SUPPLEMENTARY MATERIAL

Integration of non-target metabolomics and sensory analysis unravels metabolite signatures associated with dill sensory quality: A case study using dill (*Anethum graveolens*)

Victor CASTRO-ALVES [a,1], Irina KALBINA [a,1], Asgeir NILSEN [b], Mats ARONSSON [c], Eva ROSENQVIST [d], Marcel A K JANSEN [e], Minjie QIAN [a,2], Åsa ÖSTRÖM [b], Tuulia HYÖTYLÄINEN [a], Åke STRID [a,*]

[a] School of Science and Technology, Örebro University, SE-70182 Örebro, Sweden; [b] School of Hospitality, Culinary Arts and Meal Science, Örebro University, SE-71202 Grythyttan, Sweden; [c] Svegro AB, Torslundavägen 20, SE-17996 Svartsjö, Sweden; [d] Section of Crop Sciences, Institute of Plant and Environmental Sciences, University of Copenhagen, Højbakkegård Allé 9, DK-2630 Tåstrup, Denmark; [e] School of Biological, Earth and Environmental Sciences, Environmental Research Institute, University College Cork, North Mall, Cork, Ireland;

[1] These two authors have contributed equally to this work.

[2] Present address: College of Horticulture, Hainan University, Haikou 570228, China

Victor Castro-Alves: victor.castro-alves@oru.se
Irina Kalbina: irina.kalbina@oru.se
Asgeir Nilsen: asgeir.nilsen@oru.se
Mats Aronsson: mats.aronsson@svegro.se
Eva Rosenqvist: ero@plen.ku.dk
Marcel A K Jansen: m.jansen@ucc.ie
Minjie Qian: cookiekin@zju.edu.cn
Åsa Öström: asa.ostrom@oru.se
Tuulia Hyötyläinen: tuulia.hyotylainen@oru.se
Åke Strid: ake.strid@oru.se

* **Corresponding author**: Åke Strid, Örebro University, School of Science and Technology, SE-70182 Örebro, Sweden. E-mail address: ake.strid@oru.se

Table S1. GC Orbitrap analysis performance. Linearity of in-house standards. RI values were obtained either from NIST library or from Golm Metabolome Database.

Base peak (m/z)	Identified compound	Metabolite	RI (database)	RI (calculated)	Base peak area at different concentrations (μg/mL)					Linearity (r^2)
					0.2	1.5	5.0	10.0	15.0	
116.0890	Alanine (2TMS)	Alanine	1087	1106	1.1E+08	1.1E+09	4.9E+09	8.4E+09	1.3E+10	0.9965
144.1203	Valine (2TMS)	Valine	1227	1223	2.0E+08	1.9E+09	7.0E+09	1.1E+10	1.8E+10	0.9936
158.1359	Leucine (2TMS)	Leucine	1279	1275	nd	1.6E+09	6.5E+09	1.1E+10	1.8E+10	0.9945
158.1359	Isoleucine (2TMS)	Isoleucine	1286	1292	1.4E+08	1.6E+09	6.5E+09	1.1E+10	1.6E+10	0.9941
142.1047	Proline (2TMS)	Proline	1295	1294	nd	6.2E+08	5.0E+09	1.0E+10	1.7E+10	0.9956
174.1128	Glycine (3TMS)	Glycine	1305	1302	2.8E+08	2.5E+09	8.4E+09	1.5E+10	2.2E+10	0.9982
254.1153	Catechol (3TMS)	Catechol	1309	1312	2.5E+07	1.9E+08	7.2E+08	1.2E+09	1.8E+09	0.9933
149.0449	Succinic acid (2TMS)	Succinic acid	1310	1314	9.6E+07	6.5E+08	2.2E+09	3.7E+09	5.4E+09	0.9945
245.0660	2-Butenedioic acid (2TMS)	Fumaric acid	1353	1354	8.7E+07	8.5E+08	3.4E+09	5.7E+09	8.7E+09	0.9959
204.1234	Serine (3TMS)	Serine	1352	1362	6.3E+07	1.0E+09	4.9E+09	8.2E+09	1.4E+10	0.9902
218.1391	Threonine (3TMS)	Threonine	1387	1387	2.9E+07	3.7E+08	1.5E+09	2.5E+09	4.2E+09	0.9917
104.0620	Hydrocinnamic acid (1TMS)	Hydrocinnamic acid	1423	1418	4.0E+07	4.0E+08	1.9E+09	3.2E+09	4.9E+09	0.9951
155.0999	Glutamine [-H2O] (2TMS) BP	Glutamine	1469	1471	nd	8.3E+07	4.0E+08	6.0E+08	9.4E+08	0.9869
149.0448	Malic acid (3TMS)	Malic acid	1479	1486	1.1E+08	4.8E+08	1.6E+09	2.6E+09	4.3E+09	0.9939
179.0886	4-Hydroxybenzyl alcohol (2TMS)	4-Hydroxybenzyl alcohol	1500	1499	nd	8.3E+08	3.1E+09	4.8E+09	7.7E+09	0.9914
267.0869	Salicylic acid (2TMS)	Salicylic acid	1506	1502	1.2E+08	1.3E+09	5.1E+09	8.6E+09	1.3E+10	0.9963
128.0890	Methionine (2TMS)	Methionine	1515	1514	nd	5.5E+08	3.0E+09	5.3E+09	9.4E+09	0.9904
156.0839	L-5-Oxoproline (2TMS)	Pyroglutamic acid	1522	1515	1.6E+08	2.1E+09	7.6E+09	1.2E+10	1.8E+10	0.9914
232.1183	Aspartic acid (3TMS)	Aspartic acid	1522	1516	nd	7.8E+08	3.6E+09	6.8E+09	1.1E+10	0.9989
220.1004	Cysteine (3TMS)	Cysteine	1550	1551	nd	3.3E+08	2.5E+09	4.3E+09	7.1E+09	0.9931
292.1341	Threonic acid (4TMS)	Threonic acid	1546	1556	nd	4.8E+08	1.7E+09	2.7E+09	4.4E+09	0.9907
164.0651	Acetic acid, 3-hydroxyphenyl- (2TMS)	3-Hydroxybenzeneacetic acid	1613	1603	nd	2.1E+08	9.2E+08	1.6E+09	2.3E+09	0.9927
246.1340	Glutamic acid (3TMS)	Glutamic acid	1614	1615	nd	5.9E+07	1.1E+09	2.5E+09	3.8E+09	0.9952
267.0867	Benzoic acid, 3-Hydroxy- (2TMS)	3-Hydroxybenzoic acid	1633	1562	9.8E+07	9.2E+08	3.6E+09	5.9E+09	9.5E+09	0.9953
218.1027	Phenylalanine (2TMS)	Phenylalanine	1622	1619	nd	2.4E+08	2.5E+09	5.0E+09	7.9E+09	0.9966

