SUPPLEMENTARY MATERIAL

Agronomic, metabolomic and lipidomic characterization of Sicilian *Origanum vulgare* (L.) ecotypes

by Teresa Tuttolomondo¹, Federico Martinelli^{1,2*}, Lorenzo Mariotti³, Claudio Leto¹, Antonella Maggio⁴, Salvatore La Bella¹

¹Dipartimento di Scienze Agrarie e Forestali, viale delle scienze Ed. 4. Palermo, 90128, Italy:

²I.E.M.E.S.T. Istituto Euro Mediterraneo di Scienza e Tecnologia, Via Emerico Amari, 123, 90139 Palermo, Italy

³Dipartimento di Biologia. via Ghini 13, 56124, Pisa, Italy

⁴Dipartimento STEBICEF. Università degli Studi di Palermo, viale delle scienze, 90128 Palermo, Italy.

*Corresponding author: Tel: +393318039998. Email: federico.martinelli@unipa.it

ABSTRACT

Although *Origanum vulgare* (L.) has been deeply analyzed at phytochemical level, poor knowledge is available regarding non-volatile compounds such as lipids. The aim of this work was to characterize five wild Sicilian *Origanum* ecotypes from an agronomic, metabolomic and lipidomic perspective. Serradifalco presented higher dry weight and inflorescences/plant than the others while Favara had a significantly higher number of branches per plant and more extensive flowered stratum. Metabolomic analysis, performed with LC-MS-TOF, allowed a preliminary characterization of the non-volatile metabolome of the five oregano ecotypes *Origanum vulgare* ssp. *hirtum*. Twenty-five metabolites were identified belonging to organic acids, aminoacids, lysophosphatidylcholines, carnithines, nucleic bases and lysophosphatidylethanolamines. Lipidomic analysis identified 115 polar plant membrane glycerolipid species. 13 of them were differentially present in the two chosen ecotypes. The role of these metabolites in plant physiology and from a qualitative and pharmacological point of view was discussed.

Keywords: ecotypes, lipidomics, metabolomics, *Origanum vulgare*.

SUPPORTING INFORMATION DESCRIPTION

EXPERIMENTAL

Plant material. Five ecotypes of *O. vulgare* spp. *hirtum* (Link) Ietswaart [Sin.: *O. heracleoticum* L.] were selected based on different geographical positions in Sicily. These ecotypes were named based on the collection place: Serradifalco (S) (province of Catania), Villalba (V) (Agrigento), Naro (N) (Agrigento), Linguaglossa (L) (Caltanissetta) and Favara (F) (Agrigento).

They were identified in natural areas characterized by varying altitudes and pedological conditions. Small bush plants were collected and cultivated as mother plants to obtain cuttings that were used for the experimental field. They were cultivated from 2011 - 2013 at the University of Palermo Department of Agricultural and Forest Sciences (38°06'N 13°20'E). The experimental layout/design consisted of randomized blocks. The soil was an "ARIC Regosol" (World Reference Base for Soil Resources). Shallow, limestone-rich (pH = 7.61), with a sandy loam texture 53.92% sand, 22.65% silt and 23.45% clay and an organic matter content of 1.40%. Agronomic managements (**Tab. S4**) consisted of fertilization at pre-implantation 90 kg/ha of P and 100 kg/ha of potassium), irrigation at post-tranplantation together with weed management using mechanical weeding. No phytosanitary treatment was performed during experimentation. At harvest (end of May 2013) the following agronomic parameters were determined: dry plant weight (g), total number of branches and stems per plant ramification (n), dry weight of inflorescences, leaves and stems (g), percentages of flowers, stems and leaves, height of flower stratum (cm), length per plant (cm).

LC-Q-TOF-MS Metabolite Profiling Analysis. The procedure was similar to the one published⁶. A pool of 10 plants per block were analyzed for a total of 4 replicates for each biotype. Flowers were ground in liquid nitrogen and lyophilized. 2 ml of prechilled extraction solvent MeOH:CHCl₃:1:1 (v/v) were added to 20 mg of ground tissues, stirred at 4°C per 5 min and then centrifuged at 6000 rpm for 2 min. An aliquot of supernatant was completely dried in a SpeedVac concentrator. Derivatization was carried out with methoxyamine and N-methyl-N-trimethylsilyl-trifluoroacetamide and separation was obtained with a LC-Q-TOF-coupled to a quadrupolar mass spectrometer (MS) and equipped with a Zorbax-C 8 column. Samples (1 µl) were introduced in both splitless and split modes. In splitless manner, a flow of helium was used at 10.5 ml/min for 1 min (8 psi). In split mode (1:10), a split flow rate was employed at 10.3 ml/min. A source temperature of 230°C was used. oven temperature increased 10°C/min from 60 °C (1 min start time) to 325°C 10 min (final conditions). The run time was 37 min, following which, it was cooled to 60°C. The

