Journal of Pharmaceutical and Biomedical Analysis 193 (2021) 113727

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Unraveling the metabolome composition and its implication for *Salvadora persica* L. use as dental brush via a multiplex approach of NMR and LC–MS metabolomics

Mohamed A. Farag[a,b,*], Zeinab T. Shakour[c], Tilo Lübken[e], Andrej Frolov[d,f], Ludger A. Wessjohann[d], Engy Mahrous[a]

[a] Pharmacognosy Department, College of Pharmacy, Cairo University, Kasr el Aini St., P.B. 11562, Cairo, Egypt
[b] Department of Chemistry, School of Sciences & Engineering, The American University in Cairo, New Cairo 11835, Egypt
[c] Laboratory of Phytochemistry, National Organization for Drug Control and Research, Cairo, Egypt
[d] Leibniz Institute of Plant Biochemistry, Dept. Bioorganic Chemistry, Weinberg 3, D-06120, Halle (Saale), Germany
[e] Organic Chemistry, Technische Universität Dresden, Bergstraße 66, 06120 Dresden, Germany
[f] Department of Biochemistry, St. Petersburg State University, St Petersburg, Russia

ARTICLE INFO

Article history:
Received 1 September 2020
Received in revised form 21 October 2020
Accepted 22 October 2020
Available online 2 November 2020

Keywords:
Salvadora persica
Miswak
NMR metabolomics
Proline betaines
Stachydrine
Salvadoside

ABSTRACT

Salvadora persica L. (toothbrush tree, Miswak) is well recognized in most Middle Eastern and African countries for its potential role in dental care, albeit the underlying mechanism for its effectiveness is still not fully understood. A comparative MS and NMR metabolomics approach was employed to investigate the major primary and secondary metabolites composition of *S. persica* in context of its organ type *viz*., root or stem to rationalize for its use as a tooth brush. NMR metabolomics revealed its enrichment in nitrogenous compounds including proline-betaines *i.e*., 4-hydroxy-stachydrine and stachydrine reported for the first time in *S. persica*. LC/MS metabolomics identified flavonoids (8), benzylurea derivatives (5), butanediamides (3), phenolic acids (8) and 5 sulfur compounds, with 21 constituents reported for the first time in *S. persica*. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) of either NMR or LC/MS dataset clearly separated stem from root specimens based on nitrogenous compounds abundance in roots and is justifying for its preference as toothbrush versus stems. The presence of betaines at high levels in *S. persica* (9–12 μg/mg dry weight) offers novel insights into its functioning as an osmoprotectant that maintains the hydration of oral mucosa. Additionally, the previously described anti-inflammatory activity of stachydrine along with the antimicrobial effects of sulfonated flavonoids, benzylisothiocynate and ellagic acid derivatives are likely contributors to *S. persica* oral hygiene health benefits. Among root samples, variation in sugars and organic acids levels were the main discriminatory criterion. This study provides the first standardization of *S. persica* extract using qNMR for further inclusion in nutraceuticals.

1. Introduction

Miswak tree (*Salvadora persica* L., family Salvadoraceae) is an ever green tree that grows in different regions in Africa and Asia. For centuries, roots and stems of *S. persica* have been used as a natural toothbrush in many cultures where it is known as Arak, Kharijal, Miswak or Siwak among other names [1,2]. Additionally, leaves, fruits and seed extracts of *S. persica* have been prescribed for the treatment of fever, cold, malaria, constipation and rheumatoid arthritis [2].

Among its potential biological actions, the antimicrobial effect of *S. persica* has received the most attention. Interestingly, the extent and spectrum of *S. persica* antimicrobial activity varied widely depending on the organ used, method of extraction or geographical origin of the plant used [3]. Further investigation by Sofrata et al. identified benzyl isothiocyanate (BITC), a constituent found in *S. persica* roots and twigs, as a potent bactericidal agent especially against Gram negative microbes i.e., *Aggregatibacter actinomycetemcomitans* and *Porphyromonas gingivalis* that are involved in the etiology of different gum diseases [4]. Other secondary metabolites identified in *S. persica* include 1,4-butanediamides,

* Corresponding author at: Cairo University, College of Pharmacy, Department of Pharmacognosy, Egypt.
E-mail address: mohamed.farag@pharma.cu.edu.eg (M.A. Farag).

https://doi.org/10.1016/j.jpba.2020.113727
0731-7085/

lignin glycosides, benzylurea derivatives, alkaloids, and volatile components [2].

Metabolomics, large scale analysis of metabolites in a living organism employs different chemical analyses including chromatographic separation of metabolites coupled to mass spectrometry (MS) or direct fingerprinting using nuclear magnetic resonance NMR to analyze complex mixtures of extracted metabolites [5]. Since its emergence at the beginning of the century, the use of plant based metabolomics has continuously evolved to include applications in different fields such as quality control of herbal drugs, dereplication of bioactive compounds in herbal extracts, optimization of nutraceuticals and investigation of the effect of herbal products on the human metabolome [5]. Carefully designed metabolomics studies are currently used for identifying chemicals mediating for certain biological activities of plant extract such as detection of the anti-diabetic alkaloid trigonelline in Egyptian balanites [6].

MS based metabolomics is commonly used to provide an extensive chemical profile of secondary metabolites present in a plant extract, even those present at low concentrations [7], while NMR based metabolomics, despite being less sensitive is employed to provide a more quantifiable picture of plant major metabolites. We have previously reported on GC/MS for the aroma profiling of *S. persica* in context of its different origin and organ type [8]. As GC/MS analysis is limited to volatile constituents, herein we extend our investigation to encompass two more comprehensive platforms (UPLC-MS and NMR) to scrutinize the secondary metabolome of *S. persica* in its methanolic extract. Methanol was chosen for extraction due to its superior ability to extract a large number of secondary metabolites with a wide range of polarity. The major goal of the current study was to investigate the bioactive metabolites of *S. persica* accessions in context of its different plant organs and localities so as to set a framework for its quality control. Additionally, we attempted to rationalize for the traditional use of *S. persica* as a natural toothbrush *via* detailed metabolites profiling. Owing to the complexity of acquired data, multivariate analyses *e.g.,* hierarchical cluster analysis (HCA), principal component analysis (PCA) and orthogonal projections to latent structures (OPLS) were performed to ensure good analytical reliability and define both similarities and differences among accessions. Our results unraveled a wide array of biologically active metabolites likely to contribute to the use of *S. persica* as a natural toothbrush, and of potential to be incorporated in dental pastes.

2. Material and methods

2.1. Plant material

Four well-characterized *S. persica* L. samples representing different organs and geographical origin were selected. SP1 stem sample was obtained from cultivated *S. persica* tree in Jeddah area, Saudi Arabia, while SP2 and SP3 were two root samples obtained from different regions of Saudi Arabia. Finally, SP4 represented a commercial root sample obtained from Aswan, Egypt.

2.2. Chemicals and reagents

Methanol-d4 (99.80 % D) and hexamethyldisiloxane (HMDS) were purchased from Deutero GmbH (Kastellaun, Germany). MilliQ water was used for LC analysis. Acetonitrile (HPLC grade) and formic acid (LC–MS grade) were obtained from J. T.Baker (Deventer, The Netherlands). All other chemicals were provided by Sigma-Aldrich (St. Louis, MO, USA).

2.3. Extraction and sample preparation

Freeze-dried *S. persica* root or stem, (n = 3 each) were ground in liquid nitrogen using pestle and mortar. Extraction procedure was conducted as described previously [9]. The powder (150 mg) was mixed with 6 mL methanol containing 1 μg/mL umbelliferone (an internal standard, Sigma-Aldrich, St. Louis, MO, US) and homogenized with an Ultra- Turrax (IKA, Staufen, Germany) at 11,000 rpm, 5 × 60 s with 1 min break intervals. Extracts were vortexed (1 min), centrifuged at 3000 g (30 min), and filtered through a 22 μm pore size filter.

2.4. High resolution UPLC–MS analyses

For chromatographic separation of metabolites, an Acquity UPLC system (Waters corporation, Milford, MA, US) equipped with a HSS T3 column (100 × 1.0 mm, 1.8 μm; Waters) was used. A binary gradient of water/ formic acid, 99.9/0.1 [v/v] (A) and acetonitrile/formic acid 99.9/0.1 [v/v] (B) was used for optimal separation of eluted peaks at a flow rate of 150 μL min^{-1} as follows: 0–1 min, isocratic 95 % A, 5 % B; 1–16 min, linear from5 % to 95 % B; 16–18 min, isocratic 95 % B; 18–20 min, isocratic 5 % B. The injection volume was 3.1 μL [9]. High resolution mass of eluted compounds were detected using a Micro- TOF-Q quadrupole time-of-flight mass spectrometer (Bruker Daltonics Inc., Billerica, MA, US) in both positive and negative ion modes using Apollo II electrospray ion source under the following conditions: nebulizer gas, nitrogen, 1.6 bar; dry gas, nitrogen, 6 L min^{-1}, 190 °C; capillary, −5500 V (+4000 V); end plate offset, −500 V; funnel 1 RF, 200 Vpp; funnel 2 RF, 200 Vpp; hexapole RF, 100 Vpp; quadrupole ion energy, 5 eV; collision gas, argon; collision energy, 10 eV; collision RF200/400 Vpp (timing 50/50); transfer time, 70 l s; prepulse storage, 5 l s; pulser frequency, 10 kHz; spectra rate, 3 Hz. Internal mass calibration of each analysis was performed by infusion of 20 μL 10 mM lithium formate in isopropanol/water, 1/1 (v/v), at a gradient time of 18 min using a diverter valve.

