Supplementary Material

Supplementary Figures 1 to 5

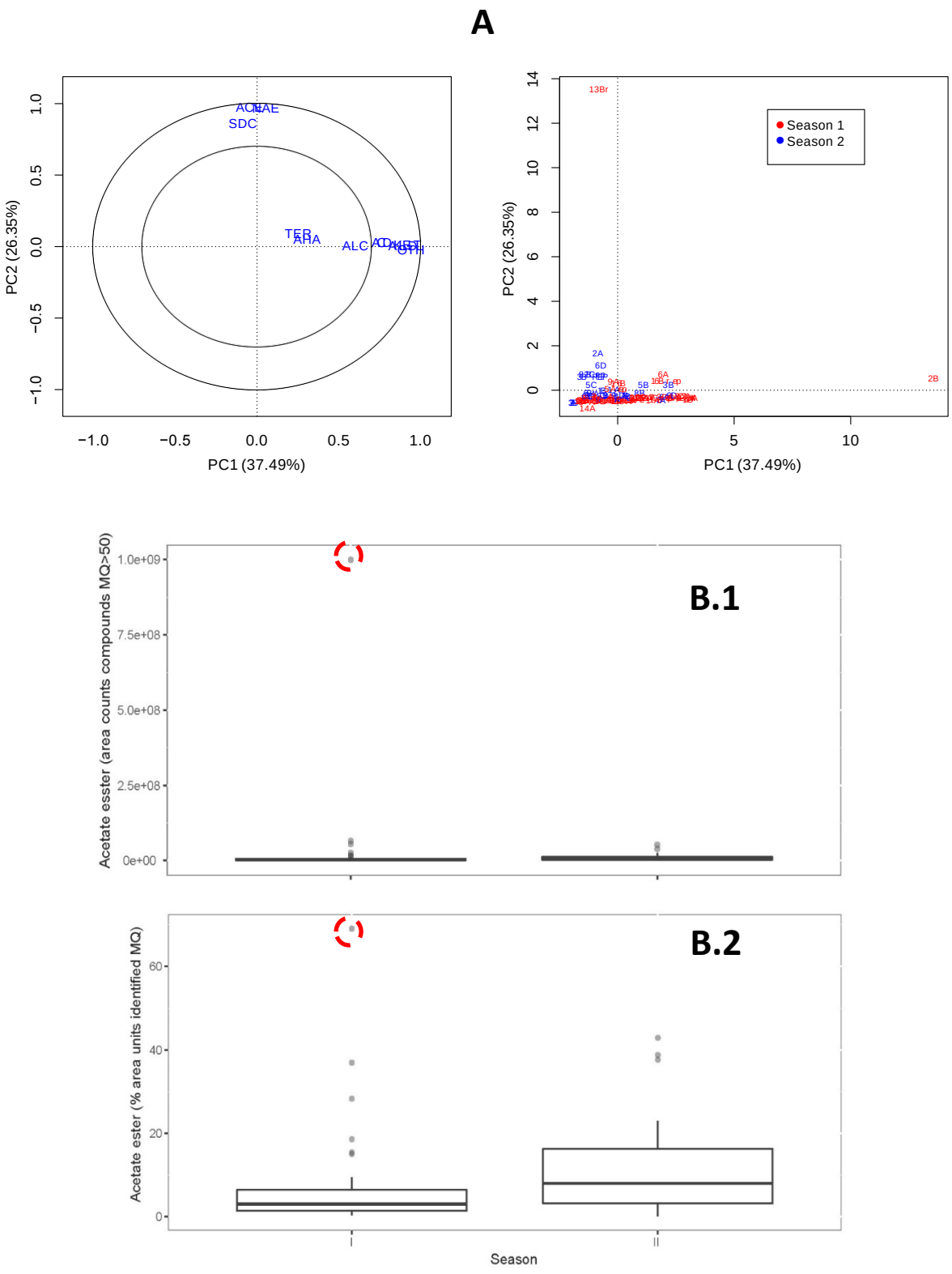
Methodology to remove strong ouliers of non-climacteric melon fruit aroma at harvest obtained by HS-SPME GC-MS analysis

Juan Pablo Fernández-Trujillo[1*], Mohamed Zarid[1] and María Carmen Bueso[2]

1 **Department of Agricultural and Food Engineering, Regional Campus of International Excellence "Campus Mare Nostrum" (CMN), Technical University of Cartagena (UPCT), Paseo Alfonso XIII, 48, ETSIA. E-30203 Cartagena, Murcia, Spain; Juanp.fdez@upct.es (J.P.F.-T.); juanp.fdez@upct.es; m.zarid@gmail.com (M.Z.)**