nd: not detected.

Continue…

Base peak (m/z)	Identified compound	Metabolite	RI (database)	RI (calculated)	Base peak area at different concentrations (μg/mL)					Linearity (r^2)
					0.2	1.5	5.0	10.0	15.0	
217.1075	Arabitol (5TMS)	Arabitol	1712	1709	9.6E+07	9.7E+08	2.3E+09	3.9E+09	5.9E+09	0.9949
267.0504	Vanillic acid (2TMS)	Vanillic acid	1756	1757	6.3E+07	5.4E+08	2.1E+09	3.5E+09	5.1E+09	0.9954
204.0996	Shikimic acid (4TMS)	Shikimic acid	1794	1804	9.2E+07	1.3E+09	5.5E+09	9.1E+09	1.3E+10	0.9918
142.1047	Ornithine (4TMS)	Ornithine	1806	1808	nd	1.2E+08	2.3E+09	4.6E+09	7.1E+09	0.9949
149.0448	Cinnamic acid, 2-hydroxy-, trans- (2TMS)	Coumaric acid	1812	1810	nd	3.6E+08	1.5E+09	2.4E+09	3.7E+09	0.9914
179.0522	Acetic acid, 3,4-dihydroxyphenyl- (3TMS)	Homoprotocatechuic acid	1823	1819	7.5E+07	1.0E+09	4.2E+09	7.0E+09	1.0E+10	0.9940
297.0608	Syringic acid (2TMS)	Syringic acid	1888	1887	nd	4.0E+08	1.7E+09	2.7E+09	4.1E+09	0.9906
179.0522	3,4-Dihydroxyhydrocinnamic acid (3TMS)	Hydrocaffeic acid	1941	1933	nd	1.4E+09	6.1E+09	1.0E+10	1.5E+10	0.9911
202.1046	Indole-3-acetic acid (2TMS)	Indole-3-acetic acid	1970	1951	nd	9.7E+08	4.5E+09	7.2E+09	1.1E+10	0.9897
202.1046	3-Indolepropionic acid (2TMS)	3-Indolepropionic acid	2070	2068	nd	3.0E+08	1.3E+09	2.1E+09	3.3E+09	0.9923
219.0472	Caffeic acid, trans- (3TMS)	Caffeic acid	2135	2125	nd	7.9E+08	4.1E+09	6.9E+09	1.1E+10	0.9947
361.1681	Sucrose (8TMS)	Sucrose	2622	>2500	1.7E+08	9.1E+08	2.8E+09	4.6E+09	6.8E+09	0.9944
368.1654	Catechin (5TMS)	Catechin	2865	> 2500	nd	1.6E+08	1.4E+09	2.7E+09	4.4E+09	0.9957

nd: not detected.