2.5. Identification of metabolites via UPLC–MS

Metabolites were characterized by their UV–Vis spectra (220–600 nm), retention times relative to external standards, mass spectra and comparison to our in-house database, phytochemical dictionary of natural products database and reference literature.

2.6. MS data processing for multivariate analysis

Relative comparison of *S. persica* metabolites profiles after UHPLC–MS was performed using XCMS data analysis software, which can be downloaded for free as an R package from the Metlin Metabolite Database (https://www.bioconductor.org/packages/release/bioc/html/xcms.html). This software approach employs peak alignment, matching and comparison, as described previously [9].

2.7. NMR acquisition

Agilent VNMRS 600 NMR spectrometer (Agilent Technologies, Palo Alto, CA, US) was used for recording all spectra at proton NMR frequency of 599.83 MHz using a 5-mm inverse detection cryoprobe. ^{1}H-NMR spectra were recorded digital resolution: 0.367 Hz/point; pulse width (pw): 3 μs (45°); relaxation delay, 23.7 s; acquisition time, 2.7 s; and number of transients, 160. Zero filling up to 128 K (lb = 0.4) was used prior to Fourier transformation. 2D NMR spectra were recorded using standard CHEMPACK 4.1 pulse sequences (gDQCOSY, gHSQCAD, gHMBCAD) implemented in Varian VNMRJ 2.2C spectrometer software. HSQC experiment

was optimized for ^{1}JCH = 146 Hz with DEPT-like editing and ^{13}C-decoupling during acquisition time. HMBC experiments were optimized for a long-range coupling of 8 Hz; a two-step ^{1}JCH filter was used (130–165 Hz).

2.8. NMR data processing for multivariate analysis

The ^{1}H-NMR spectra were automatically Fourier transformed to ESP files using ACD/NMR Manager lab version 10.0 software (Advanced Chemistry Development Inc., Toronto, Canada). The spectra Hexamethyl disiloxane (HMDS) was used as a reference internal standard set at 0.062 ppm for ^{1}H-NMR, while the solvent carbon signal for CD_3OD at 49 ppm was used as a reference for ^{13}C-NMR, respectively. For multivariate analysis, ^{1}H-NMR spectra were divided into buckets of equal width (0.04 ppm) within the region of δ 11.4 to −0.4 ppm. Spectral intensities after exclusion of the regions between δ 5.0–4.7 and δ 3.4–3.25 corresponding to residual water and methanol signals, respectively, were used for multivariate analyses. PCA was performed with R package (2.9.2) using custom-written procedures after referencing to HMDS signal and exclusion of solvent regions.

2.9. NMR quantification

For the quantification of metabolites, HMDS was added to a final concentration of 0.94 mM. Peak area of selected proton signals belonging to the target compounds and the peak area of the IS (HMDS) were integrated manually for all the samples. Concentration for each metabolite was calculated as mg/g of dry powder using the method proposed by Farag et al. [9]. Values were expressed as mean ± S.D (n = 3).

2.10. PCA and OPLS analysis of UPLC-MS dataset

PCA and HCA were performed on the MS-data using custom script under the R 2.9.2 environment. SIMCA-P Version 13.0 (Umetrics, Umeå, Sweden) was used for partial least squares-discriminant analysis (OPLS-DA). Analysis of the resulting S-plot was used to identify discriminating markers with covariance (p) and correlation (pcor). All variables were mean centered and scaled to Pareto variance. To prevent OPSL models over fit, *p* value was calculated for all models at less than 0.05.

3. Results and discussion

3.1. NMR analysis of S. persica metabolome

^{1}H-NMR was utilized to characterize and quantify major peaks in *S. persica* extract in an effort to identify major constituents to contribute for its biological effects. The ^{1}H-NMR spectrum of *S. persica* extract was characterized by overcrowded signals in the mid-spectrum region (δ_H 3–5) which represented mainly sugars, alcohols and amino acids, Fig. 1. In contrast, few NMR signals with much lower intensities were observed in the downfield region, generally characteristic for phenolic/aromatic compounds (δ_H 5–8), while the upfield region characteristic for sterols/terpenoids (δ_H 0–3) showed major variations among different samples, most notably in the spectrum of stem samples (SP1) where the most upfield region showed considerably high intensity signals of fatty acids, Fig. 1A. To assign ^{1}H-NMR signals, 2D experiments such as 2D^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HMBC and ^{1}H-^{13}C-HSQC were used to assign signals in ^{1}H-NMR spectrum to their corresponding compounds.

3.1.1. Primary metabolites

Some signals characteristic for common plant sugars could be readily assigned in the ^{1}H-NMR spectrum such as those for the anomeric protons of sucrose at δ_H 5.39 (d, *J* = 3.8 Hz), α-glucose at δ_H 5.10 (d, *J* = 3.7 Hz), *β*-glucose δ_H 4.47 (d, *J* = 7.8 Hz) and the methyl group signal characteristic for rhamnose δ_H 1.12 (3H,d, *J* = *6.4* Hz) in accordance with their spectroscopic data published in many studies [9,10]. The identity of the aforementioned sugars was further confirmed by examining their 2D^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HMBC and ^{1}H-^{13}C-HSQC correlations (Table 1).

The presence of fatty acids was inferred from their characteristic broad peak at δ_H 1.26–1.30 which appeared coupled in the ^{1}H-^{13}C-HSQC spectrum to methylene carbon at δ_C 30.4 representing the repeated methylenes of the aliphatic chain [9], Fig. S1. Additionally, unsaturation of some fatty acid chains was confirmed by the presence of a triplet at $\delta_{H/C}$ (1H, 5.34/130, *t*, *J* =6.4 Hz) which could be correlated to the adjacent methylene at $\delta_{H/C}$ 2.04/28.2 and the aliphatic methylenes at δ_C 30.4 in the HMBC spectrum, Fig. S2. Additionally, ^{1}H-^{13}C-HSQC and HMBC spectra revealed that the majority of aliphatic chain carbons were parts of amide derivatives rather than free acids. While the free fatty acid is characterized by its carbonyl at δ_C 178 and the adjacent α-methylene at $\delta_{H/C}$ 2.27/35.22, amide derivatives could be identified by the upfield shift of the amide carbon at δ_C 174.8 and downfield shift of the adjacent methylene proton at δ_H 2.33 (Table 2, Figs. S1 & S2). Other identified primary metabolites in the ^{1}H-NMR spectrum are shown in Table 1 which include carboxylic acids i.e., acetic, lactic and succinic acid, and amino acids i.e., alanine in addition to β-sitosterol. 2D spectra showing the diagnostic signals for these metabolites are shown in Figs. S3–S5.

3.1.2. Proline-betaines

^{1}H-NMR peaks with the highest integration values were found as a large cluster between 3.00–3.50 ppm and could not be assigned in the 1D spectrum due to its overlapping with solvent peak of CD_3OD at δ_H 3.32. In the ^{1}H-^{13}C-HSQC spectrum, five prominent cross peaks were observed in this region at $\delta_{H/C}$: 3.161/46.3, 3.166/49.1, 3.338/52.8, 3.352/49.7 and 3.411/54.87, Fig. 2. The chemical shift in both proton and carbon dimensions suggested that these peaks corresponded to methyl groups attached to heteroatoms (O or N). Upon examining ^{1}H-^{13}C-HMBC spectra, it was clear that within this cluster two pairs of methyl groups can be identified that is the first at δ_H 3.166 and 3.411 representing compound **1** while the two other signals observed at δ_H 3.161 and 3.338 ppm were annotated to compound **2**, Fig. 2A. Within each pair, the two methyl groups shared the same HMBC correlations to two carbon atoms in the ppm range of δ_C 68–78, Fig. 2B, confirming their attachment to a heteroatom. Interestingly, as these methyl groups are attached to a heteroatom, they displayed strong HMBC correlations (2–3 bonds) to one another, Fig. 2B, indicating that the only possible arrangement for each pair is to be attached to a quaternary nitrogen. Furthermore, the lack of chemical and magnetic equivalence for these germinal methyl groups suggested the diastereotopic nature of these methyls due to their presence next to a stereogenic carbon. Additional HMBC correlations from these methyl pairs were observed to a methine carbon (CH) at $\delta_{H/C}$ 4.313/77.8 for compound **1** and $\delta_{H/C}$: 4.055/77.5 for compound **2** which in turn displayed HMBC correlation to a carboxylic group at δ_C 170.8 for compound **1** and δ_C 171 for compound **2,** respectively, Figs. S6 & S8. These key cross peak correlations annotated compounds **1** and **2** as α-aminoacids of proline nucleus. By further examination of ^{1}H-^{13}C-HMBC and ^{1}H-^{1}H-COSY correlations, compound **1** was identified as 4-hydroxyproline-betaine (4-hydroxystachydrine), Figs. 2, S6 & S7, while compound **2** was identified as proline-betaine (Stachydrine), Figs. 2, Fig. S8 & S9 [11]. Both compounds are first time to be reported in *Salvadora spp*,