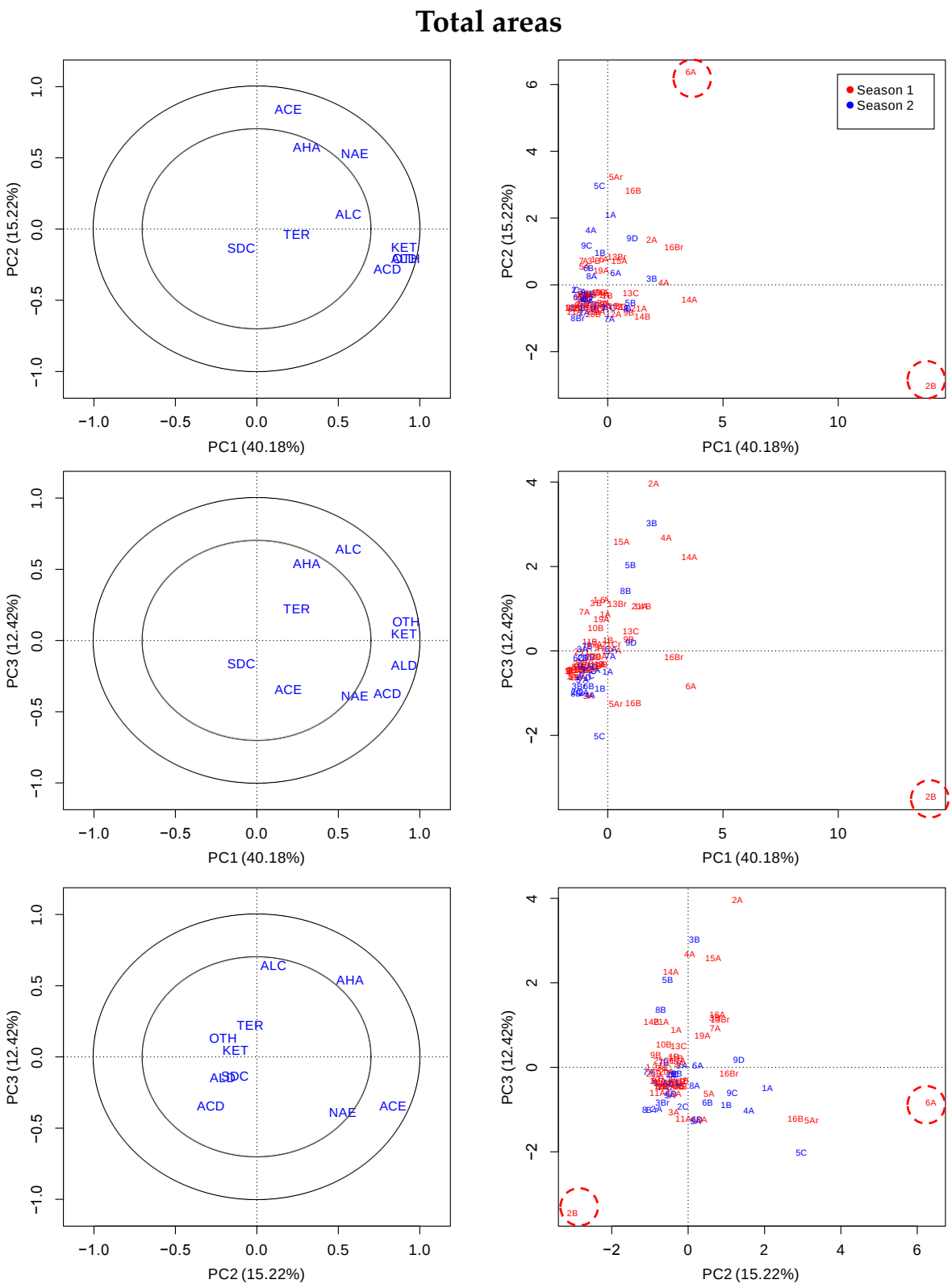
2 **Department of Applied Mathematics and Statistics, CMN, UPCT, Doctor Fleming s/n, ETSII. E-30202 Cartagena, Murcia, Spain. mcarmen.bueso@upct.es**

***Correspondence:**
Juan Pablo Fernández Trujillo
Juanp.fdez@upct.es



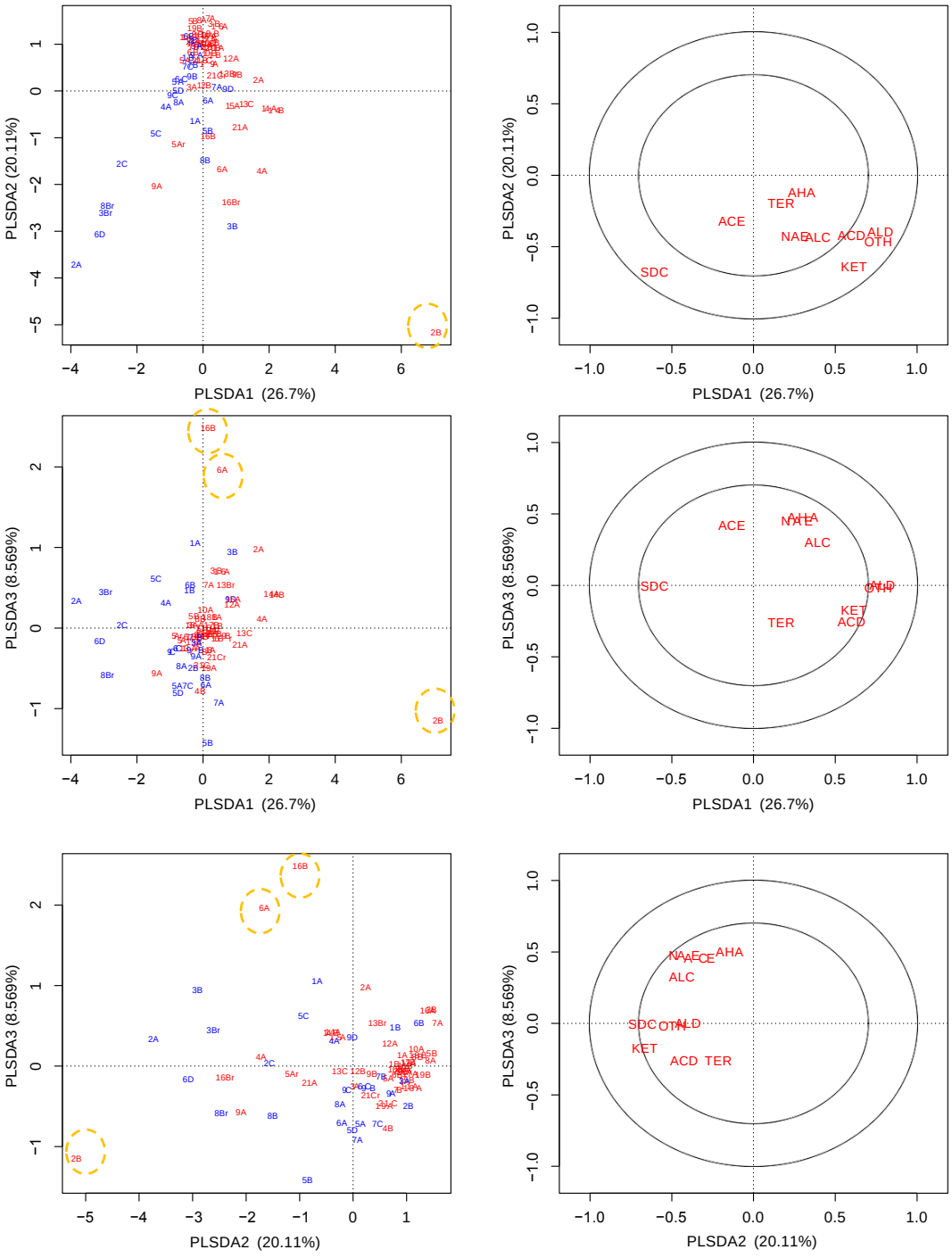
Supplementary Figure 1. A. Correlation (left) and score (right) plots of the first two components of the PCA (with 63.84% of variance explained) applied to compound classes of aroma variable based on total areas (only compounds with match quality > 50 without exogenous compounds), obtained in two seasons. This graph contained an outlier (former 13Br, S1) removed in a previous version of this manuscript because it was a mistake due to the confusion with a climacteric cultivar of melon with abundance of esters and sulfur derived compounds. Another data (2B, S2) was considered a real outlier. AEC, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components. **B.** Example of outliers in acetate esters (ACE) of two aroma variables obtained in two seasons. Bold red circles indicate the fruit 13Br fruit of season 1. **B.1.** Box whisker plot of area counts of ACE with MQ>50. **B.2.** Box whisker plot of percentages of ACE based on total areas of compounds with match quality >50 and without exogenous compounds.

