Authenticity assessment of some unconventional oils from the Moroccan desert

vorgelegt von

Said El Harkaoui, M. Sc.
ORCID: 0000-0002-2791-2566

an der Fakultät III - Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Cornelia Rauh
Gutachter: Prof. Dr. Stephan Drusch
Gutachter: Dr. Bertrand Matthäus
Gutachter: Prof. Dr. Simon Hammann

Tag der wissenschaftlichen Aussprache: 12. Januar 2026

Berlin 2026

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Bertrand Matthäus, for his guidance throughout this PhD journey and for welcoming me to the Department of Safety and Quality of Cereals at the Max Rubner-Institut in Detmold. Thank you for granting me the freedom to explore different research directions, the opportunity to work with various analytical facilities, and for the many insightful discussions. I especially appreciate what I have learned from you about navigating projects involving multiple partners.

I would also like to thank my supervisor, Prof. Dr. Stephan Drusch, for the deep scientific discussions about the results and their interpretation, including from a statistical perspective. Your careful proofreading has elevated every one of my texts to a new level, and your help in building a clear structure and coherent storyline was invaluable.

My sincere thanks go to Prof. Dr. Sascha Rohn for always being available for questions and discussions. Your prompt, clear, and constructive feedback has been extremely helpful in moving the work forward. I have appreciated your careful and supportive proofreading, which always felt like a true peer review.

I would like to express my appreciation to Prof. Zoubida Charrouf for showing me the other side of scientific work, the close collaboration with farmers and cooperatives, and the social aspects that you so passionately pioneer. Thank you for making your vast network available to me throughout the project.

I am also very grateful to Prof. Dr. Simon Hammann for his willingness to evaluate this thesis.

Many thanks to Prof. Hanae El Monfalouti, Prof. Badr Kartah, and Prof. Said Gharby for the valuable discussions and your help with sample collection. Thank you for the opportunity to collaborate and supervise some of your students. It was a pleasure to work closely with your teams.

My warmest thanks go to all the scientists, PhD students, and staff of the lipid group at the Max Rubner-Institut in Detmold, as well as colleagues from other departments, for welcoming me so warmly at the beginning of my PhD a particularly difficult time as it coincided with the COVID-19 pandemic. I am grateful for your extraordinary support and for the many great moments we shared in the lab, in the office, during business trips, and outside of work. Your encouragement to learn German has played an important role in my success in the telc B2 exam with gut erfüllt.

Special thanks to Ina, Ludger, Klaus, Philipp, Katharina, Theresa, Maximilian, and Christina for the many casual scientific chats and exchanges. Thank you Nelli for the nice photos you did for my samples, Heike for your support with sterol measurements, Dominik for always being the person to call when when searching for stuff in the lab, and Ralph for your help with the ToF and, not to forget, the great coffee cups.

I would also like to thank all the cooperatives involved in the project for their support, valuable input from the field and the market, and the open communication throughout the process.

Many thanks also to the members of the KIDA team: Micha, Aaron, Cristina, and Julia for your support and the great opportunity to collaborate in the field of authenticity and machine learning.

From the bottom of my heart, I thank my mother and father and brothers for their unwavering trust and support not only during the PhD, but also throughout all the years of my studies even when I was often far away from home. Thank you for always encouraging me to go further and achieve more. I owe my deepest gratitude to my wife, Latifa, for her endless support, for moving to Germany with me, and for her patience during stressful times. Thank you for bringing our lovely son, Mohamed-Amin into the world, he is the most wonderful gift.

Finally, I gratefully acknowledge the financial support from the Federal Ministry of Agriculture, Food and Regional Identity (BMLEH), based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) [support program: FKZ 2819DOKA03].

Declarations

During the preparation of this work I used OpenAI's ChatGPT (version GPT-4o, accessed via chat.openai.com) in order to improve the readability and language of the manuscript. All scientific content and interpretations were my own and all the outputs were carefully reviewed and revised by me before inclusion in this work.

Abstract

Investigating new sources of vegetable oils is particularly relevant for countries like Morocco, where drought and heat limit the production of traditional oilseed crops, meeting only 20% of national demand. This dissertation characterizes balanites (*Balanites aegyptiaca*), cactus (*Opuntia ficus-indica*), and date palm (*Phoenix dactylifera*) as unconventional sources of vegetable oil in the Moroccan desert focusing on their chemical composition and authenticity assessment.