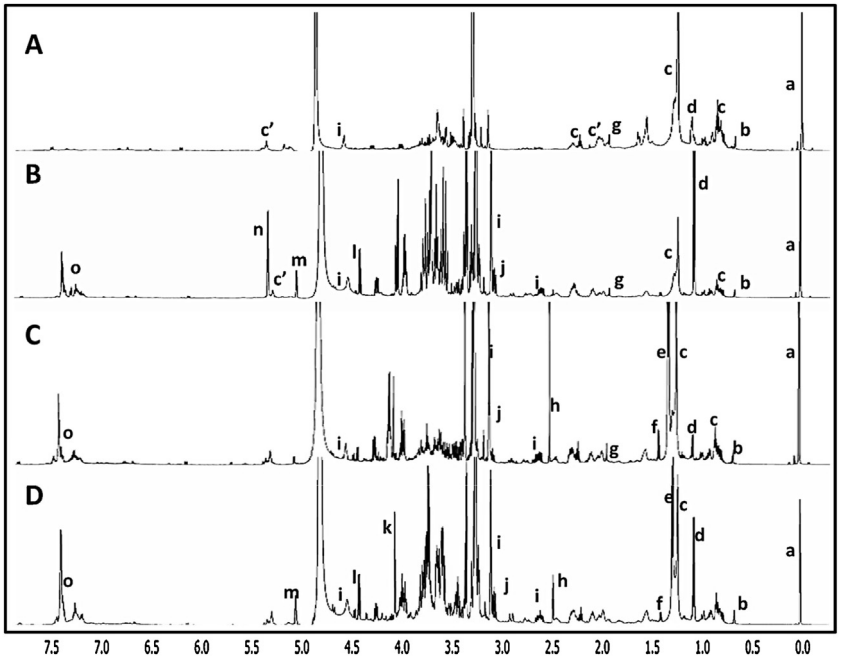

Fig. 1. ^{1}H-NMR spectra of representative samples of 4 *S. persica.*
^{1}H-NMR spectrum in the region δ_H -0.5–8.0 of four biological specimens A: SP1, stem sample, B:SP2, root sample from S.A., C: SP3: root sample from S.A., and SP4: root sample from Egypt.
Lower case letters indicate signals assigned for different metabolites, a:HMDS, b:*β*-sitosterol, c: fatty acids, c': unsaturated fatty acids d: rhamnose, e: lactic acid, f: alanine, g: acetic acid, h: succinic acid, i: 4-hydroxystachydrine, j: stachydrine, k: benzylamine, l: *β*-glucose, m: α-glucose, n: sucrose, o: collective aromatic signals of benzyl derivatives.

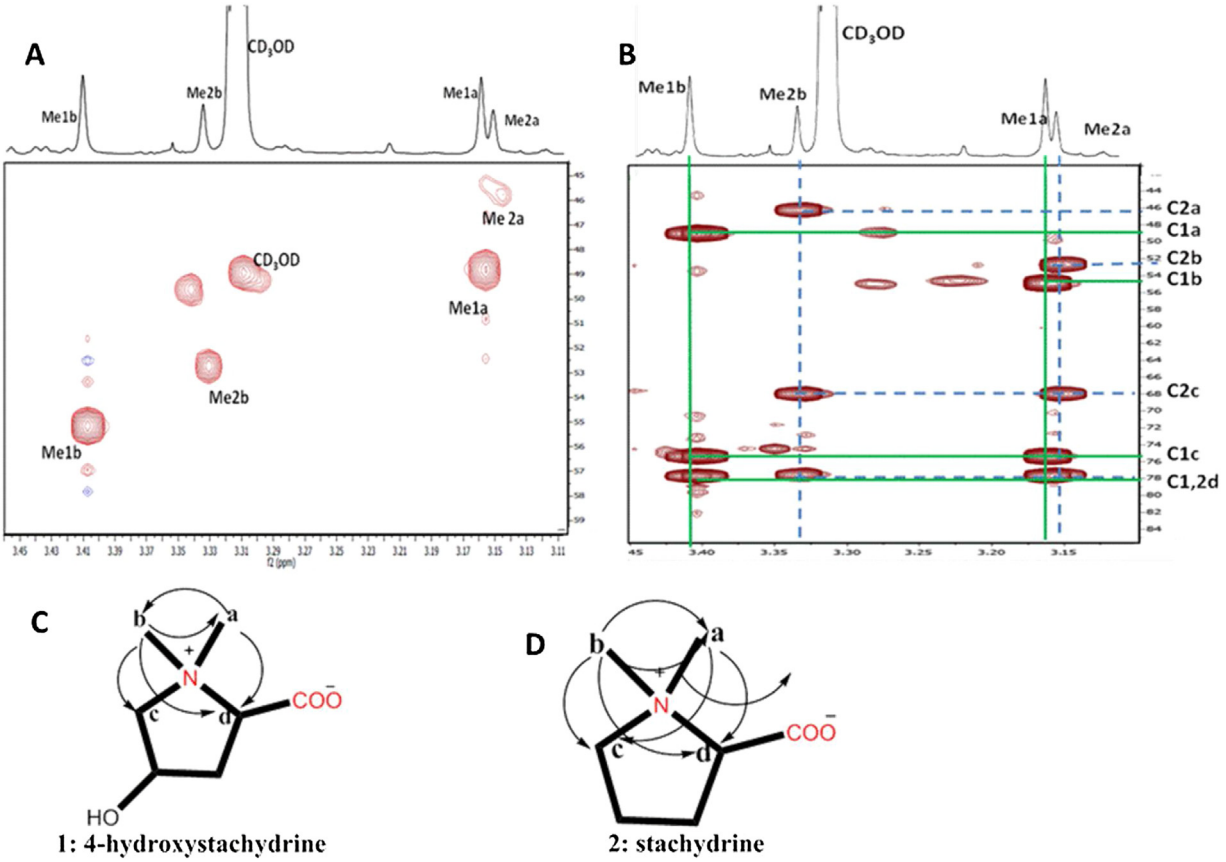

Fig. 2. Identification of proline-betaines in *S. persica* extract.
A: ^{1}H-^{13}C-HSQC spectrum from the region (δ_H 3.10–3.46, δ_C 44–60) showing characteristic methyl signals for two proline-betaines **1** and **2**, B: ^{1}H-^{13}C-HMBC spectrum from the region (δ_H3.10–3.46, δ_C 42–85) showing HMBC correlations from the methyl group protons to adjacent carbons of compound **1** (green solid lines) and **2** (blue dashed lines), C and D: Chemical structures of compound **1**, 4-hydroxystachydrine, and **2**: stachydrine with HMBC correlations of the methyl groups indicated by arrows.

Table 1
NMR Assignment of major metabolites identified in *Salvadora persica* extract.

Metabolite	Position	δ_H ppm	Multiplicity, J(Hz)	HSQC	HMBC
Acetic acid	C2	1.98	*s*	–	**C1:**175.7
Succinic acid	C2,3	2.55	*s*	30.5	**C1,C4**:177
Lactic acid	C2	4.16	*q, 6.9*	68.5	**C1**:180, **C3:**21.8
	C3	1.37	*d, 6.9*	21.8	**C1:**180, **C2:**68.5
	C2	4.313	*dd,10.3,7.2*	77.8	**C1**:170.8,**C6**:49.1, **C6'**:54.9**, C3**:38.1
4-Hydroxy proline-betaine	C3	2.658	*ddd,14.9,10.3,7.2*	38.1	**C1**:170.8, **C2**:77.8, **C3:**75.1,
4-Hydroxy stachydrine		2.349			
	C4	4.594	*br*	67.6	**C2:**77.8
	C5 C7	4.029	*overlap*	75.1	**C4**:67.6**, C3:** 38.1
		3.411	*s*	54.9	**C2**:77.8, **C5**:75.1, **C6':** 49.1
	C8	3.166	*s*	49.1	**C2:**77.8,**C5:** 75.1, **C6:**54.9
Proline-betaine	C2 C3 C4 C5 C7 C8	4.055	*overlap*	77.5	**C1:**171, **C3:**26.8
Stachydrine		2.322	*m*	26.8	**C5:**68.2**, C4:**19.9
		2.504	*m*	19.9	**C3:**26.7
		2.147	*m*	68.2	**C4:**19.9
		3.716	*overlap*	46.3	**C2**:77.5**, C5**:68.2, **C6'**:52.8
		3.535	*dd, 3, 9.3*	52.8	**C2:**77.5**, C5:**68.2, **C6:**46.3,
		3.161	*s*		
		3.338	*s*		
Fatty acids	t-CH_3	0.89	*t,7.1*	14.4	30.2, 21.4
	$(CH_2)_n$-	1.26-1.3	*m*	30.2	30.2, 21.4
	CH = CH**CH2**-CH=	5.35	*t,6*	130.7	28.2
		2.04	*m*	28.2	130.7, 30.2
	C2	2.27	*t*	35.2	**C1:**178, 30.2, 26.1
Fatty amide	$(CH_2)_n$- C2	1.26-1.3	*m*	30.2	30.2, 21.4
		2.33	*m*	34.9	**C1:**174.8, 30.2
Rhamnose	C6	1.13	*d,6.4*	19.7	**C5:**69.7, **C4**:72.2
Sucrose	C1	5.4	*d,3.8*	93.6	**C**2':105.1,**C2:** 74.5
	C5'	3.77	*m*	83.6	**C**6':64.9
	C1'	3.6	*m*	64.1	**C**2':105.4, **C3':**79.4
	C3'	4.1	*m*	79.4	**C**4':76.2**, C1'**:64.1
α-Glucose	C1	5.12	*d,3.6*	93.8	**C2:**75.25**, C3:**73.4
β-Glucose	C1	4.48	*d, 7.7*	98.2	**C2:**78.5
Alanine	C3	1.47	*d, 7.2*	17.3	**C1**:175.2**, C2:**51.7
Benzyl urea	C1'	4.26	*s*	52.4	**C1**: 161.4**, C2':** 137.8, **C3',7'**:131,
	C1	4.73	*d,*	71.7	**C2**:139.2,**C3,7**:130.4, **C1'**:101.2
Salvadoside		4.91	*overlap*		
	C1'	4.5	*d, 7.8*	101.2	**C1**: 71.7
	C3,7	7.45	*overlap*	130.4	**C3,5**:129.3**, C7:**71.4,
Benzyl amine	C1	4.12	*s*	44.2	**C2**: 135**, C3,7:** 130.7,
	C3,7	7.48	*overlap*	130.7	**C1:**52.7
BITC	C1'	4.84	*overlap*	49.9	**C1**':137.7**, C1**:129.2,
β-Sitosterol	C18	0.72	*s*	12.1	**C14**:58**, C13:**43.6, **C12**:40.9
	C26/27	0.84	*d,6.5*	19.9	**C24**:51.6, **C25**:33.6, **C26/27**:19.9

