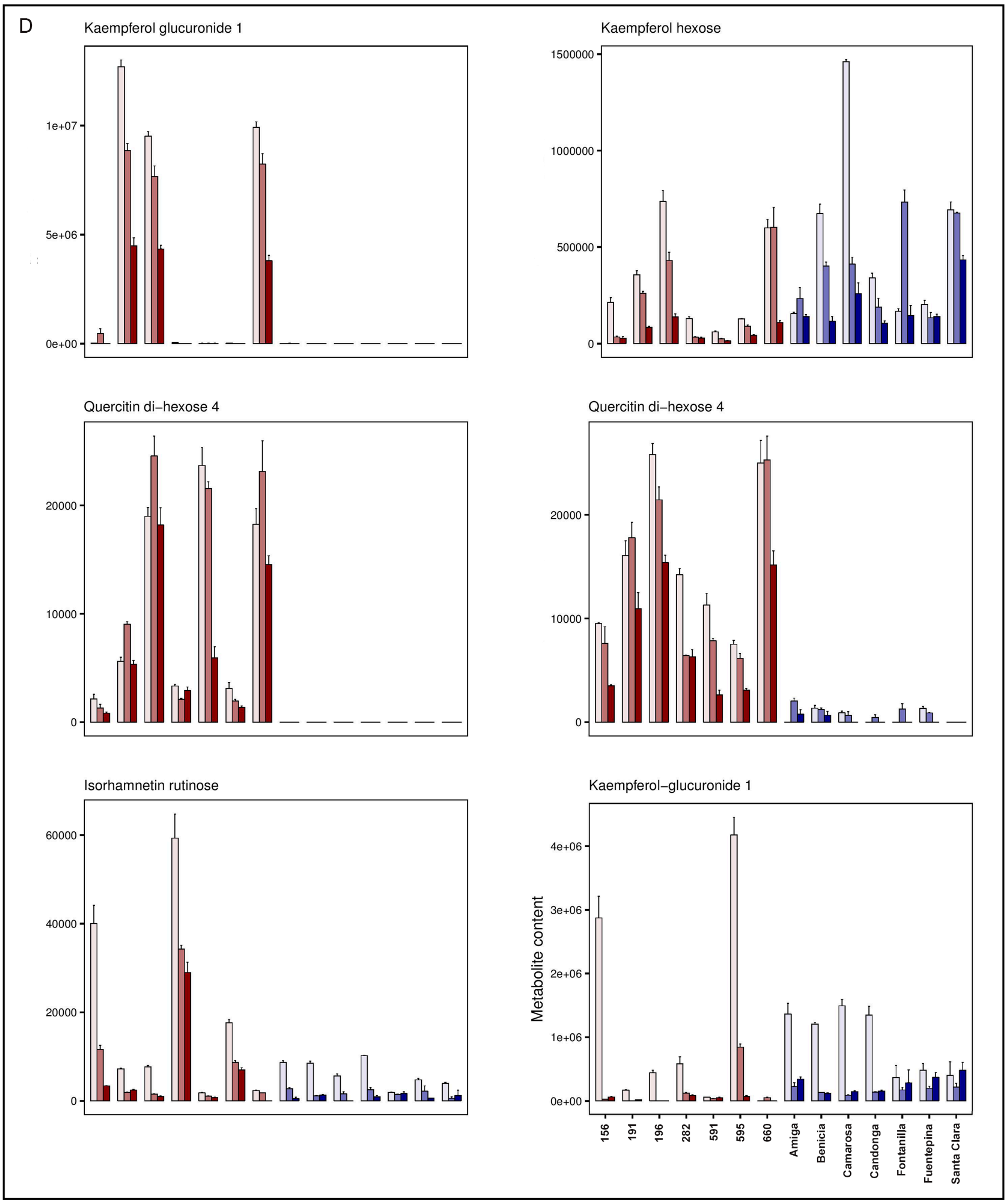
Supplementary Figure S5. Comparative network analysis using domesticated (D) and wild (W) strawberry varieties separately for each metabolomics platform. For every maturation stage (green, white and red), the correlation-based metabolic network from D was compared to that from W. The values shown inside the networks are the corresponding densities. (A) The comparative analysis using only primary metabolites (GC-MS) shows slight, unsystematic differences in density in the different stages. (B) The comparative analysis using only secondary metabolites (LC-MS) shows systematic differences in density that exceed more than 2-fold of D in all stages. Filled circles represent primary metabolites (GC-MS) and open circles represent secondary metabolites (LC-MS).