Figure 2. A-C



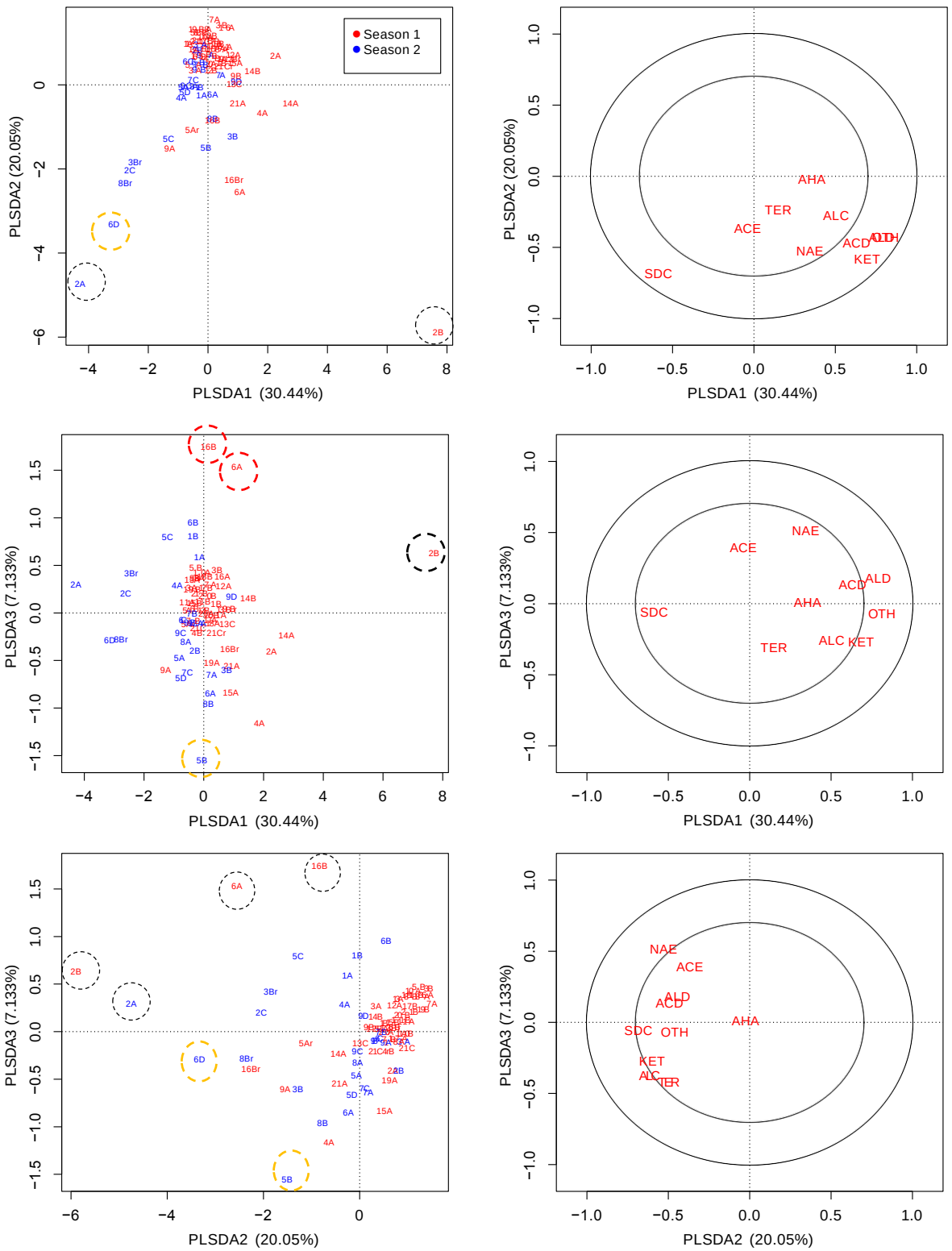
Supplementary Figure 2.A Results of the principal component analysis (PCA) to verify potential outliers (dashed lines in different colors) in compound classes of aroma variable based on total areas (only compounds with match quality > 50), obtained in two seasons. The areas of compound classes were not normalized to the response of the internal standard (1-phenylethanol). ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components.