Table 2
Secondary metabolites identified in *S. persica* and their potential biological activity relevant to periodontal health.

Metabolite	Detection LC/MS	Detection NMR	Activity	Reference
Stachydrine[a]	–	++	Anti inflammatoryAntioxidant	[28]
4-Hydroxy stachydrine[a]	+	++	Anti inflammatory	[29]
Salvadosides	++	++	Not known	
N-Benzyl-2-phenylacetamide	+	–	Antimicrobial	[13]
BITC	–	+	AntibacterialAntiviral (HSV-1)	[4] [15]
Ellagic acid[a]	+	–	Antibacterial	[18]
Sulfated flavonoids[a]	+	–	Antiviral Anti inflammatory	[19] [20]
Persicaline	+	–	Antioxidant	[16]
β-Sitosterol	+	+	Anti inflammatory	[30]

[a] Identified in *S. persica* for the first time, +: present, ++, present in high concentration.

albeit previously reported in many plants including citrus fruits and *Leonurus* genus [12]. The presence of additional betaines at lower concentration may be possible due to the existence of minor cross peaks within the same ppm range of compounds **1** and **2**.

3.1.3. Benzyl derivatives

Due to the quantitative nature and the relative low sensitivity of NMR, only plant metabolites with the highest abundance are detected. By examining ^{1}H-NMR spectrum of *S. persica* extract, signals indicative for phenolic compounds such flavonoids, coumarins and phenolic acids were detected only at very low intensity, thus their presence as major secondary metabolites in *S. persica* was excluded, Fig. 1. This narrow ppm range of the aromatic signals and the fact that they are heavily overlapping suggested the abundance of benzyl derivatives rather than phenolics.

The benzylic region in the ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C HMBC spectra showed four main benzyl derivatives that were distinguished based on their chemical shifts and HMBC correlations of their corresponding benzylic CH_2 carbon, Fig. S10. The first benzyl derivatives were annotated as benzylureas/benzylamide derivatives with the benzylic CH_2 observed spanning a narrow δ_H range of 4.22–4.30 and δ_C 52, Fig. S10. Among these signal which all showed HMBC cor-

relation to an aromatic carbon at 131 ppm, at least one (δ_H 4.25, 2H, s) showed a clear cross peak at the downfield carbon at a δ_C:161.4, characteristic for benzylurea Fig. S11 [13]. Further distinction of these benzyl derivatives based on the substitution of the aliphatic chain or the aromatic ring was not possible as these substitutions are more than 3 bonds far beyond the limit of detection in HMBC experiment.

The downfield chemical shift of another benzylic CH_2 carbon at $\delta_{H/C}$ 4.73,4.91/71.7 was found to be consistent with that of benzylglycosides including salvadoside, 1-*O*-benzyl-*β*-*D*-glucopyranoside-2-sulfate, previously reported in *S. persica* [14]. The presence of the glycosidic linkage was also confirmed by HMBC correlations of the benzylic protons to the anomeric carbon of *β*-glucose at δc 101.2, Fig. S12. Similarly, *β*-glucose anomeric proton at δ_H 4.5 showed HMBC correlations to the benzylic carbon at δ_c 71.7 confirming the assignment of salvadoside signals.

Additional benzylic compounds detected in the ^{1}H-NMR spectrum included benzylamines which has been previously reported in *S. persica* via GC/MS [8]. Benzylamine was identified in the ^{1}H-^{13}C-HSQC from its key benzylic carbon δ_H 4.12 (2H, s), δ_C 44.2, Fig. S10. These benzylic protons were assigned to benzylamine based on literature and due to their strong HMBC correlations to two aromatic carbons at δ_C 110.7 and 135, while showing no additional correlations despite its relatively good signal intensity, Fig. S13. Finally, the characteristic signal for the benzylic carbon for benzylisothiocyanate (BITC) could not be unequivocally assigned in the ^{1}H-NMR spectrum due to its position at δ_H 4.84 underneath the solvent (water) signal, Fig. S10. However, presence of BITC was confirmed from the ^{1}H-^{13}C-HSQC spectrum by a methylene cross peak at $\delta_{H/C}$ 4.84/49.9 and HMBC correlations observed for these protons to two carbons at 128.9 and 136.8 in agreement with previous reports [15,16], Table 2, Fig. S14.

3.2. Quantitative NMR

Due to the inherent unbiased nature of NMR, quantification of major metabolites could be obtained as long as they display at least one well resolved discriminatory signal. In that regard, relative quantitation of several metabolites in *S. persica* extract was possible relative to the internal standard HMDS. Stachydrine level was quantified from its characteristic methyl group (3 H) at δ_H 3.353, whereas 4-hydroxystachydrine was quantified based on its resolved signal (1 H) at δ_H 4.596.

For benzylic derivatives, NMR could only discriminate based on the general type of benzylic derivatives as further substitution at the conjugated aliphatic moiety may not affect the chemical shift of the benzylic protons to considerable degree. Accordingly, the molecular weights for *N,N'* dibenzylurea and salvadoside were used to calculate benzylureas and benzylglycosides amounts, respectively.

For sugars, integration of the well resolved anomeric proton signal was used to estimate their concentration, while lactic acid, acetic acid and alanine were quantified from their well resolved methyl groups (3 H) at δH 1.37, 1.98 and 1.47, respectively, whereas succinic acid was quantified through integration of its α-CH2 (4H, 2.55 ppm).

Generally, samples derived from stem had much lower secondary bioactive metabolites level than root samples suggesting its inferior quality as toothbrush or chewing stick, Table S1. NMR quantification revealed that proline-betaines appeared to be the most abundant group of secondary metabolites in all samples. 4-Hydroxy-stachydrine was the major secondary metabolite detected at highest levels in root samples (7.3–10.4 μg/mg) versus lowest in stem (3.76 μg/mg), Table S1. Salvadoside was the second most abundant secondary metabolite in both root (6.48–8.38 μg/mg) and stem (2.84 μg/mg). Larger variation was observed in primary metabolites i.e., sugars and organic acids exemplified in sucrose which displayed wide concentration range from 4.07–26.45 μg/mg. Interestingly, higher benzylamine levels were always detected along high lactic acid levels and vice versa, Table S1. Therefore, the detrimental effect exhibited by lactic acid on teeth may be offset by the presence of other ionized molecules such as benzylamine and betaines.

3.3. Characterization of secondary metabolites in S. persica via UPLC–ESI-MS

Although NMR provided a clear distribution of major metabolite classes in *S. persica* with quantification, LC/MS was further employed for the detection of minor constituents. Methanol extract of the different accessions of *S. persica* were analyzed using a high resolution UPLC–PDA–ESI–qTOF-MS, which allowed for the detection of 73 peaks as listed in Table S2. To provide a comprehensive insight onto metabolites composition, both negative and positive ionization modes were employed. A total of 76 metabolites were detected which belonged to various chemical classes including phenolic acids, sulfur compounds, flavonoids, alkaloids, phenolic diterpenes and fatty acids (Fig. 3). 45 compounds were detected in positive ion mode mostly belonging to nitrogenous compounds and coumarins versus 24 compounds detected in negative ion mode, mainly phenolic acids, sulfated compounds and flavonoids, Table 2.