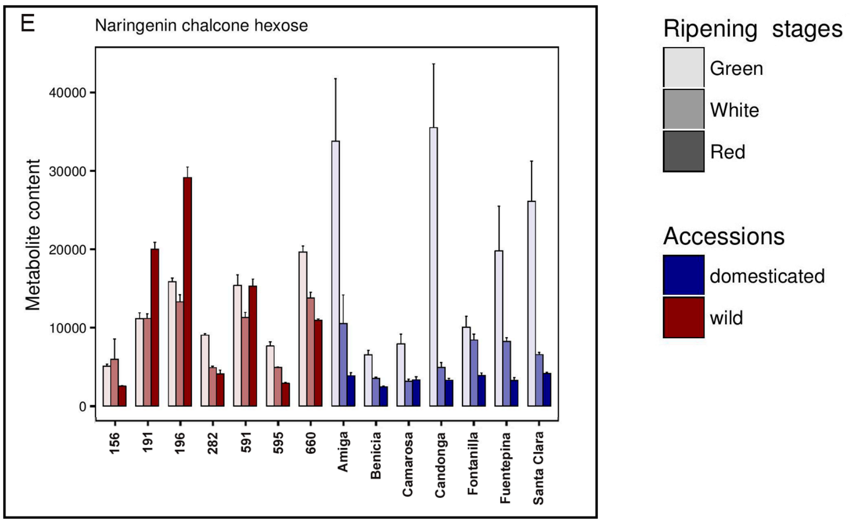


Supplementary Figure S6. Relative abundance in green, white and red fruits of the metabolites which exhibited significant population