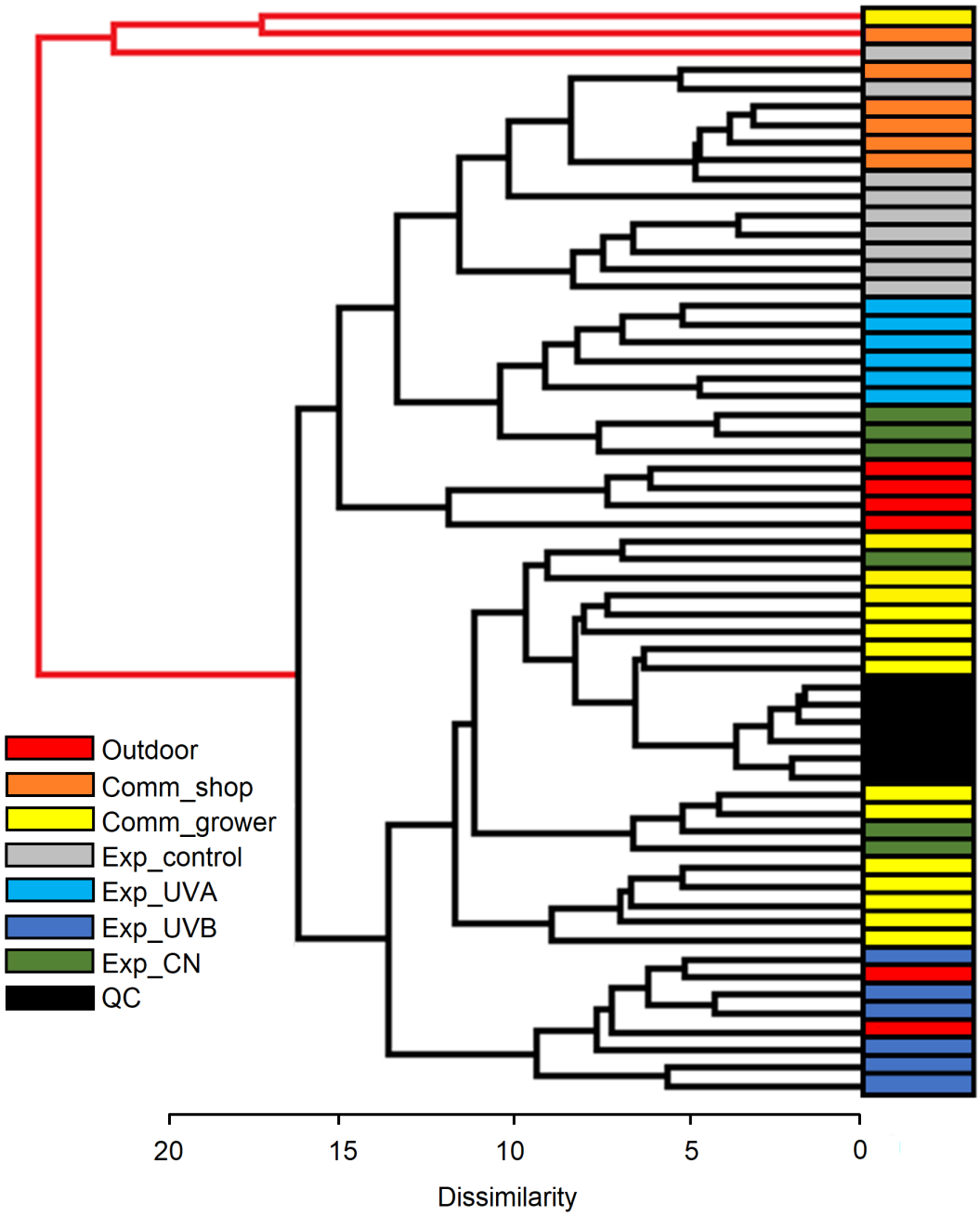

Fig S1. GC-Orbitrap analysis performance. Hierarchical Cluster Analysis (HCA) revealed that quality control samples (QC) from distinct GC-Orbitrap analysis batches were clustered together with low dissimilarity values compared to other samples. HCA also indicates three biological replicates (5.5% of samples) with high dissimilarity compared to other samples (these were one Comm_shop, one Comm_grower, and one Exp_control). These replicates with high dissimilarity were excluded from the dataset before further analysis. Dissimilarity values refers to Euclidean measured distances (complete-linkage clustering).