3.3.1. Phenolic acids

MS spectral analysis revealed for the presence of ellagic acid and its derivatives as major phenolic components in *S. persica* extract (peaks 9, 11, 12, 16, 34). The MS/MS spectrum of peak 11 displayed fragmentation pattern characteristic of ellagic acid with fragment ion *m/z* 257 $[M-H-CO_2]^-$, *m/z* 241 $[M-H-CO_2 -OH]^-$, *m/z* 213$[M-H-2CO_2]^-$ and *m/z* 181$[M-H-2CO_2-2OH]^-$, along with pyran ring derivatives at *m/z* 210, 167, 151 and 139, Fig. S15. Peak 9 exhibited $[M-H]^-$ ion at *m/z* 287.0421 and yielded fragment ions typical of ellagic acid at *m/z* 257$[M-H-CO]^-$, 241 $[M-H-CO-OH]^-$, 213[M−H-CO−CO2]- and 181$[M-H-CO-CO_2-2OH]^-$ with one less oxygen annotated as deoxyellagic acid, Fig. S16. Peaks 12 and 16 exhibited [M−H]– at *m/z* 315.0832 and *m/z* 329.0712 [M−H]–, respectively indicating one and two extra methyl groups in peak 12 and 16, annotated as methoxyellagic acid and di-*O*-methylellagic acid, respectively, Figs. S17 and S18 [17]. An acylated ellagic acid with gallic acid moiety was annotated in peak 34 $[M-H]^-$ at *m/z* 453.0912 showing typical 152 amu loss of gallic acid moiety appearing at *m/z* 301 and identified as ellagic acid gallate. These ellagic acid derivatives might play a role in *S. persica* oral health owing to their potent antimicrobial effect against common oral pathogens e.g. *P. gingivalis* and *S. mutans* [18].

Other phenolic acids detected in the positive ion mode included *O*-coumaroylquinic acid (peak 31, *m/z* 339.4321, $C_{16}H_{19}O_8^+$), *O*-caffeoylshikimic acid (peak 38, *m/z* 337.1532, $C_{16}H_{17}O_8^+$) and caffeoyl-*O*-hexoside (peak 55, *m/z* 343.2511, $C_{15}H_{19}O_9^+$). These phenolic acids are reported herein for the first time in *S. persica* L.

3.3.2. Flavonoids

8 Flavonoid glycoside peaks were identified (22–25, 27, 29, 52, 56) based on their UV spectra and corresponding to sugar losses. Kaempferol aglycone (*m/z* 285.0329, $C_{15}H_9O_6$) was detected in peaks 22, 23, 24 annotated as kaempferol-*O*-dihexoside, kaempferol-O-hexosyl-pentoside and kaempferol-*O*-hexoside, respectively. Likewise, two isorhamnetin glycosides were detected in peaks 52 and 56 with $[M+H]^+$ at *m/z* 463.4321 and 479.3272 yielding fragment ion of isorhamnetin at *m/z* 317 annotated as isorhamnetin-O-deoxyhexoide and isorhamnetin-O-hexoside, respectively. Abundance of flavonols was visible from the

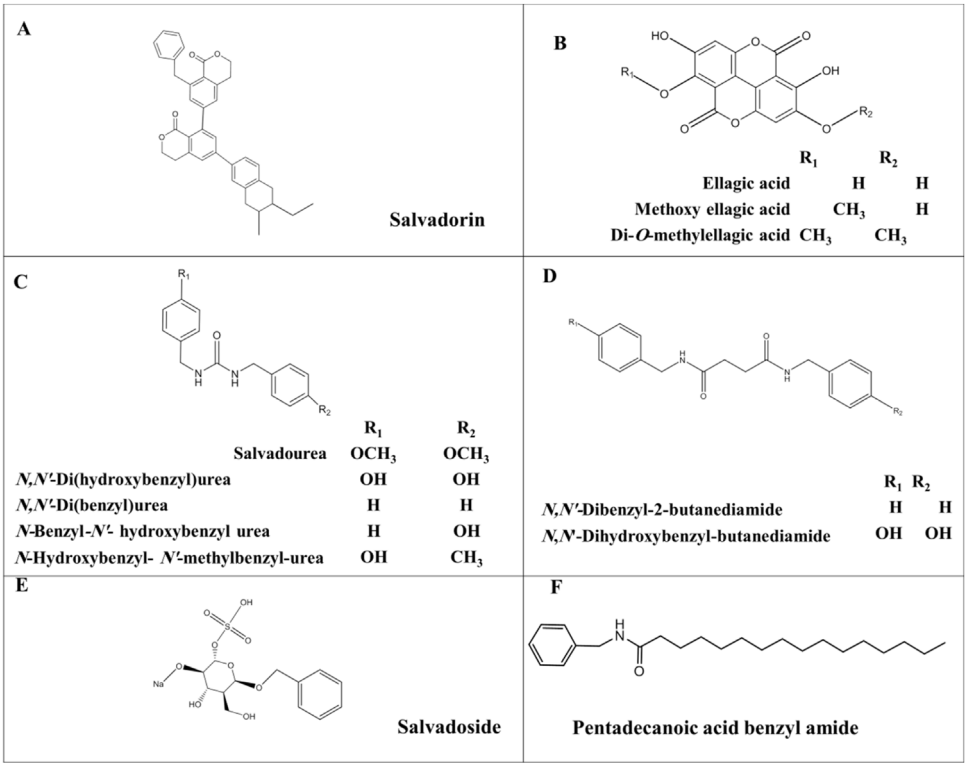

Fig. 3. Examples of natural product classes identified in *S. persica* using UPLC-MS.
Classes identified included coumarins. A: phenolic acid B: urea derivatives, C: butanediamide derivative, D: sulfur compound, E: fatty acid amide, F: with selected compound(s) discussed in the manuscript

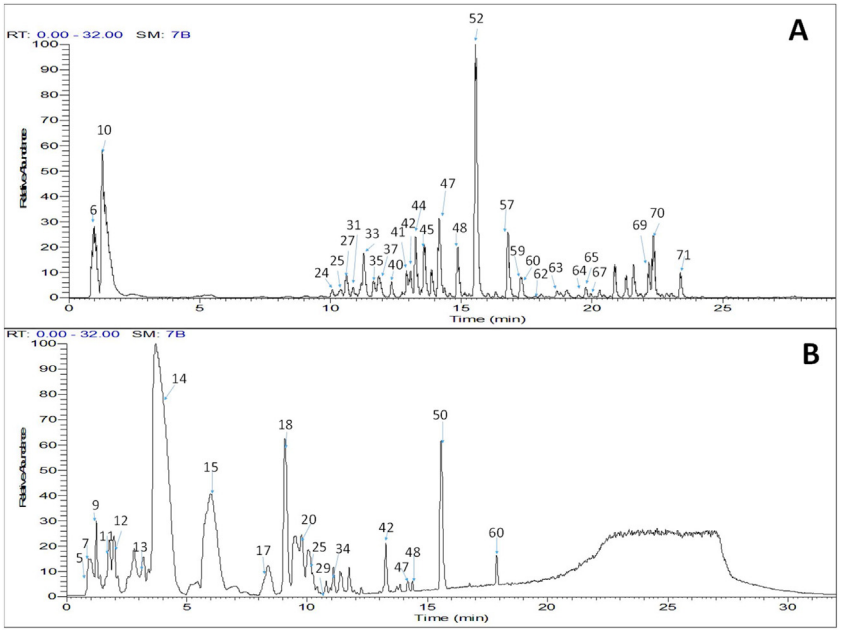

Fig. 4. Representative UPLC-ESI-MS base peak chromatograms (BPC) of the methanol extracts of *S. persica.*
BPC of *S. persica* extract A: in positive ion mode and B: negative ion modes. Chromatographic conditions are described under "material and methods" section. The identities, R_t values, and basic UV and MS data of all peaks are listed in Table 2.

chromatogram in the elution region (100–300 s), especially in *S. persica* stem extract (Fig. 4).

In addition to flavonols, apigenin glycosides were identified in peaks 25, 27 and 29, based on aglycone fragment ion at *m/z* 271 $[M+H]^+$. In the negative ionization mode, a flavonoid sulfate was identified (peak 32 $[M-H]^-$ at *m/z* 659.1822 for $C_{26}H_{27}O_{18}S^-$) yielding characteristic cleavage of sulfate group at *m/z* 579 $[M-H-80]^-$ along with further sequential loss of a hexose *m/z* 417 for $[M-H-162-80]^-$and deoxyhexose *m/z* 433 for $[M-H-146-80]^-$ and finally yielding *m/z* 271 corresponding to naringenin aglycone. Consequently, peak 32 was annotated as naringenin-O-hexosyl-deoxyhexoside sulfate. This is the first report of a sulfated flavonoid

in *S. persica* which may be of special significance due to its reported antifungal, anticoagulant, anti-inflammatory, antiviral activities [19,20].

Flavonoids and phenolic acids were not detected in the ^{1}H-NMR spectrum of the extract (Fig. 1) due to the relative low concentration of these compounds in the extract which fell below detection limits of ^{1}H-NMR. However, their higher ionizability together with high sensitivity of MS allowed for their detection using UPLC/MS (Table S2), highlighting the advantage of complementing both techniques.