carrier gas was flowed at a constant rate of 1 ml/min. The MS was set at 150°C, with a 5.90 min solvent delay time, the scan was performed from 50 to 600 atomic mass units. Prior to the acquisition, the MS was autotuned using FC43 according to the instrument manual.

Data acquisition and peak identification. Relative concentrations of metabolites were determined by peak areas. The identification of metabolites was performed as a service from Fiehn Lab (Metabolomics facility, UC Davis) using the BinBase algorithm (rtx5) using these settings: validity of chromatogram (<10 peaks with intensity >10⁷ counts s⁻¹), unbiased retention index marker detection (MS similarity>800, validity of intensity range for high m/z marker ions), retention index calculation by 5th order polynomial regression. Spectra were cut to 5% base peak abundance and matched to database entries from most to least abundant spectra using the following standard matching filters: retention index window ±2.000 units (equivalent to about ±2 s retention time), validation of unique ions and apex masses, mass spectrum similarity must fit criteria dependent on peak purity and signal/noise ratios and a final isomer filter. Quantification was reported as peak height using the unique ion as default, unless a different quantification ion is manually set in the BinBase administration software BinView. A quantification report table was obtained for all database entries that were positively detected in more than 10% of the samples of a study design class (as defined in the miniX database) for unidentified metabolites. The Agilent Fiehn LC/MS Metabolomics RTL Library was used for metabolite identifications. This library contains searches for more than 700 common metabolites in relation to their LC/MS EI spectra and retention time indexes. All peak detections were manually checked for false positive and false negative assignments, corrected with retention time locking to decrease run-to-run retention time variation. Universal or class-specific internal standards in GC-MS based analysis, within each chemical class, were used.

Lipidomic analysis. An aliquot of completely dried supernatant was used for lipidomics analysis. Data were acquired using the following parameters. As a column, a Waters Acquity UHPLC CSH C18 (100 mm length x 2.1 mm internal diameter; 1.7 µm particles) was used. The mobile phase A was composed of 60:40 v/v acetonitrile:water + 10 mM ammonium formate + 0.1% formic acid. Mobile phase B was 90:10 v/v isopropanol:acetate. Column temperature was set at 65°C, flow-rate at 0.6 ml/min, 1.67 µl (for ESI +) and 5 µl were injected for ESI at 4°C. ESI capillary voltage was +3.5 kV; scan range: m/z 60–1700 Da. spectral acquisition speed: 2 spectra/s. Mass resolution: 10.000 (for ESI +) on an Agilent 6530 QTOF MS.

The analytical UHPLC column was protected by a short guard column which we replaced

after 400 injections while the UHPLC column was replaced after 1.200 serum or plasma) extract injections. An automatic valve was used switching after each injection to reduce sample carryover for highly lipophilic compounds such as triacylglycerols (TAGs) from 29% to 0.1%. This valve switching employs a dual solvent wash, first with a water/acetonitrile mixture (1:1. v/v) and subsequently with a 100% isopropanol wash.

Lipidomic data analysis. Raw data were processed in an untargeted qualitative manner by Agilent’s software MassHunter Qual v. B.05.00 to find peaks in up to 300 chromatograms. Peak features are then imported into Mass Profiler Professional for peak alignments to seek which peaks are present in multiple chromatograms, using exclusion criteria by the minimum percentage of chromatograms in which these peaks are positively detected. In a tedious manual process, these peaks were then collated and constrained within the MassHunter quantification software v. B.05.01 on the accurate mass precursor ion level, using the MS/MS information. The LipidBlast library was used to identify lipids with manual confirmation of adduct ions and spectral scoring accuracy. MassHunter Quan enables back-filling of quantifications for peaks that were missed in the primary peak finding process, hence yielding data sets without missing values. Peaks were identified in manual comparison of MS/MS spectra and accurate masses of the precursor ion to spectra given in the Fiehn laboratory’s LipidBlast spectral library 67.