Total areas

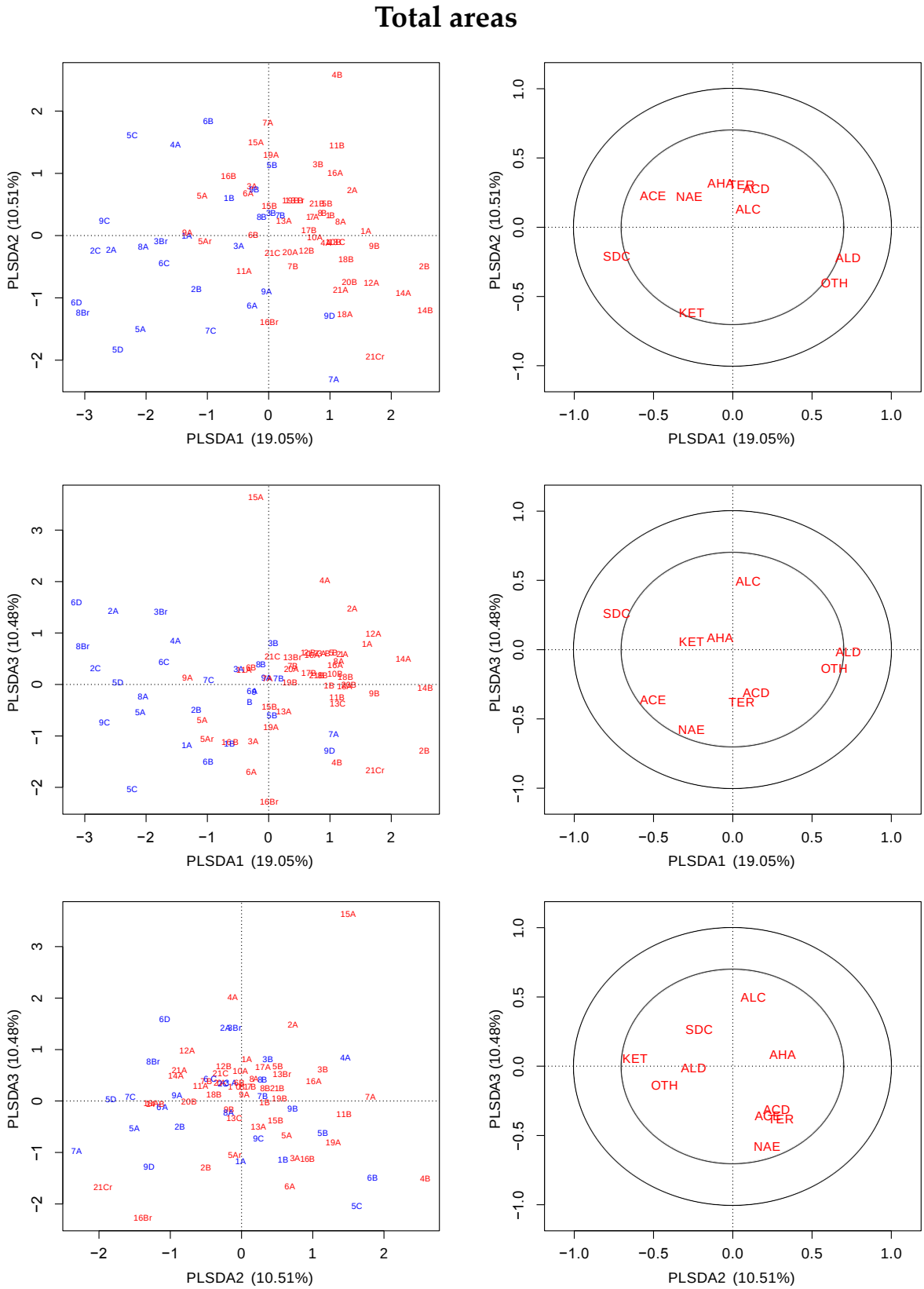


Supplementary Figure 2.B Results of the partial least-squares discriminant analysis (PLS-DA) to verify potential outliers (dashed lines in different colors) in compound classes of aroma variable based on total areas (only compounds with match quality > 50), obtained in two seasons. The areas of the compound classes were normalized to the response of the internal standard (1-phenylethanol) each season. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components.

Total areas



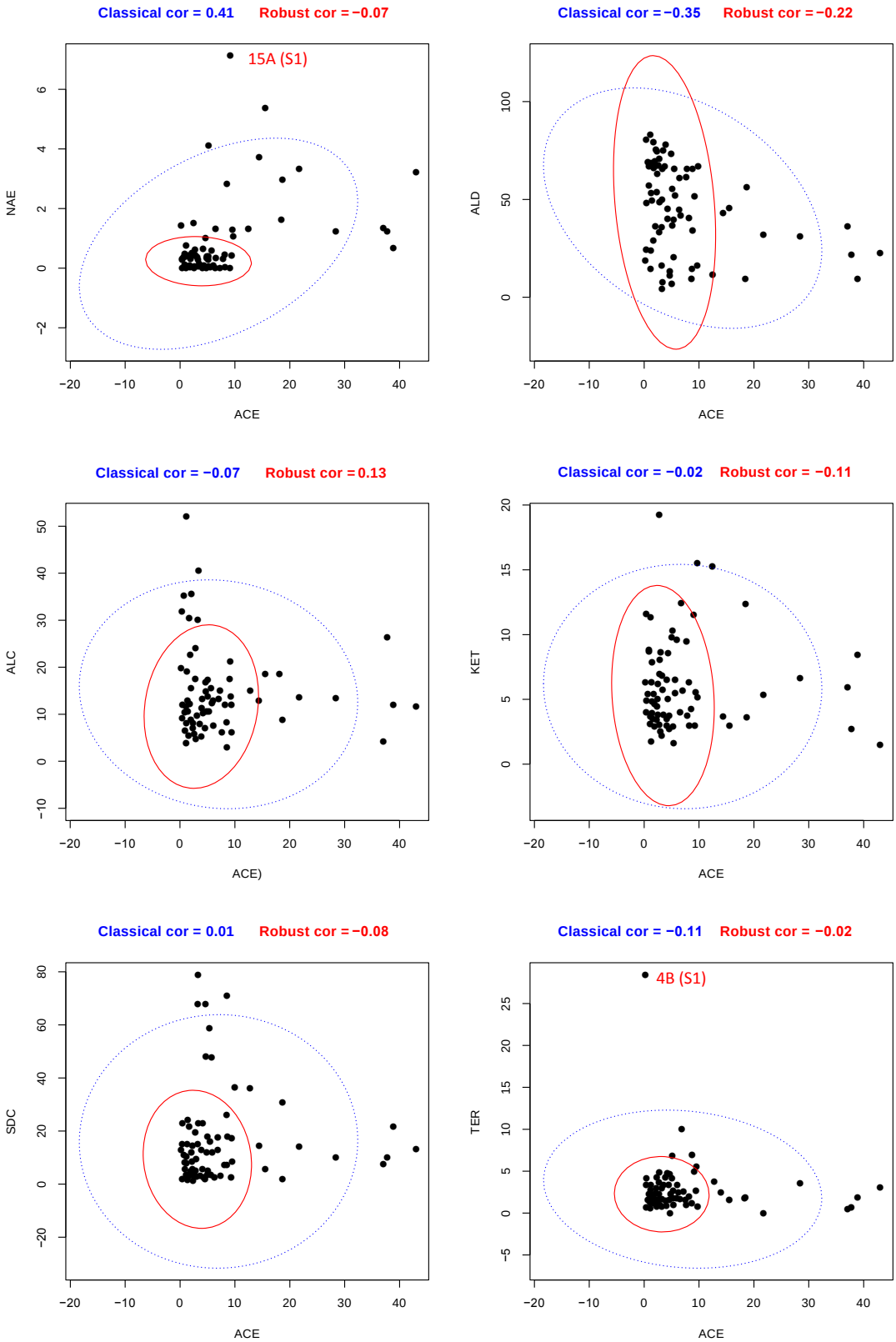
Supplementary Figure 2.C Results of the partial least-squares discriminant analysis (PLS-DA) to verify potential outliers (dashed lines in different colors) in compound classes of aroma variable based on total areas (only compounds with match quality > 50), obtained in two seasons. The areas of the compound classes were not normalized to the response of the internal standard (1-phenylethanol) each season. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components.