3.3.3. Sulfur-containing compounds

S. persica is known to biosynthesize sulfur-containing compounds including persicaline and salvadoside which were detected in peak 21 and 50 from their predicted molecular formula [16]. The positive ion MS spectrum of persicaline (peak 22, *m/z* 400.1623, $C_{21}H_{25}N_2O_4S^+$) showed sequential loss of water molecules at *m/z* 382, 364 and 346 in addition to the characteristic ion for dibenzyl-dihydroimidazolethione at *m/z* 280, Fig S19. Salvadoside (peak 50, m/z 373.2041) exhibited fragment ions at *m/z* 293 due to the loss of sulfate moiety in addition to the loss of benzene ring appearing at *m/z* 295 and the characteristic peaks at *m/z* 91 and 108 corresponding to benzyl and benzyl alcohol moieties, respectively, Fig. S20. Other salvadoside derivatives were identified in peaks 13 and 14 for the first time in *S. persica* at *m/z* 511.1131 for $C_{19}H_{27}O_{14}S^-$ and *m/z* 349.0812 for $C_{13}H_{17}O_9S$- annotated as *O*-benzyl dihexosylsulfate and *O*-benzyl hexosylsulfate conjugates, Figs. S21 and S22.

Butanediamides

Three butanediamide derivatives were identified in the positive ion mode in peaks 17, 53 and 59 as evident from characteristic fragmentation of the butanediamide based moiety, Figs. S23, S24. Their tandem MS showed intense product ions at *m/z* 205 (benzyl butanediamide), *m/z* 144 (dimethyl butanediamide), *m/z* 120 (methylbenzyl amine), *m/z* 117 (butanediamide), *m/z* 108 (benzylamine), *m/z* 91 (benzyl moiety). By comparison with the literature, peak 17, *m/z* 223.9912 $[M+H]^+$, peak 53, *m/z* 329.1321$[M+H]^+$ and peak 59, *m/z* 297.1033$[M+H]^+$ were annotated as *N*-hydroxybenzyl-butanediamide, *N,N'*-dihydroxybenzyl-butanediamide and *N, N'*-dibenzyl-butanediamide, respectively [13].

3.3.4. Urea compounds

Four urea derivatives were detected in *S. persica* for the first time along with salvadourea (bis-(methoxybenzyl)-urea) [21]. Urea derivatives were detected in peak 33 (*m/z*: 301.1321$^+$, salvadourea), peak 40 (*m/z* 273.1123$^+$, *N,N'*-di(hydroxyphenyl)-urea), peak 42 (*m/z* 271.1322, *N*-hydroxybenzyl-*N'*-methylbenzyl-urea, peak 49, (*m/z* 241.1323$^+$, *N,N'*-dibenzylurea) and peak 54, (*m/z* 257.1102$^+$, *N*-benzyl-*N'*-hydroxybenzyl-urea), Table S2. A characteristic fragmentation pattern of such metabolite class include fragment ions appearing at *m/z* 241 for dibenzylurea $(C_{15}H_{17}N_2O)^+$, *m/z* 163 for methyl benzylurea $(C_9H_{11}N_2O)^+$, *m/z* 150 for benzylurea $(C_8H_{10}N_2O)^+$, *m/z* 108 for benzylamine $(C_7H_6NH_2)^+$, *m/z* 91 for benzyl moeity $(C_7H_7)^+$, dimethyl urea at *m/z* 89 and urea at *m/z* 62 $(CO(NH_2)_2)$, Figs. S25, S26.

3.3.5. Phenolic diterpenes

MS spectral analysis also revealed for the presence of carnosol (peak 70, *m/z* 331.0443, $(C_{20}H_{27}O_4^+)$ along with carnosic acid (peak 68, *m/z* 333.1522 $(C_{20}H_{29}O_4^+)$. The respective fragment ions at *m/z* 287 and 289 $[M+H-CO_2]^+$ in the tandem MS of peak 67 and 68 accounted for the decarboxylation of the lactone and free acid, respectively. Asides, MS^2 spectra exhibited fragmentation pattern typical of phenolic diterpenes at *m/z* 241, 215, 205 and 165, Figs. S27 and S28. Both peaks 67 and 68 showed UV maxima at 248, 280 and 428 nm indicative for the *O*-benzoquinone chromophore.

Alkaloids. Alkaloids are widely distributed in *Salvadora* genus and were readily annotated in ESI-MS from their even mass weights. Hydroxystachrydine was detected in peak 3 (*m/z* 158.9821, $C_7H_{12}O_3N^-$, $[M-H]^-$), showing the loss of hydroxy and carboxy moieties (-18 and −44 Da) at *m/z* 141 and 115, respectively, in addition to ions at *m/z* 87 and 71 characteristic of hydroxylpyrrolidine and pyrrolidine moieties, respectively Fig. S29. This alkaloid was also identified in the ^{1}H-NMR as the major constituent of the extract along with its biosynthetic precursor prolinebetaine (Fig. 2). Other alkaloids detected included *N*-benzylamine; (peak 1, *m/z* 106.1670, $C_7H_8N^-$, $[M-H]^-$), *N*-benzyl-2-phenylacetamide (peak 37, *m/z* 226.3421, $C_{15}H_{16}NO^+$, $[M+H]^+$) and reticuline (peak 48, *m/z* 330.2034, $C_{19}H_{24}NO_4^+$,$[M+H]^+$), Table S2.

Coumarins. Examination of *S. persica* UPLC/MS chromatogram revealed the presence of salvadorin, a dimeric dihydroisocoumarin, (peak 60, *m/z* 555.2481$[M+H]^+$, $C_{38}H_{35}O_4^+$) yielding fragment ions at *m/z* 537, 464, 449, 435, 316 and 281 corresponding to the following losses $[M+H-H_2O]^+$, $[M+H-C_7H_7]^+$, $[M+H-C_7H_7-CH_3]^+$, $[M+H-C_7H_7-C_2H_5]^+$, $[M+H-C_7H_7-C_9H_8O_2]^+$ and $[M+H-C_7H_7-C_{14}H_{15}]^+$, respectively, Fig. S30. Identification of salvadorin was further confirmed by UV λ_{max} at 225, 240 and 315 nm suggesting a dihydroisocoumarin skeleton.

Hydroxy-methoxy-methyl-*H*-chromen-one-*C*-hexoside was identified in peak 28, *m/z* 369.1235 $[M+H]^+$ showing loss of water at *m/z* 351$[M+H-H_2O]^+$, loss of methoxy at *m/z* 321 and the characteristic typical loss of *C*-hexoside at *m/z* 261 $[M+H-90-OH]^+$, *m/z* 249 $[M+H-120]^+$and *m/z*$[M+H-120-OH]^+$, Fig. S31 [22].

Amino acid/dipeptides. Peak 20, with an $[M-H]^-$at *m/z* 365.1621 showed the loss of 162 Da (hexoside) with fragment ions at *m/z* 203 and 157 characteristic of tryptophan annotated as tryptophan-*N*-hexoside. To the best of our knowledge, this compound is identified for the first time in *S. persica*.

Fatty acid conjugates and sterols. MS spectra of the late eluting peaks revealed for several fatty acids and sterols detected mostly in positive ionization mode. These molecules included arachidic acid (peak 64; *m/z* 313.2712, $C_{20}H_{41}O_2^+$), stigmasten-diol (peak 65; *m/z* 433.1323, $C_{29}H_{53}O_2^+$), hydroxy-octadecadienoic acid (peak 66; *m/z* 297.1034, $C_{18}H_{33}O_3^+$), *β*-sitosterol (peak 71; *m/z* 415.2311, $C_{29}H_{51}O^+$), stigmasterol (peak 73;*m/z* 413.3821,$C_{29}H_{49}O^+$). Among these compounds, *β*-Sitosterol, *β*-sitosterol-*O*-hexoside, stigmasterol, palmitic acid, linoleic acid and arachidic acid were previously identified [8,14].

Two fatty acyl amides were identified from their even mass weights indicating the presence of a nitrogen atom in peaks 69 and 70 at *m/z* 346.3151 and 360.3241 $[M+H]^+$, Fig. S32. The mass difference of 14 Da (CH_2), suggested that they were homologues with different alkyl chain length. Tandem MS analysis annotated peaks 69 and 70 as pentadecanoic acid and hexadecanoic acid benzyl amides, respectively post the respective loss of benzyl amide moieties at *m/z* 212 and 226, in peak 69 and 70. Further fragment ion corresponding to *N*-benzylformamide $[M+H-C_8H_9NO]^+$ appeared at *m/z* 137 in peaks 69 and 70 after the loss of fatty acid alkyl chains. This is the first report for the presence of oxygenated fatty acids and fatty acyl amides in *Salvadora* species and suggests that UPLC-MS represents a useful platform for fatty acids profiling in *Salvadora sp*.