effect. (A) galloyl glucoses, (B) anthocyanins, (C) ellagitannins and ellagic acid precursor/derivative, (D) flavonols, flavanols, (E) flavanones, (F) propelargonidins, (G) terpenoid derivatives and H) primary metabolites.

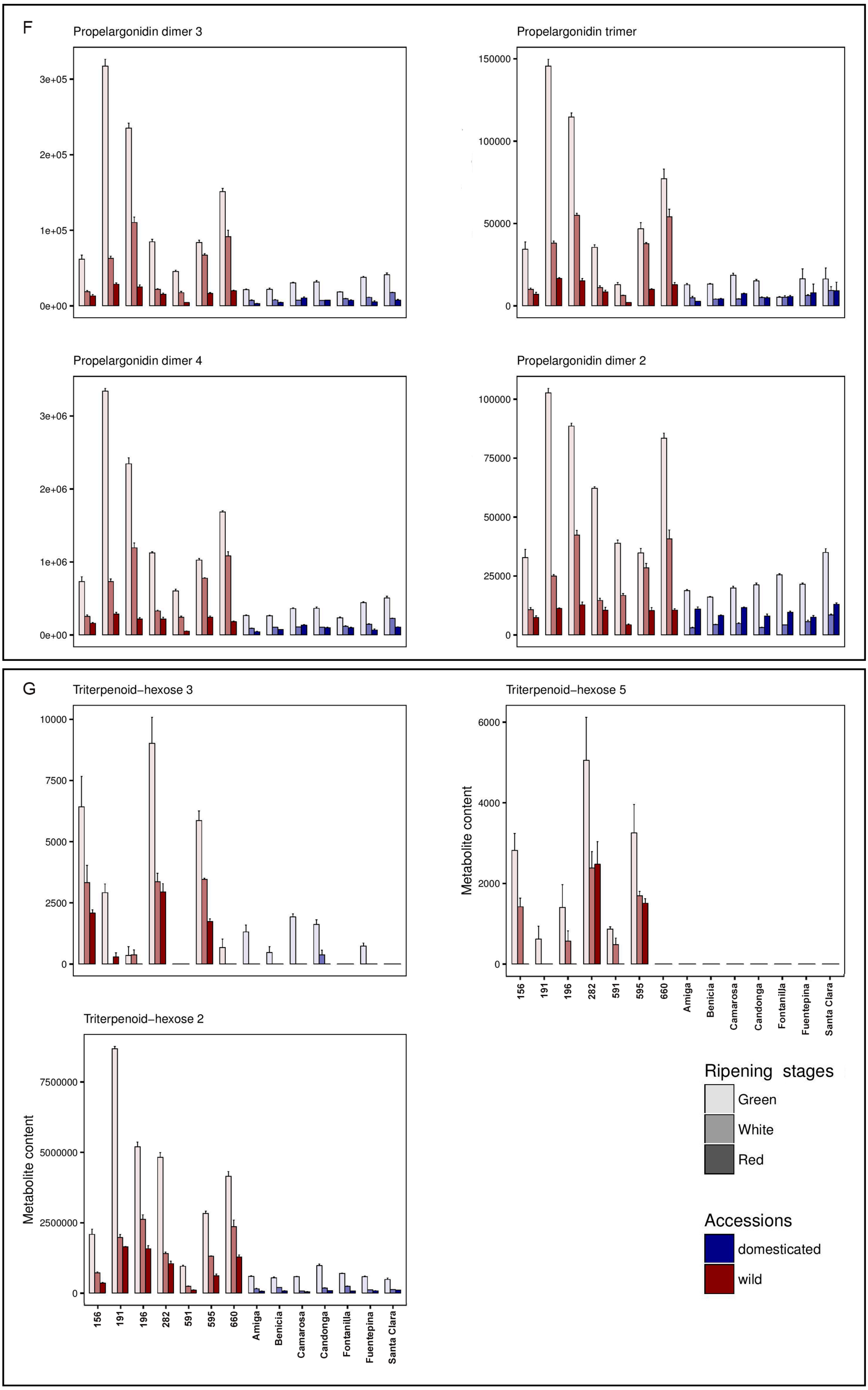


Supplementary Figure S6

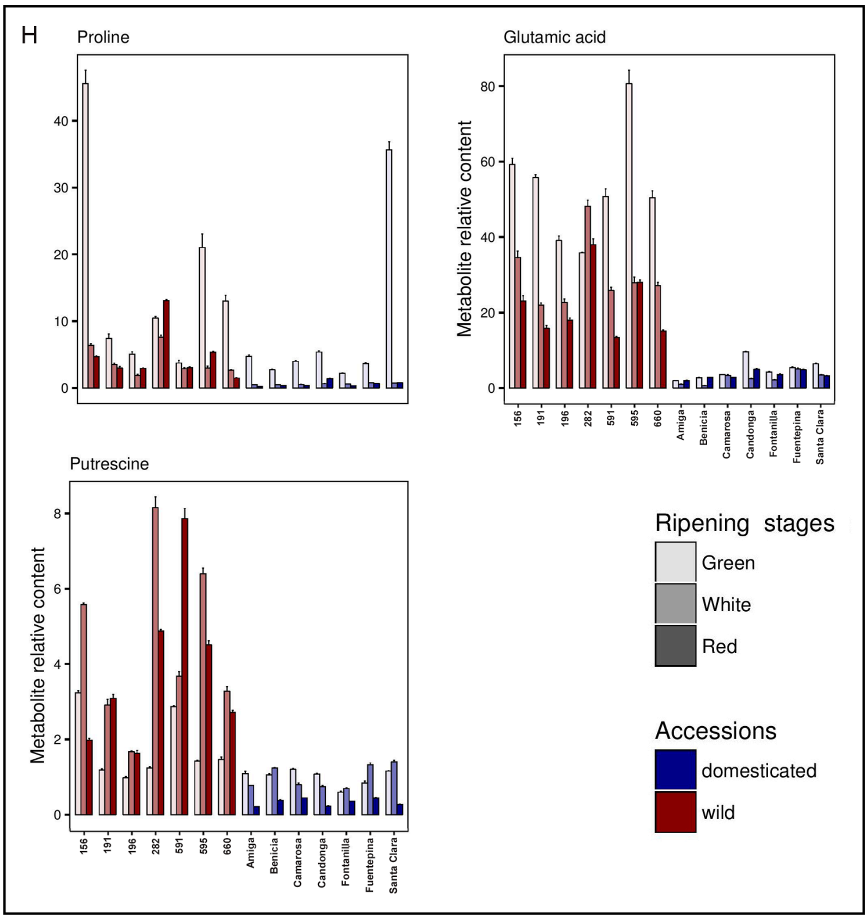




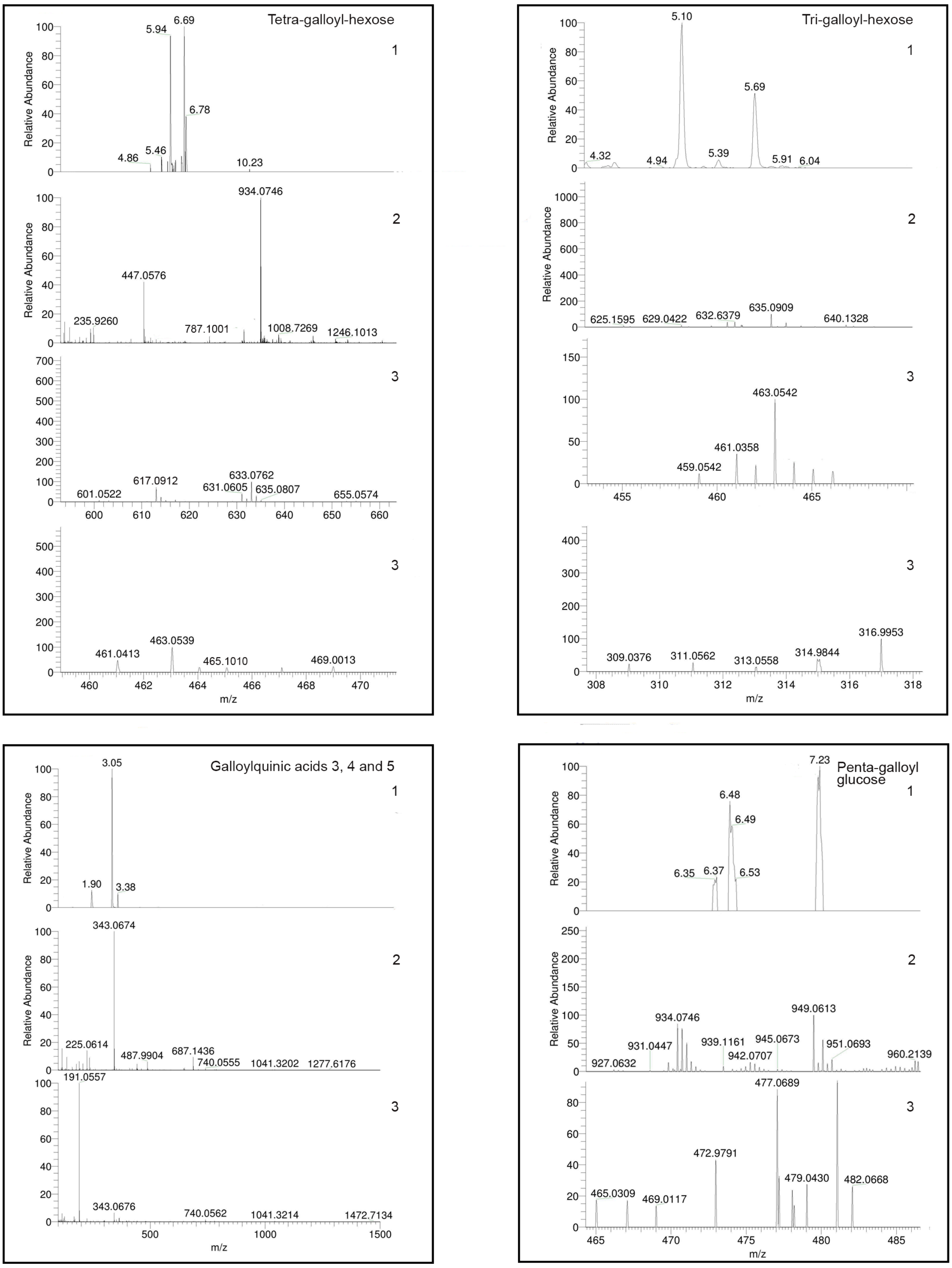