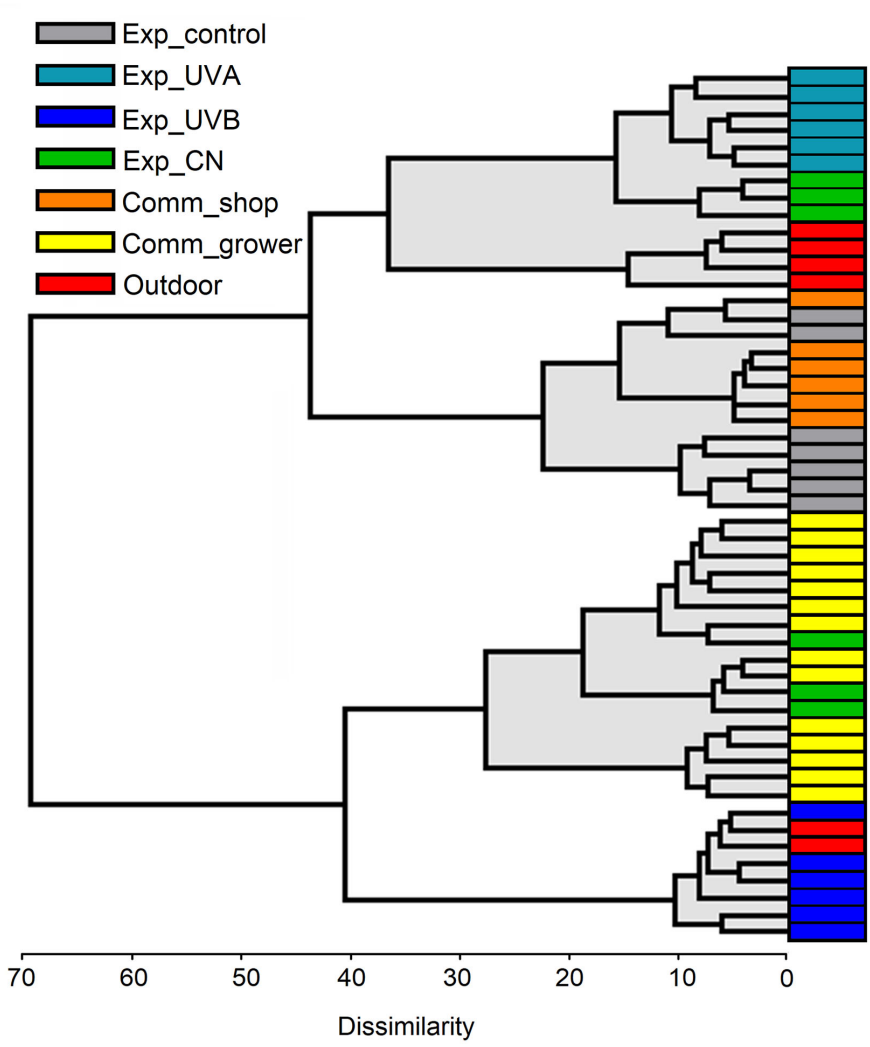

Fig S2. Metabolite profiling of dill obtained over two consecutive years. Hierarchical cluster analysis (HCA) indicates four main clusters of samples, as observed in PCA (Fig. 2A). Quality control samples and dill samples with high dissimilarity values were excluded from the analysis. Dissimilarity values refers to Euclidean measured distances (complete-linkage clustering).

Table S2. Sensory descriptors. The most frequent descriptors (> 15% of frequency) defined by the assessors during sensory analysis.

Descriptors	Frequency (%)
Year 1 (29 assessors)	
Sour*	14 (48%)
Grass*	14 (48%)
Sweet*	13 (45%)
Bitter*	12 (41%)
Spicy*	9 (31%)
Dill*	8 (27%)
Lemon*	7 (25%)
Peppery	6 (21%)
Aniseed, liquorice, fennel	6 (21%)
Fresh	5 (17%)
Celery	5 (17%)
Year 2 (17 assessors)	
Grass*	17 (100%)
Bitter*	13 (76%)
Spicy*	10 (58%)
Dill*	7 (41%)
Sour*	6 (35%)
Vegetable	6 (35%)
Sweet*	5 (29%)
Lemon*	5 (29%)
Harsh	5 (29%)
Herb	4 (24%)
Earthy	3 (18%)
Floral	3 (18%)

*: descriptors with high frequency (at least 25% frequency) in both sensory panels.