3.4. Unsupervised PCA and HCA analyses of S. persica L

To characterize for metabolites heterogeneity in *S. persica* in context of its different origin and organ type, multivariate data analyses exemplified in PCA and HCA were employed to classify different accessions based on their chemical profile. For data derived

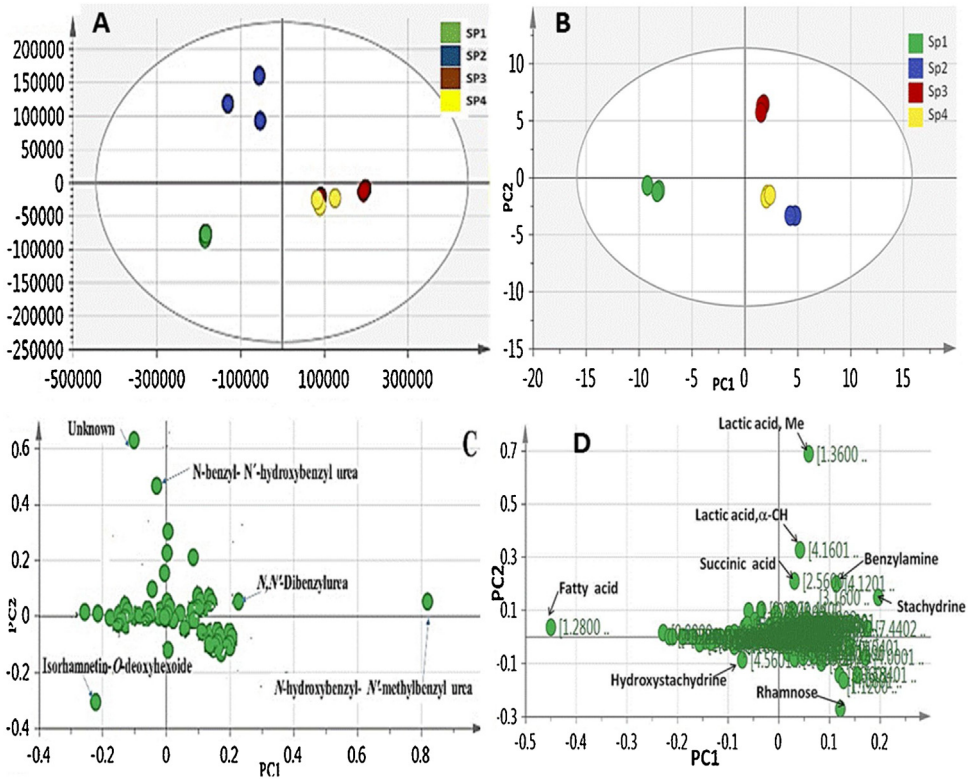

Fig. 5. Multivariate analysis of LC/MS and NMR metabolomics data sets for four *Salvadora persica* samples.
Principal component analyses of the different *S. persica* (n = 3). A, scoring plot based on UPLC–MS(*m/z* 100–1000) along two vectors of principal component 1 (PC1 = 62 %) and principal component 2 (PC2 = 18 %). B: PCA scoring plot based on ^{1}H-NMR spectra of all investigated samples along principal component 1 (PC1, 57 %) and principal component 2 (PC2, 30 %), C: the corresponding PCA LC/MS loading plot along PC1 and PC2, D: ^{1}H-NMR PCA loading plot along PC1 and PC2 axes.

from both UPLC/MS and NMR analysis, triplicate biological measurements from same sample were found to be highly reproducible as indicated by PCA scoring plots and HCA dendograms obtained from these data sets, Fig. 5, Fig. S33. Both models enabled clear segregation of stem samples from roots as indicated in PCA plots derived from either LC/MS or ^{1}H-NMR dataset with stem samples positioned in the lower left quadrant with negative PC1 and PC2 score values clearly distant from root samples, Fig. 5A & B. Likewise, HCA derived from either UPLC/MS or NMR resulted in the separation of stem samples in one cluster as the most distant branch from that of root, Fig. S33.

The PCA loading plot derived from UPLC/MS data (Fig. 5C) revealed that MS peaks assigned for isorhamnetin-*O*-deoxyhexoside contributed negatively to PC1 as found more abundant in stem specimens. In contrast, roots appeared enriched in *N*-hydroxybenzyl-*N'*-methylbenzyl urea and *N,N'*-dibenzyl urea to contribute positively to PC1. No difference in alkaloids or in butanediamides was observed among specimens as revealed from PCA loading plot derived from LC/MS data.

Although NMR segregated samples in a similar pattern to that of UPLC/MS, PCA loading plot based on NMR dataset (R^2 0.865664, Q^2 0.744264) was quite different, likely due to differential metabolites detection by each technique [9]. Stem samples were separated with negative PC1 values attributed to enrichment in aliphatic chain signal of fatty acids atδ 1.28 ppm. The loading plot also revealed that signals of betaines (δ_H 3.161) and benzylamine (δ_H 4.12) contributed positively to PC1 and were more abundant in root samples, Fig. 5D. Additionally, SP3 samples which was positioned with highest PC1 values were segregated from other samples along PC2 due to their enrichment in organic acids (i.e., lactic and succinic acid). The abundance of lactic acid was previously reported in the same specimen using GC/MS [8].

3.5. Supervised OPLS analysis

A further classification model (Fig. S34) was built using OPLS-DA by modeling root accessions versus stem based on LC/MS data. The performance of the developed model was validated from model parameters "R^2 (0.9856) and Q^2 (0.9577)" indicating model validity and good prediction power, respectively. *N*-hydroxybenzyl-*N'*-methylbenzyl urea *m/z* 271.1322 $[M+H]^+$ and *N, N'*-dibenzylurea *m/z* 241.1323 $[M+H]^+$ were more abundant in root samples, whereas stem samples showed higher levels of isorhamnetin-*O*-deoxyhexoide *m/z* 463.4321 $[M+H]^{+\cdot}$ and in agreement with PCA results (Fig. 5). Likewise, S-loading plot of supervised OPLS-DA analysis of ^{1}H-NMR data (R^2 0.977893 and Q^2 0.96875) revealed the enrichment of stem samples in aliphatic fatty acids/amides due to signals at δ_H 1.28, Fig. S35.

Proline-betaines were identified for the first time in *S. persica* extract using LC/MS and NMR metabolomics platforms. NMR quantification further revealed for their abundance in all specimens as major secondary metabolite up to 12.63 mg/g. This is the first report of proline-betaines occurrence in that species, albeit the presence of certain pyrrolidines in *S. persica* was previously detected [23]. Betaines are zwitterions derived from amino acids, most commonly glycine, widely distributed in higher plants. They function as an osmoprotectant (compatible solutes) and to be secreted as a response to stress conditions, likely typical in arid areas where *S. persica* grows [24]. In the human body, these osmoprotectants function to help preserve mucous membranes as betaines can be excreted in human tears to protect ocular mucosa [25]. More relevant, certain betaines were added to commercial toothpaste to protect against dry mouth conditions and further to protect against the detergent effects of toothpaste surfactant in clinical trials [26]. Most interestingly, a mixture of a prototype betaine and the sugar

alcohol erythritol, the later previously identified in miswak from our first analysis using GC/MS, was found to cause rapid detachment of *Streptococcus mutans* biofilm involved in the formation of dental plaque [8,27]. Such multiplex approach for metabolites detection using NMR now provides more exact roles of these meabolites, i.e. betaines in *S. persica* action mechanism.

The major betaines detected in *S. persica* were stachydrine and 4-hydroxystachydrine which function mainly as osmoprotectants. However, previous studies revealed that 4-hydroxystachydrine displayed anti-inflammatory activity while stachydrine exhibited antioxidant, anti-tumor and anti-inflammatory effect *via* the release of nitric oxide and inhibition of NF-κB [28]. These major betaines in *S. persica* reported biological effects validate its use for dental care and supports more in depth investigation of its exact biological effect regarding periodontal health.

The presence of several other secondary metabolites more enriched in root samples justify further for its traditional preference over stems as natural toothbrush. Benzylureas, *N,N* benzyl butanediamide, alkaloids and amides were detected using both LC/MS and NMR. Some of these nitrogenous metabolites have well documented biological activities that augment the benefits of *S. persica* for dental hygiene including antimicrobial, antioxidant and anti-inflammatory effect as summarized in Table 2. However, many of the compounds identified herein have not been investigated for their biological activities in pure form yet.

For further quality control purposes and standardization, simple rapid NMR metabolomics platform could be employed to determine *S. persica* drug quality based on relative intensities displayed by ^{1}H-NMR peaks in the three different spectral regions *viz* fatty acids (δ_H 1–2), betaines (δ_H 3–4) and benzyl derivatives (δ_H 7.2–7.5), as revealed in Table 1. Based on the classification model results (Fig. 5), a good quality root specimen should have highest intensities between δ_H 3–4 and δ_H 7.2–7.5 while showing less intensity signals between δ_H 1-2. Such approach can indeed be used for the screening or standardization of *S. persica* when present in dental preparations and or to assess other factors on its metabolites composition i.e., processing or storage. Screening of other *Salvadora* genotypes representing different species within that genus could also help identify other, less explored drugs.

4. Conclusion

For centuries, *S. persica* twigs and roots have been used in different cultures as a tooth brushing stick (miswak) for dental care [2]. The use of *S. persica* in dental care was traditionally justified by the mechanical friction due to its fibers in a way similar to the modern toothbrush and its ability to eliminate bad odors from mouth [8]. In the present study, two powerful metabolomics platforms were employed to develop a comprehensive profiling of the *S. persica* metabolome to justify for its folk use in oral care. Our results suggest multiple benefits from the use of *S. persica* including antimicrobial, anti-inflammatory and hydrating activities due to the abundance of several metabolites such as ellagic acid derivatives, flavonoids, proline-betaines, benzylisothiocyanate and urea derivatives. This study also presents a robust NMR platform that can be used for quality control and standardization of *S. persica* and its pharmaceutical products.