Supplementary Figure S6



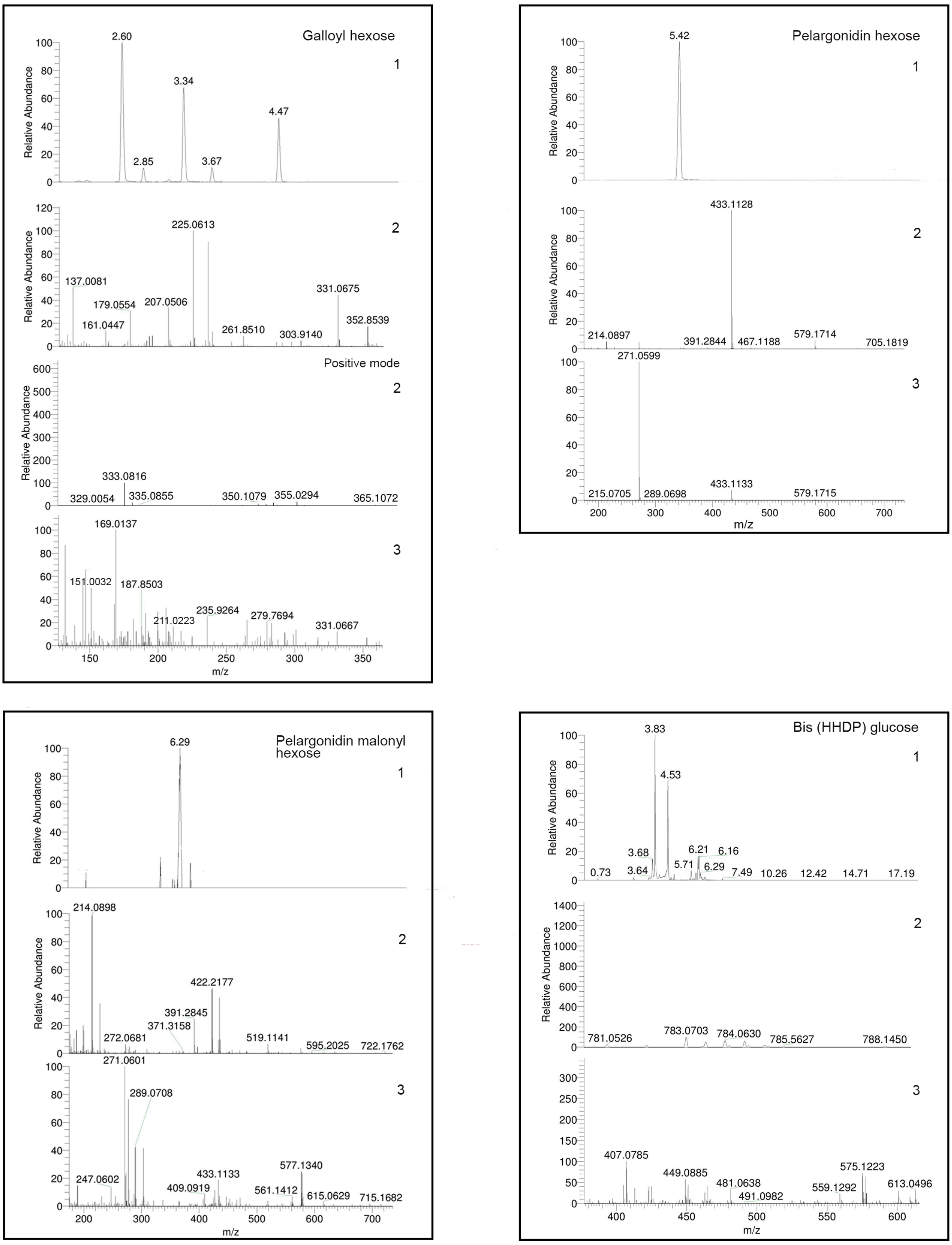
Supplementary Figure S6



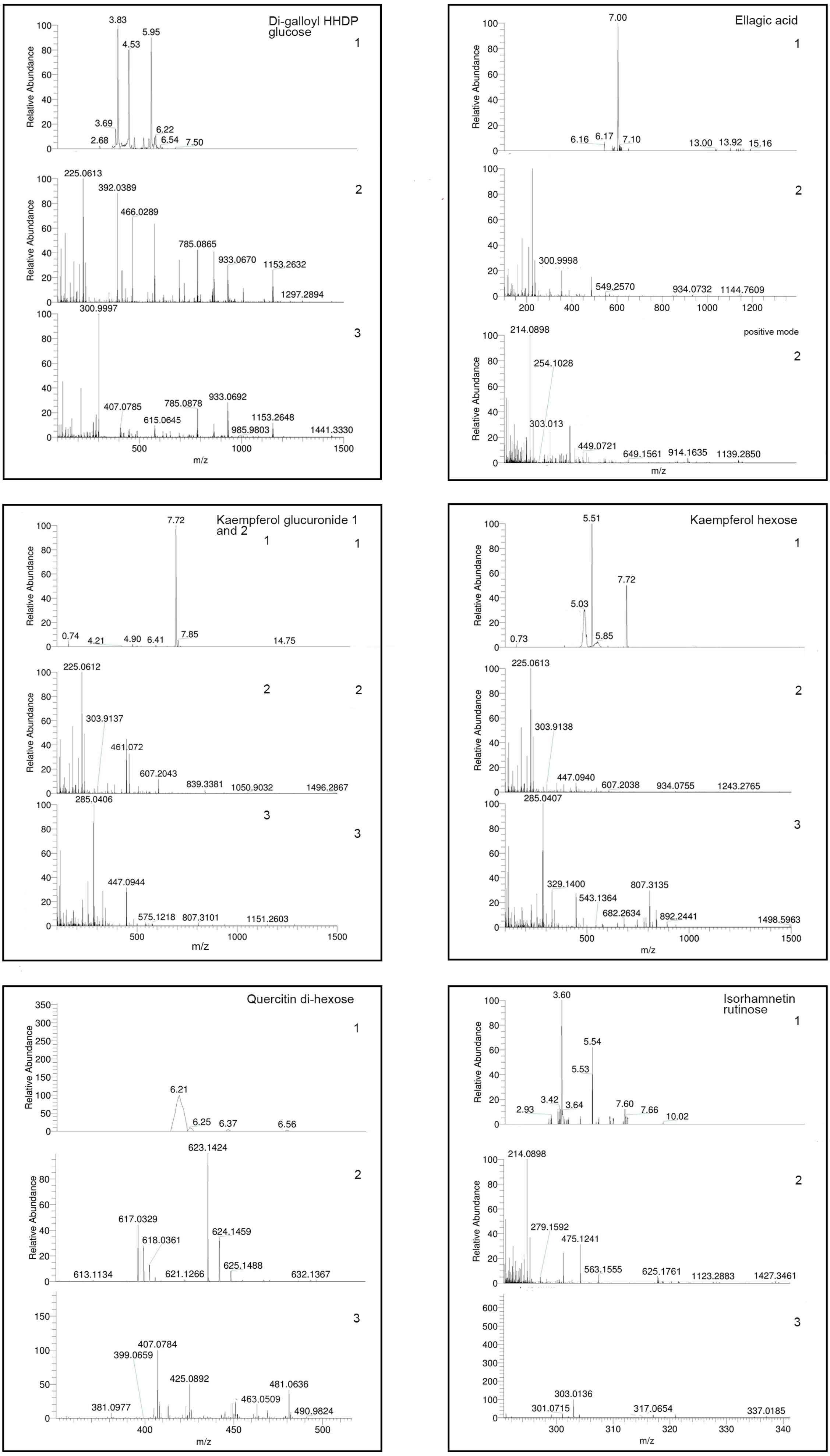
Supplementary Figure S6