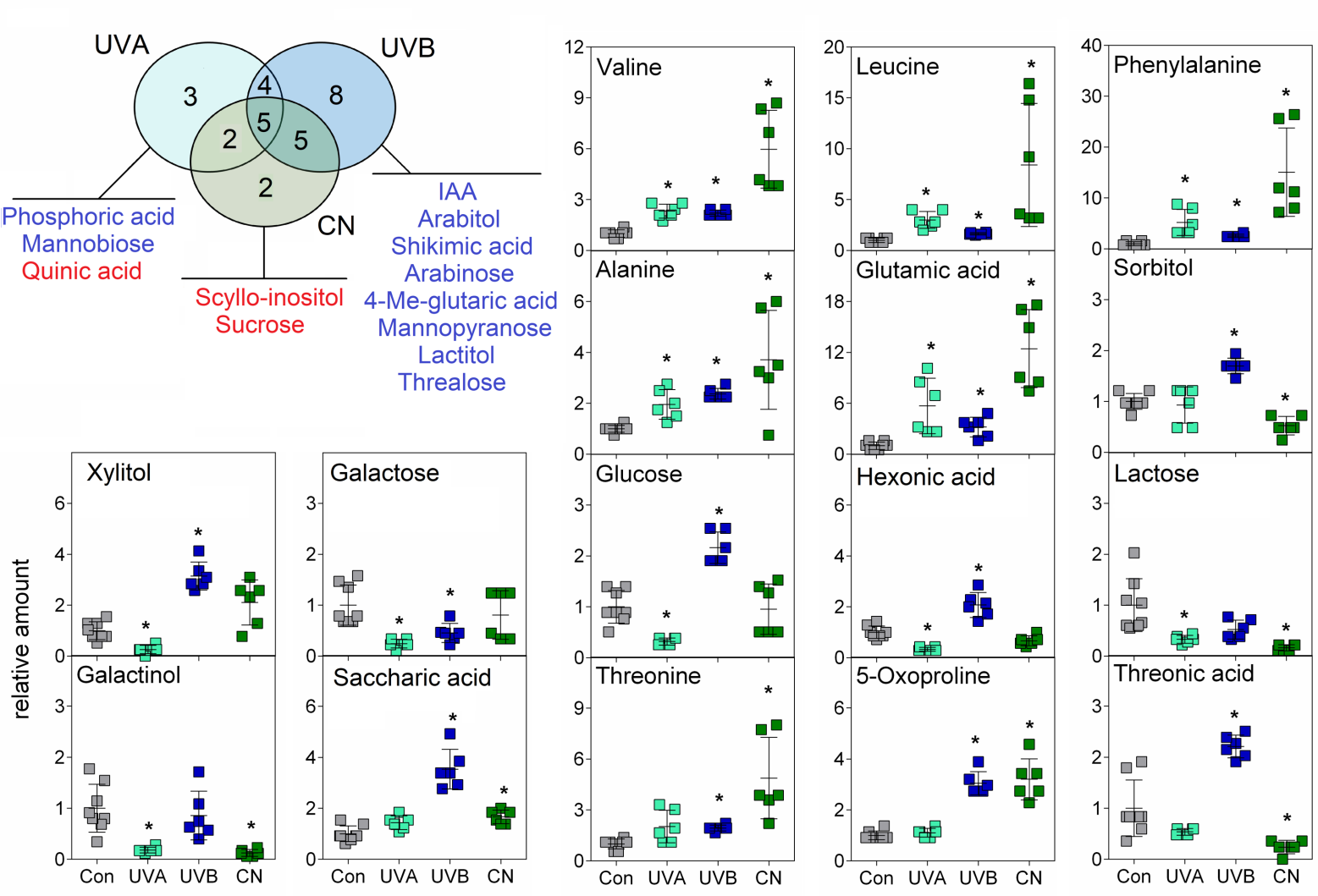

Fig S3. Differences in the metabolite profile of the Exp_control and Exp_UVA, Exp_UVB and Exp_CN. Venn diagram showing the number of compounds with significant differences when compared to the Exp_control. Metabolites that significantly changed in only one of the treatments (n = 13) are shown in blue (higher levels compared to Exp_control) and red (lower levels compared to Exp_control). The relative levels of the 16 metabolites with significant differences between Exp_control in at least two treatments are shown in the scatter plots. *: significant difference compared to the control (t-test, $p < 0.01$). Results represent mean ± SD in relation to control (n = 8/6, control/treated samples).