Authors statement

Mohamed Ali Farag, designed experiments, collected samples, run the models and revised the paper

Zeinab T. Shakour, interpreted and identified LCMS results and co-wrote the paper

Engy Mahrous, interpreted and identified NMR results and co-wrote the paper

Andrea Porzel, revised NMR assignment and the paper

Tilo Lübken, revised NMR assignment

Andrej Frolov, analysed LCMS for samples

Ludger Wessjohann, conceptualization, and project management, paper revision

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jpba.2020.113727.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] A.H. Sofrata, R.L.K. Claesson, P.K. Lingström, A.K. Gustafsson, Strong Antibacterial effect of miswak against oral microorganisms associated with periodontitis and caries, J. Periodontol. 79 (2008) 1474–1479, http://dx.doi.org/10.1902/jop.2008.070506.
[2] M.Z. Aumeeruddy, G. Zengin, M.F. Mahomoodally, A review of the traditional and modern uses of *Salvadora persica* L. (Miswak): toothbrush tree of Prophet Muhammad, J. Ethnopharmacol. 213 (2018) 409–444, http://dx.doi.org/10.1016/j.jep.2017.11.030.
[3] S. Siddeeqh, A. Parida, M. Jose, V. Pai, Estimation of antimicrobial properties of aqueous and alcoholic extracts of *Salvadora persica* (Miswak) on oral microbial pathogens - an invitro study, J. Clin. Diagn. Res. 10 (2016) FC13–FC16, http://dx.doi.org/10.7860/JCDR/2016/22213.8524.
[4] A. Sofrata, E.M. Santangelo, M. Azeem, A.K. Borg-Karlson, A. Gustafsson, K. Pütsep, Benzyl isothiocyanate, a major component from the roots of *Salvadora persica* is highly active against Gram-negative bacteria, PLoS One 6 (2011), http://dx.doi.org/10.1371/journal.pone.0023045, e23045.
[5] G. Valentino, V. Graziani, B. D'Abrosca, S. Pacifico, A. Fiorentino, M. Scognamiglio, NMR-based plant metabolomics in nutraceutical research: an overview, Molecules 55 (2020) 1444, http://dx.doi.org/10.3390/molecules25061444.
[6] M.A. Farag, A. Porzel, L.A. Wessjohann, Unraveling the active hypoglycemic agent trigonelline in *Balanites aegyptiaca* date fruit using metabolite fingerprinting by NMR, J. Pharm. Biomed. Anal. 115 (2015) 383–387, http://dx.doi.org/10.1016/j.jpba.2015.08.003.
[7] A.A. Maamoun, R.H. El-akkad, M.A. Farag, Mapping metabolome changes in *Luffa aegyptiaca* Mill fruits at different maturation stages via MS-based metabolomics and chemometrics, J. Adv. Res. (2019), http://dx.doi.org/10.1016/j.jare.2019.10.009, In press.
[8] M.A. Farag, S. Fahmy, M.A. Choucry, M.O. Wahdan, M.F. Elsebai, Metabolites profiling reveals for antimicrobial compositional differences and action mechanism in the toothbrushing stick "miswak" *Salvadora persica*, J. Pharm. Biomed. Anal. 133 (2017) 32–40, http://dx.doi.org/10.1016/j.jpba.2016.11.018.
[9] M.A. Farag, A. Porzel, E.A. Mahrous, M.M. El-Massry, L.A. Wessjohann, Integrated comparative metabolite profiling via MS and NMR techniques for Senna drug quality control analysis, Anal. Bioanal. Chem. 407 (2015) 1937–1949, http://dx.doi.org/10.1007/s00216-014-8432-1.
[10] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based metabolomic analysis of plants, Nat. Protoc. 5 (2010) 536–549, http://dx.doi.org/10.1038/nprot.2009.237.
[11] G. Blunden, S.M. Gordon, W.F.H. McLean, G.R. Keysell, β-stachydrine, a novel betaine from Griffithsia flos*culosa*, Phytochemistry 22 (1983) 293, http://dx.doi.org/10.1016/S0031-9422(00)80108-9.
[12] K. Kuchta, R.B. Volk, H.W. Rauwald, Stachydrine in *Leonurus cardiaca*, *Leonurus japonicus*, *Leonotis leonurus*: detection and quantification by instrumental HPTLC and 1H-qNMR analyses, Pharmazie 68 (2013) 534–540, http://dx.doi.org/10.1691/ph.2013.6527.
[13] A.T. Khalil, Benzylamides from *Salvadora persica*, Arch. Pharm. Res. 29 (2006) 952–956, http://dx.doi.org/10.1007/BF02969277.
[14] M.S. Kamel, K. Ohtani, M.H. Assaf, R. Kasai, M.A. El-Shanawani, K. Yamasaki, A.A. Ali, O. Tanaka, Lignan glycosides from stems of *Salvadora persica*, Phytochemistry 31 (1992) 2469–2471, http://dx.doi.org/10.1016/0031-9422(92)83301-E.
[15] N. Al-bagieh, Antiherpes simplex virus type 1 activity of benzylisothiocyanate, Biomed. Lett. 47 (1992) 67–70.
[16] M. Farag, W.M. Abdel-Mageed, O. Basudan, A. El-Gamal, Persicaline, a new antioxidant sulphur-containing imidazoline alkaloid from *Salvadora persica* roots, Molecules 23 (2018) e483, http://dx.doi.org/10.3390/molecules23020483.

[17] W. Lu, J. Zou, J. Chen, Y. Zhu, Q. Ya, W. Zhao, Two new compounds from *Cleidion brevipetiolatum*, Fitoterapia 78 (2007) 714–716, http://dx.doi.org/10.1016/j.fitote.2007.06.007.
[18] W.T.Y. Loo, L.J. Jin, M.N.B. Cheung, L.W.C. Chow, Evaluation of Ellagic acid on the activities of oral bacteria with the use of adenosine triphosphate (ATP) bioluminescence assay, Afr. J. Biotechnol. 9 (2010) 3938–3943, http://dx.doi.org/10.5897/AJB2010.000-3270.
[19] M. Correia-da-Silva, E. Sousa, M.M.M. Pinto, Emerging sulfated flavonoids and other polyphenols as drugs: nature as an inspiration, Med. Res. Rev. 34 (2014) 223–279, http://dx.doi.org/10.1002/med.21282.
[20] Y.C.F. Teles, M.S.R. Souza, M. de F.V. De Souza, Sulphated flavonoids: biosynthesis, structures, and biological activities, Molecules 23 (2018) e480, http://dx.doi.org/10.3390/molecules23020480.
[21] A.B. Ray, L. Chand, S. Dutta, C. Salvadourea, New urea derivative from *Salvadora persica*, Chem. Ind. (London, United Kingdom). 12 (1975) 517–518.
[22] S.K. El-Desouky, A new cytotoxic chromen-C-glucoside from the leaves of *Salvadora persica*, Int. J. Pure Appl. Biosci. 5 (2017) 277–281, http://dx.doi.org/10.18782/2320-7051.2926.
[23] G.C. Galletti, G. Chiavari, Y.D. Kahie, Pyrolysis/gas chromatography/ion-trap mass spectrometry of the 'tooth brush' tree (*Salvadora persica* L.), Rapid Commun. Mass Spectrom. 7 (1993) 651–656, http://dx.doi.org/10.1002/rcm.1290070719.
[24] S.D. McNeil, M.L. Nuccio, A.D. Hanson, Betaines and related osmoprotectants. Targets for metabolic engineering of stress resistance, Plant Physiol. 120 (1999) 945–950, http://dx.doi.org/10.1104/pp.120.4.945.
[25] Q. Garrett, N. Khandekar, S. Shih, J.L. Flanagan, P. Simmons, J. Vehige, M.D.P. Willcox, Betaine stabilizes cell volume and protects against apoptosis in human corneal epithelial cells under hyperosmotic stress, Exp. Eye Res. 108 (2013) 33–41, http://dx.doi.org/10.1016/j.exer.2012.12.001.
[26] I. Rantanen, J. Tenovuo, K. Pienihäkkinen, E. Söderling, Effects of a betaine-containing toothpaste on subjective symptoms of drymouth: a randomized clinical trial, J. Contemp. Dent. Pract. 4 (2003) 11–23.
[27] J.H. Lim, S.H. Song, H.S. Park, J.R. Lee, S.M. Lee, Spontaneous detachment of *Streptococcus mutans* biofilm by synergistic effect between zwitterion and sugar alcohol, Sci. Rep. 7 (2017) e8107, http://dx.doi.org/10.1038/s41598-017-08558-x.
[28] F. Cheng, Y. Zhou, M. Wang, C. Guo, Z. Cao, R. Zhang, C. Peng, A review of pharmacological and pharmacokinetic properties of stachydrine, Pharmacol. Res. 155 (2020) e104755, http://dx.doi.org/10.1016/j.phrs.2020.104755.
[29] R.M. Perez, Anti-inflammatory activity of compounds isolated from plants, Sci. World J. 1 (2001) 713–783, http://dx.doi.org/10.1100/tsw.2001.77.
[30] M. Valerio, A.B. Awad, β-sitosterol down-regulates some pro-inflammatory signal transduction pathways by increasing the activity of tyrosine phosphatase SHP-1 in J774A.1 murine macrophages, Int. Immunopharmacol. 11 (2011) 1012–1017, http://dx.doi.org/10.1016/j.intimp.2011.02.018.