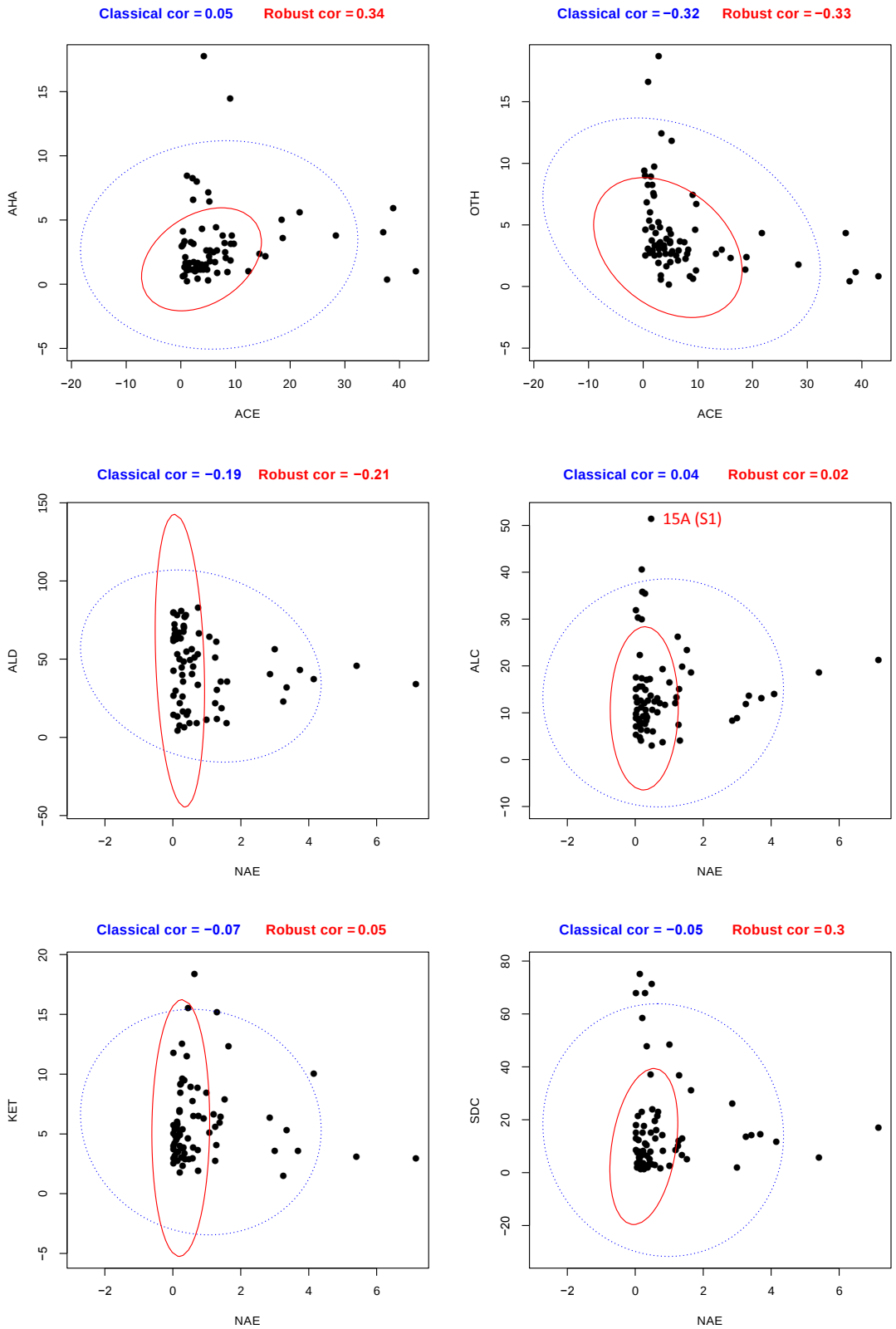
Supplementary Figure 3. Results of the partial least-squares discriminant analysis (PLS-DA) to verify potential outliers in compound classes of aroma variable based on total area counts of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components.