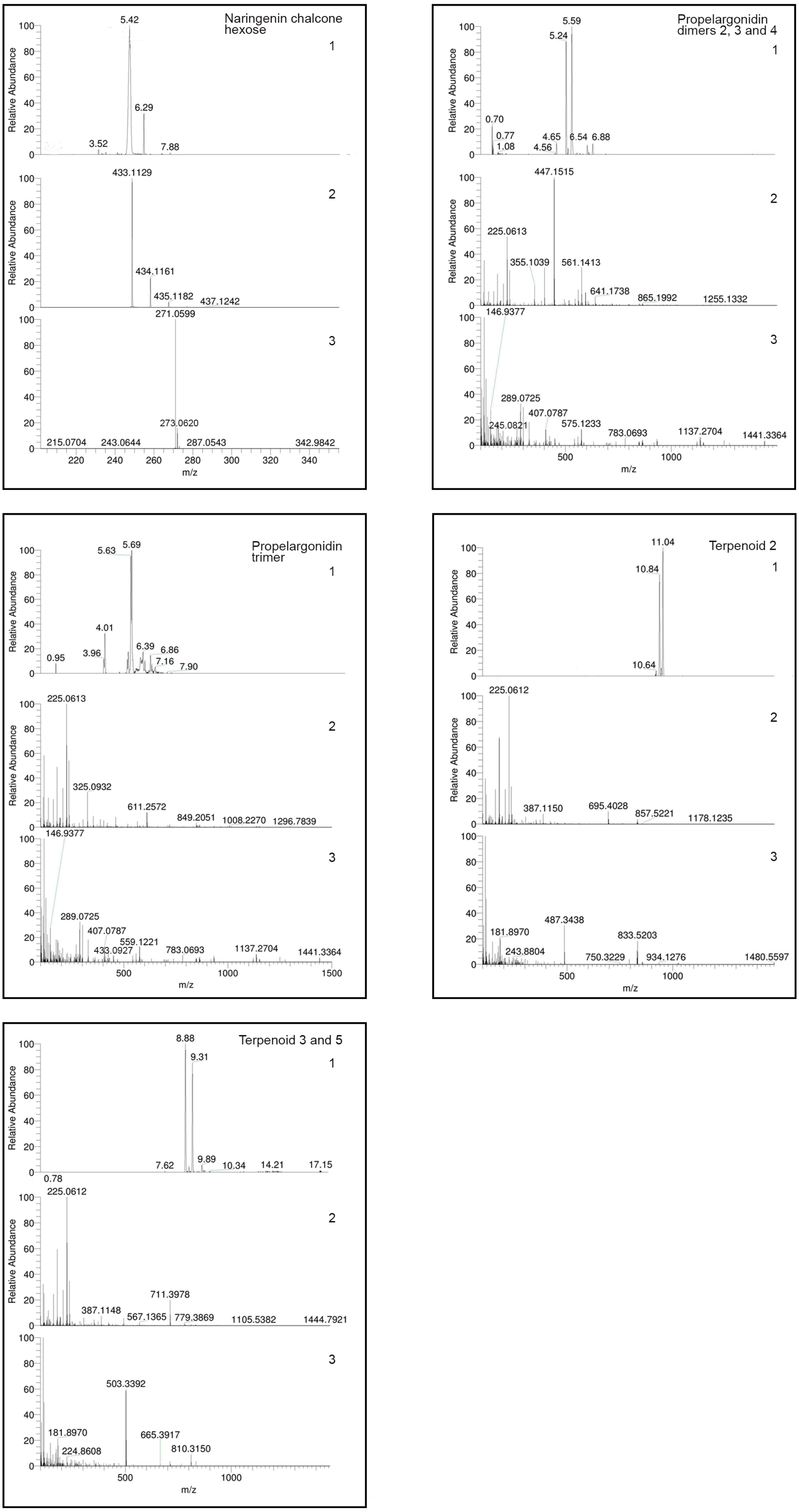
Supplementary Figure S7. Chromatograms (1), MS (2) and MS/MS (3) spectra of the metabolites which exhibited significant population effect. Spectra are shown for negative ionization mode, except where positive ionization mode is indicated. Spectra of different isomers for the same compound are only represented once.