Statistical analysis. Agronomic, metabolomic and lipidomic normalized data were statistically analyzed using ANOVA and t-test to identify significant differences between ecotypes. Principal component analyses (PCA) was used to reduce the dimensionality of the gene expression data using SAS software. Data analysis was performed to alleviate possible bias caused by the amount of collected material for each class or other confounding factors. Principal component analysis was then applied to the ratio matrix of agronomic, metabolite or lipid data list to examine the contribution of each target parameter to the separation of the sample classes. A biplot was constructed based on the first two principal components. Graphics were performed using Graphpad Prism software.

SUPPLEMENTARY TABLES

Table S1. Agronomic parameters for the five *Origanum* Sicilian ecotypes. Average and standard deviations of the four randomized blocks for each ecotype were indicated. Different letters showed significant differences in ANOVA (p-value < 0.05).

Agronomic	Serradifalco	Villalba	Naro	Linguaglossa	Favara

parameter					
Dry weight (g)	958.4±131.4 a	688.9±47.3 b	721.5±132 b	812.1±112.5 ab	697.3±128.8 b
Ramification/ plant	433±68.0 b	295±20.4 d	361.25±29.7 c	324±49.3 cd	643±34.6 a
Inflorescences (g)	264.1±34.9 a	192.9±24.8 b	168.2±30.5 b	208.6±38.5 b	170.3±34.9 b
Stems (g)	502.9±69.6 a	354.0±24.8 ab	438.0±79.1 ab	333.2±190.1 b	424.2±70.2 ab
Leaves (g)	191.4±28.3 a	142.0±10.0 b	115.4±28.2 bc	187.1±31.8 a	102.9±24.8 c
% In flower	27.6±0.6a	27.9±2.2 a	23.3±1.7 b	25.8±4.7 ab	24.4±0.8 ab
% Stems	52.5±0.9 b	21.9±2.0 c	15.9±1.8 d	51.1±7.6 b	61.0±1.5 a
% Leaves	20.0±0.4 d	50.2±0.9 b	60.8±3.3 a	23.1±2.9 c	14.7±0.9 e
Stems/clump	519.0±73.2 a	552.0±103.7 a	337.3±85.1 b	509.5±61.8 b	300.3±62.9 b
Length Stem/Plant	49.8±2.4c	52.4±6.7 c	66.3±0.5 ab	69.1±4.5 a	63.3±1.3 b
Flowered stratum	18.1±2.9 b	17.4±1.0 b	17.9±3.0 b	17.5±1.7 b	24.6±0.8 a

Table S2. Metabolic analysis of the five *Origanum* ecotypes. Average of peak heights and standard deviations were indicated. Different letters showed significant differences in ANOVA (p-value < 0.05).

Metabolites	Serradifalco	Villalba	Naro	Linguaglossa	Favara
2-Aminobenzoic acid	142014.59 ± 22724.60 b	158908.49 ± 15995.43 b	222971.65 ± 15163.30 a	326956.27 ± 66419.12 a	195182.08 ± 8428.30 c
(E)-2-methyl-glutaconic acid	151629.93 ± 25578.62 a	121219.26 ± 8280.79 ab	106138.04 ± 12438.94 bc	95932.21 ± 8192.17 c	62972.05 ± 3228.20 d
4-Guanidinobutanoic acid	666649.66 ± 10984.77 a	355103.80 ± 527.97 b	511358 ± 22679.12 c	744834.85 ± 45206.79 a	526219.30 ± 76538.43 c
Acetylcholine	1321648.67 ± 130755.90 a	733257.91 ± 17677.54 b	1618312.83 ± 256491.10 a	1612445.88 ± 121333.51 a	994166.67 ± 278383.12 ab
Adenosine	380611.52 ± 53184.84 d	928809.03 ± 54870.12 b	259195.04 ± 64924.30 d	2975010.33 ± 213247.89 c	3743107.17 ± 305693.99 a