Supplementary Figure 4. A-F

Correlation analysis

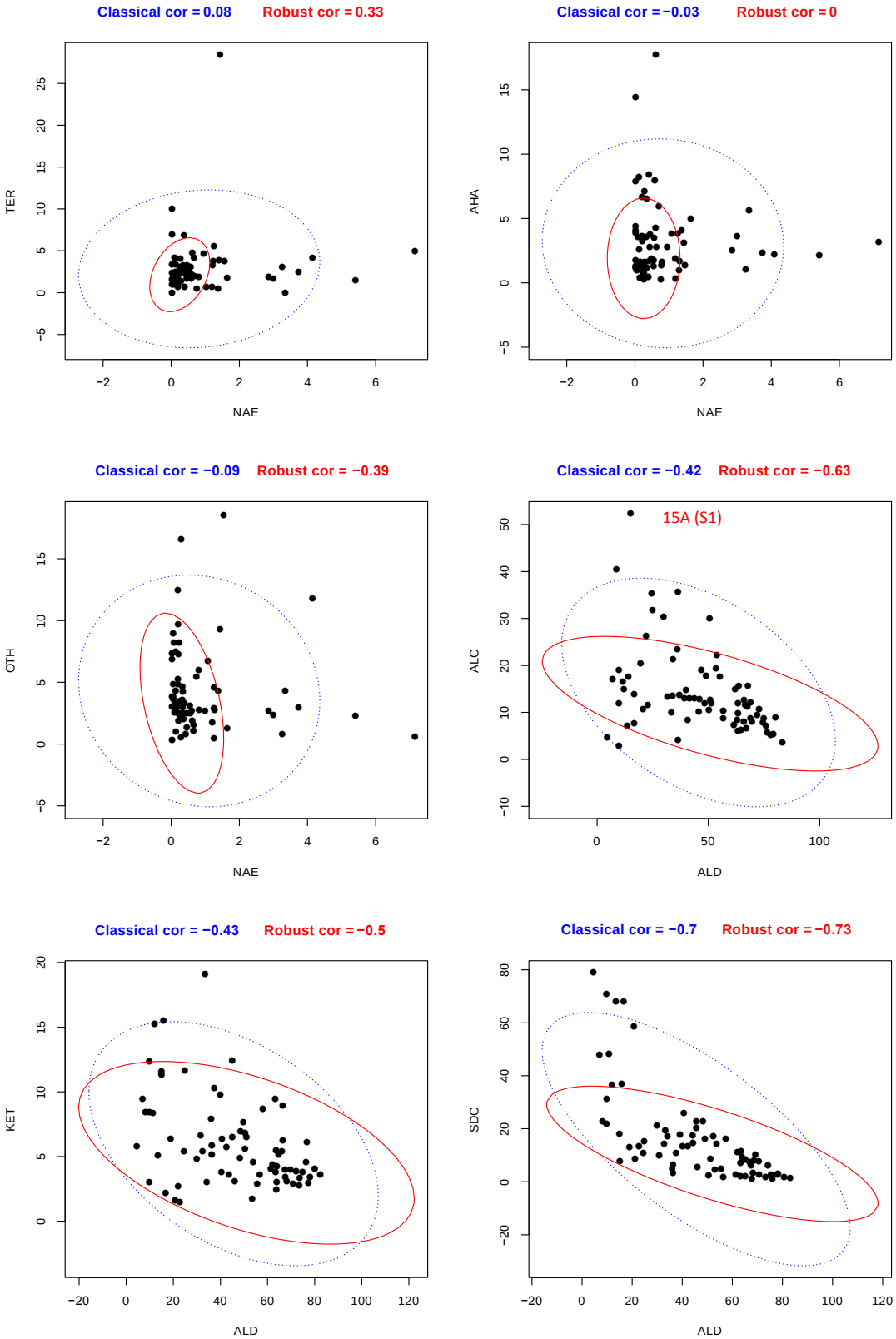


Supplementary Figure 4.A. Results of the correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Classical and robust 97.5% confidence ellipses of the data (blue and red, respectively). Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. The strong outliers labeled are the fruits 4B and 15A from season 1.

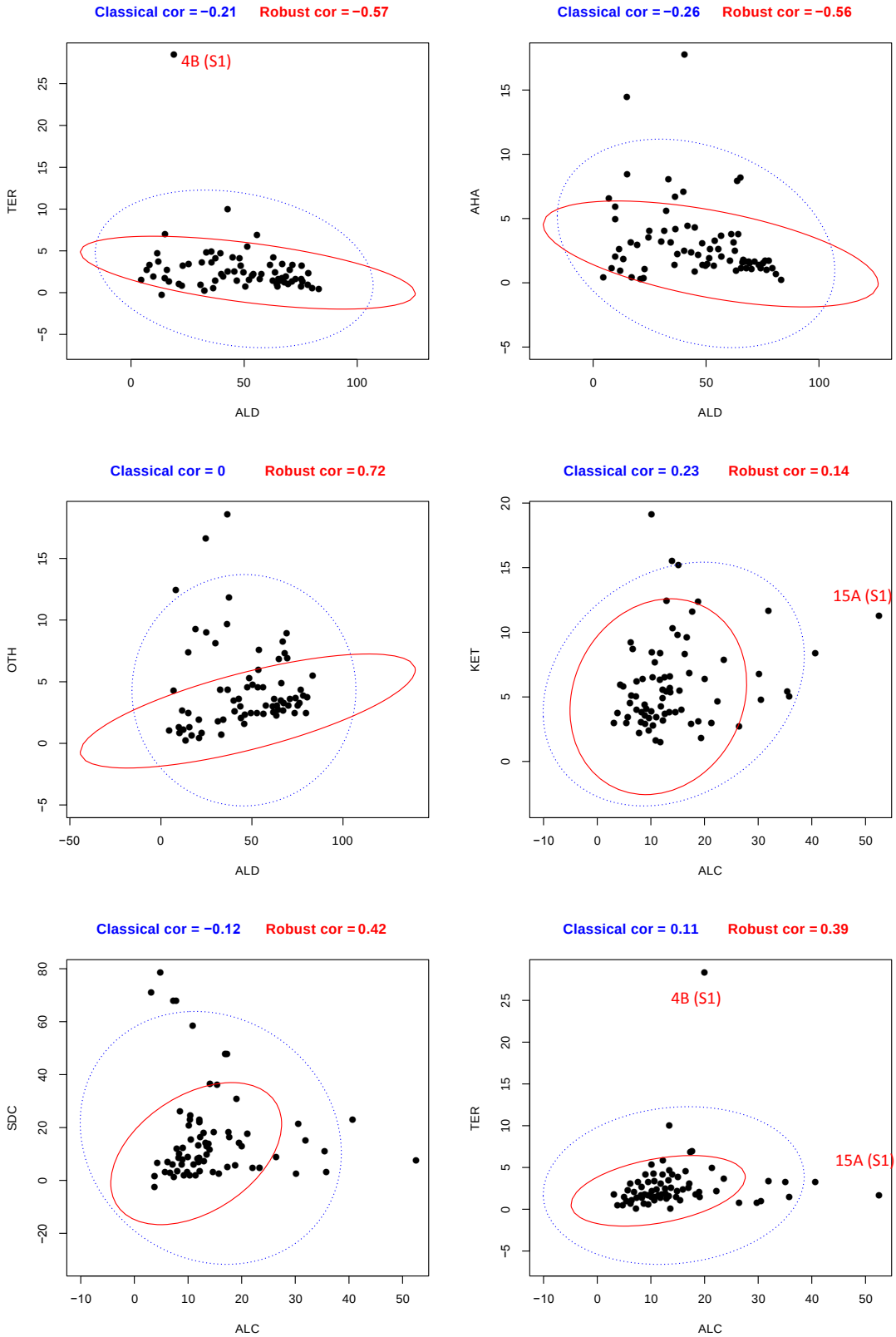


Supplementary Figure 4.B. Results of the correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. The strong outliers labeled are the fruits 4B and 15A from season 1.