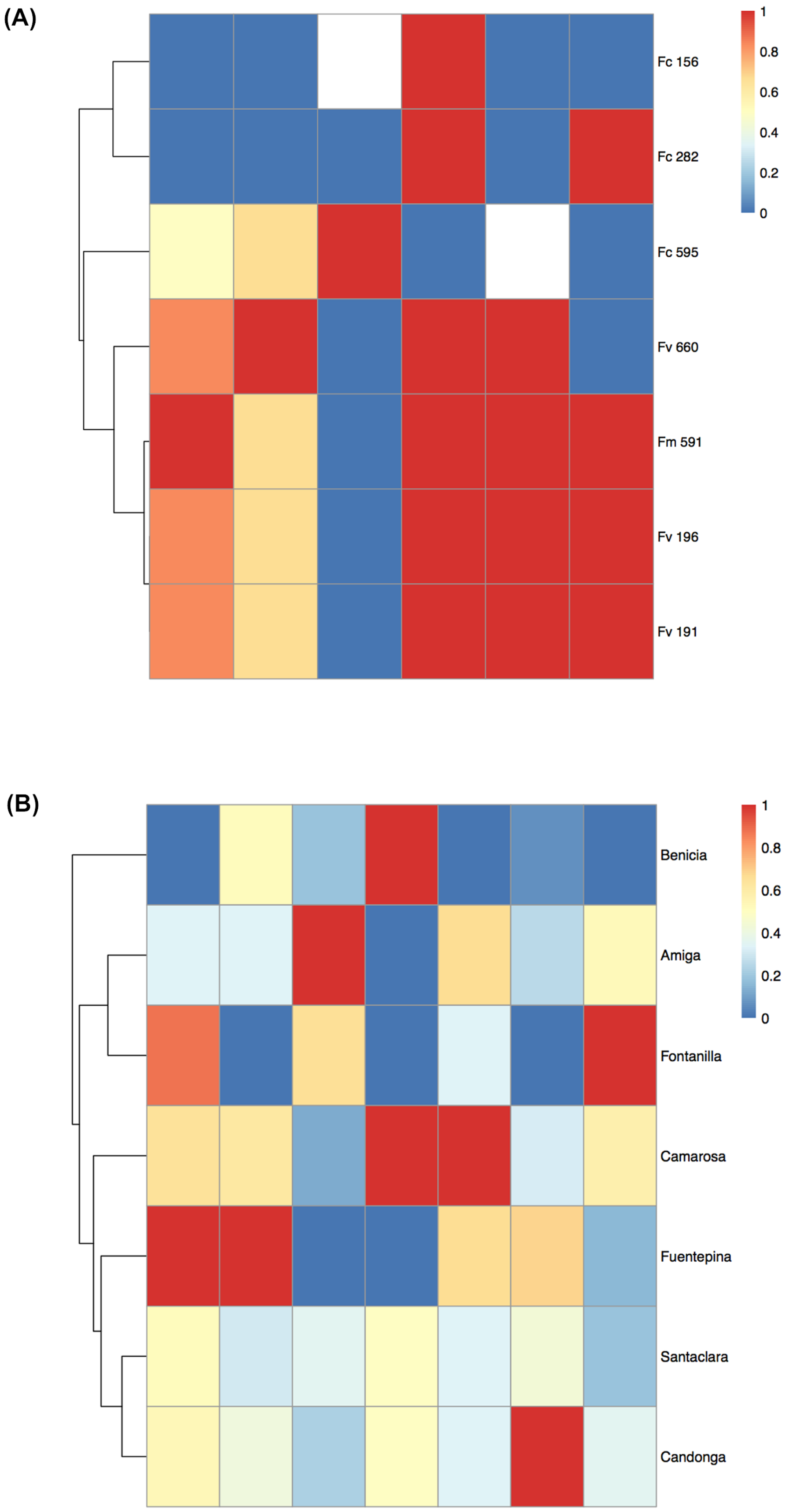
Supplementary Figure S7.



Supplementary Figure S7.



Supplementary Figure S7.



Supplementary Figure S8. Clustering of accessions based on complex traits**.** (A) wild accessions, based on growth habitat, plant vigor, plant density, number of flowers, color, and shape. (B) domesticated accessions, based on Brix, yield, plant diameter, weight, firmness, color, and shape. Heatmaps are based on correlation matrix after data normalization, and include a dendrogram from hierarchical clustering.