Carnitine C0:0	2163331.71	1434521.88	2473161.38	2460248.88	1686494.96
	± 181543.51 a	± 383265.81 b	± 468694.31 a	± 2836.74 a	± 337679.49 b
Carnitine C2:0	1144607.94	251731.69	1269586.67	1032009.04	815136.15
	± 43202.01 a	± 40165.82 b	± 222589.79 a	± 37349.16 a	± 114089.50 c
Carnitine C3:0	87777.08	44335.27	73762.55	54044.07	36838.50
	± 4158.29 a	± 9409.35 b	± 34054.47 ab	± 3727.59 b	± 7724.69 b
Carnitine C4:0	40269.84	16945.32	47750.70	28882.29	17009.45
	± 1106.58 c	± 2947.82 b	± 7686.97 a	± 1746.47 d	± 1133.94 b
Carnitine C5:0	20981.53	24442.72	27232.46	19322.52	9179.97
	± 1766.118 a	± 11779.567 a	± 1282.929 a	± 4626.018 a	± 115.664 b
Cytosine	117949.27	143648.82	96705.32	83004.15	102216.18
	± 20320.19 a	± 22439.36 ab	± 15691.14 ab	± 7126.44 b	± 5793.37 a
Glycerophosphocholine	4648106.75	1887913.63	3302525.41	2418841.50	5672490.00
	± 468127.67 c	± 84664.90 b	±1413266.31 cb	± 601430.03 cb	± 144660.48 a
Leucine	25935.38	17557.10	29300.78	320197.39	32645.29
	± 5039.26 c	± 2288.61 b	± 12331.61 bc	± 2915.78 a	± 11065.67 bc
LPC 16:0	82533.84	144689.90	36083.63	390366.50	30750.04
	± 68148.18 a	± 28121.89 ab	± 7351.15 ac	± 155915.41 abc	± 5200.91 ac
LPC 18:0	258888.71	20322.68	11482.85	33134.43	8163.99
	± 193656.86 a	± 2098.78 ab	± 710.15 ac	± 304.19 ad	±1267.79 ac
LPC 18:2	739069.97	601986.62	468093.50	505820.16	320869.30
	± 33175.28 a	± 197704.71 abc	± 49734.44 b	± 13597.71 b	± 8186.59 c
LPC 20:0	27593.43	40446.29	50697.46	75947.27	23903.55

	± 10070.59 a	± 13132.47 ab	± 5087.94 bc	± 14289.60 ac	± 5880.50 a
LPC 20:1	25796.58	24388.20	17423.52	22872.62	8841.17
	± 6491.73 a	± 3324.21 ab	± 2304.81 ab	± 4778.74 a	± 784.81 ac
LPC 20:5	27875.54	36503.44	23741.73	22259.71	22337.60
	± 6835.96 a	± 202.13 a	± 4985.34 a	± 8127.21 a	± 8744.80 a
LPE 16:0	128165.08	324081.72	160313.02	485205.26	182699.38
	± 41760.66 b	± 12416.48 a	± 19595.96 b	± 82897.25 a	± 5289.82 a
LPE 18:3	107696.25	146705.45	53188.96	396284.22	43808.75
	± 25018.64 b	± 72130.54 bc	± 6943.17 c	± 68383.55 a	± 31973.22 bc
PC 32:3	27138.20	86519.66	97036.29	438262.78	86316.04
	± 3027.17 c	± 2663.64 b	± 22243.23 b	± 2895.11 a	± 6305.56 b
Phenylalanine	265127.09	133804.00	104505.73	92155.48	70991.68
	± 24318.49 a	± 7428.60 b	± 25279.35 bc	± 18224.94 c	± 5310.93 c
Proline	67498.38	48269.07	107744.97	118397.15	21213.70
	± 2907.079 a	± 4196.886 b	± 7053.915 c	± 25258.523 abc	± 2526.872 d
Tryptophan	533138.80	173200.69	375680.65	550948.60	384490.64
	± 22933.78 a	± 26064.11 b	± 19067.56 c	± 54075.87 a	± 28903.71 c

Table S3 (excel file). Lipidomic analysis of two *Origanum vulgare* (L.) ecotypes. Two biological replicates were used. Peak area for each metabolite was indicated.

Table S4. Main agronomic parameters of experimental scheme.

Agronomic Characteristics	Details
Transplantation Date	nov-11
Planting	50 cm on row. 100 cm between rows

Experimental scheme	Randomized blocks (four replicates)
Tillage	Plowing at 20-25 cm. harrowing
Weed control	Mechanical
Irrigation	At post-tranplantation
Fertilization at pre-implantation	90 kg/ha di P ₂ O ₅
	100 kg/ha di K ₂ O
Phytosanitary treatments	None

Figure S1. Principal component analysis of agronomic data for the five analyzed *Origanum vulgare* (L.) ecotypes. Vectors for ach agronomic parameter was indicated: Ramf/pl (ramifications/plant), FlowSr (flowered stratum), LengthP (length/plant), Stem/P (stem/plant), Inflow inflorescence (weighth), DW (dry weighth).

Figure S2. Differentially regulated lipids between the two analyzed ecotypes (p-value < 0.1).