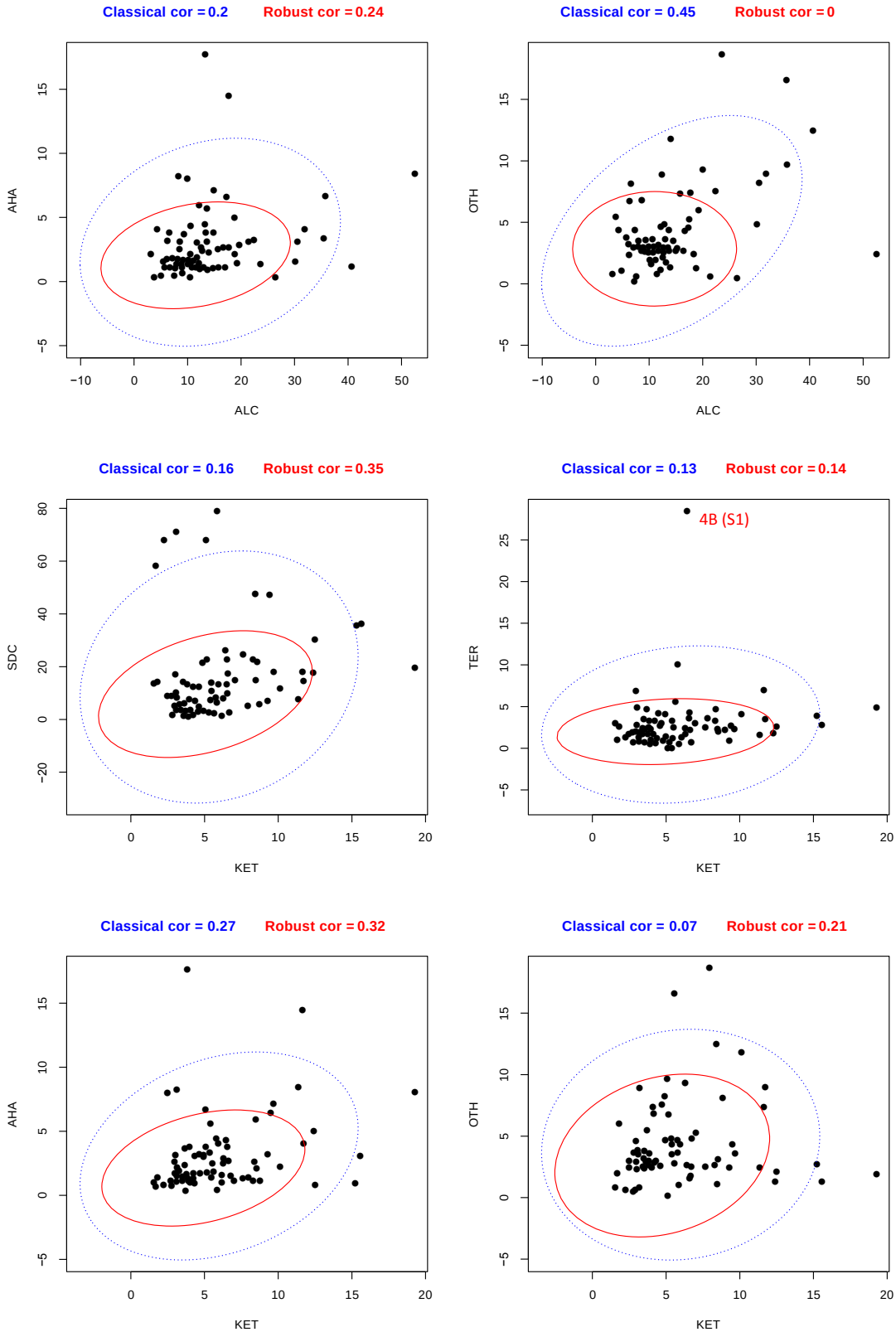
Strong outliers labeled are the fruits 4B and 15A of season 1.



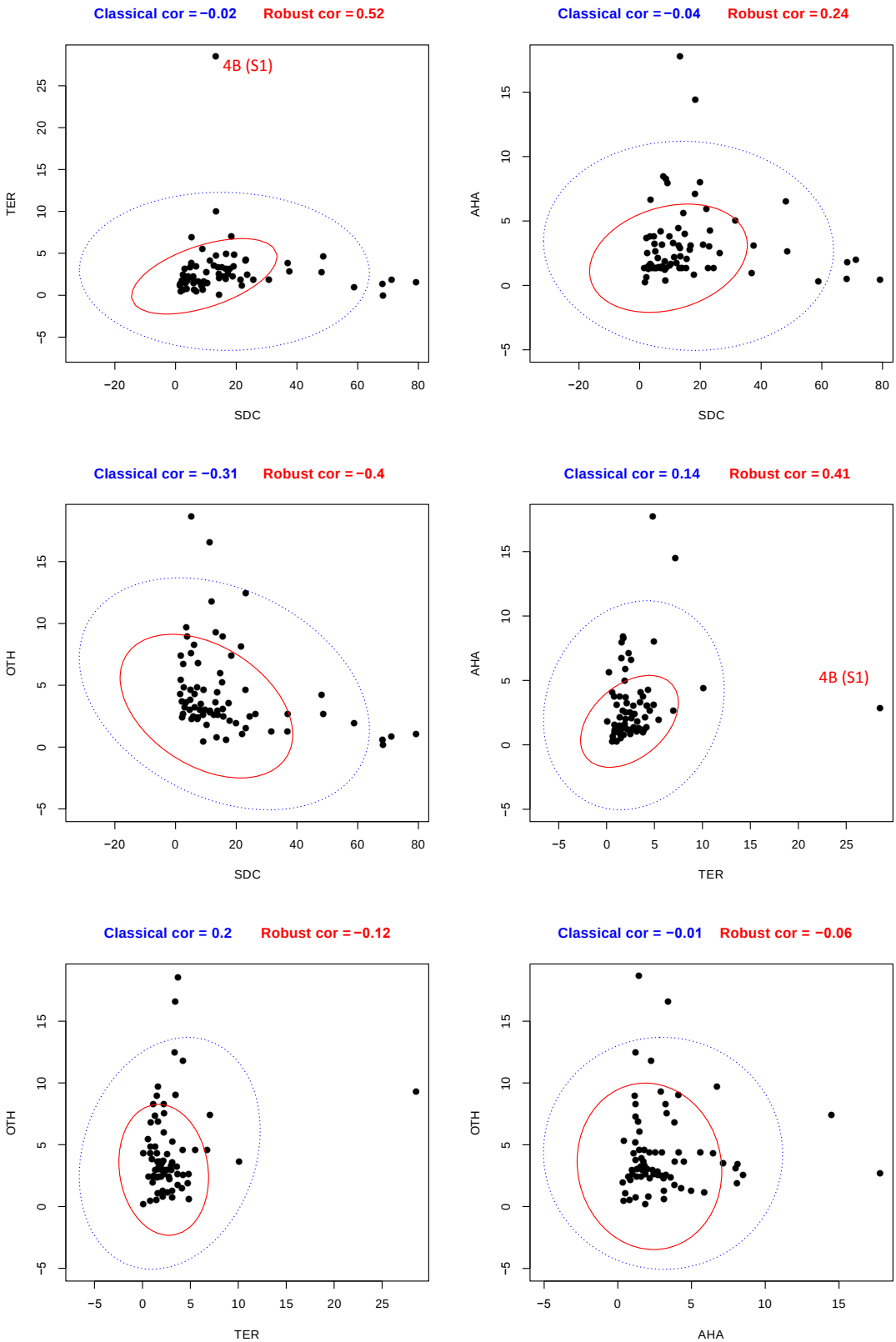
Supplementary Figure 4.C. Results of the correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. The strong outliers labeled are the fruits 4B and 15A of season 1.



Supplementary Figure 4.D. Results of the correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Classical and robust 97.5% confidence ellipses of the data (blue and red, respectively). Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. . Strong outliers labeled are the fruits 4B and 15A of season 1.

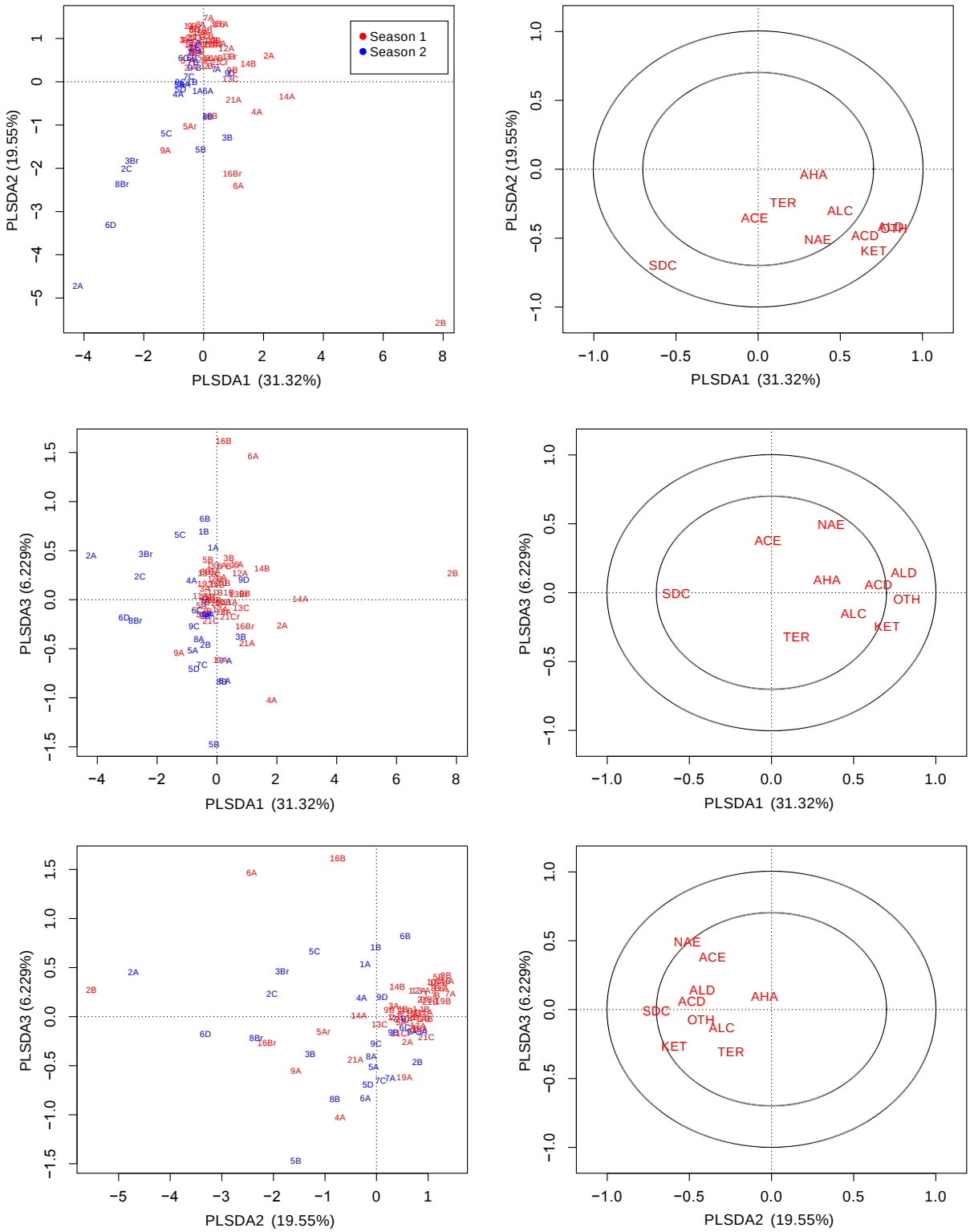


Supplementary Figure 4.E. Results of the correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Classical and robust 97.5% confidence ellipses of the data (blue and red, respectively). Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. The strong outlier labeled is the fruits 4B of season 1.



Supplementary Figure 4.F. Results of correlation analysis to verify potential outliers in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compound. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Classical and robust 97.5% confidence ellipses of the data (blue and red, respectively). Robust method used 50% of the observations for Minimum Covariance Determinant (MCD estimations. The strong outlier labeled is the fruits 4B of season 1.

Percentages of areas



Supplementary Figure 5. Results of partial least-squares discriminant analysis (PLS-DA) after removing strong outliers of fruits 4B and 15A (season 1) in compound classes of aroma variable of percentages of different compound classes based on total areas of such compounds with match quality >50, and without exogenous compounds, obtained in two seasons. ACE, acetate ester; NAE, non-acetate ester; ALD, aldehyde; ALC, alcohols, KET, ketones; SDC, sulfur derived-compounds; TER, terpenes; AHA, alkanes and aliphatic compounds; OTH, other compounds. Circles represent r^2 = 50% and 100% variability explained by the components.