The first objective was to define the chemical identity of these oils by analysing their main lipid groups: fatty acids (FA), triacylglycerols (TAGs), tocochromanols, and phytosterols. In Balanites kernel oil (BKO) and cactus seed oil (CO), linoleic acid was the dominant fatty acid, followed by oleic, palmitic, and stearic acids. A distinctive feature of CO was its relatively high content (~5%) of vaccenic acid, an isomer of oleic acid. Date seed oil (DSO), by contrast, exhibited a more diverse FA profile, including oleic, lauric, myristic, palmitic, linoleic, and stearic acids which partly explains the high oxidative stability of DSO. TAG analysis supported the FA results, with DSO showing the most diverse profile (34 TAGs). Regarding tocochromanols, BKO was rich in α-tocopherol, CO in γ-tocopherol, while DSO was characterized by high tocotrienol content, especially α-tocotrienol. All three oils shared a similar phytosterol composition, dominated by β-sitosterol, campesterol, and Δ5-avenasterol, with CO showing the highest total content.

To explore the effect of geographical origin, it was hypothesized that oil composition varies by collection area and that some variation could be linked to climate conditions such as water deficit and/or temperature. Despite limitations in regional sample size, multivariate statistical analysis was used to identify origin-related trends. The degree and nature of these variations depended on the region and the oil type. For instance, Moroccan BKO samples differed clearly from those collected in Sudan and Mauritania, mainly due to major compounds from all four chemical classes (FA, TAGs, tocochromanol, and phytosterol). In contrast, for DSO at a smaller geographical scale (three Moroccan palm groves), especially minor TAGs were key to distinguishing origin. These findings were discussed in relation to known biosynthetic pathways and stress-related enzymatic responses.

Still within the scope of geographical origin, an untargeted metabolomic approach was applied to DSO as a case study, to determine whether the polar metabolite profile could reflect geographical variation, by revealing clustering trends and potential origin-specific markers not captured by classical lipid analysis. Samples from three Moroccan palm groves were analysed by UHPLC-ESI-QTOF-MS in both positive and negative ionisation modes. PCA results showed a similar clustering trend as observed with lipid composition, with samples from *Allougoum* forming a distinct group compared to *Alnif* and *Errachidia*. Based on these results, an OPLS-DA model was used to identify the discriminative features. Among the top 50 discriminative features, 25 metabolites from various chemical classes were tentatively identified and hydroxy fatty acids were the most represented class. These compounds, reported for the

first time in DSO, expand the current understanding of its chemical profile. The data processing workflow used was based entirely on open-source tools and can be readily applied to other oils such as BKO and CO.

For the second aspect of authenticity, adulteration detection, the study aimed to build a machine learning-based model that does not rely on large sets of physical mixtures. Instead, simulation methods were used to generate synthetic data for model training and testing. The hypothesis was that integrating analytical data with simulation and machine learning could enable reliable adulteration detection. CO was used as a test case, with refined sunflower oil (SO) as adulterant. After analysing FA, TAG, and tocochromanol profiles of pure CO and SO, two simulation methods were tested: Monte Carlo (MC) and Conditional Tabular Generative Adversarial Network (CTGAN). MC performed consistently well, even with small datasets, whereas CTGAN was less effective. Using a weighted sum formula, the simulated oils were used to create multiple levels of adulteration. These simulated mixtures were used to train Random Forest (RF) and Neural Network (NN) models. RF outperformed NN, achieving 94% accuracy on simulated data and 90% on real test samples, with better interpretability and lower computational demand. Thus, combining MC simulation with RF is proposed as a robust approach for oil adulteration detection. The methodology, implemented in Python and shared as open-source code, can be easily adapted to other oils with minimal retraining.

Overall, this dissertation presents a multi-approach framework for assessing Moroccan oils authenticity, integrating lipid profiling, untargeted metabolomics, and machine learning. It advances understanding of chemical composition, demonstrates traceability of origin, and offers a robust strategy for detecting adulteration.

Kurzfassung

Die Erschließung neuer Quellen für pflanzliche Öle ist hat besonders relevant für Länder wie Marokko eine große Bedeutung. Hier ist der , wo Dürre und Hitze den Anbau traditioneller Ölpflanzen durch Hitze und Trockenheit stark eingesschränkent und kann derzeit nur etwa 20 % des nationalen Bedarfs gedeckt deckenwerden können. In Ddiese Dissertation werden charakterisiert Balanites (Balanites aegyptiaca), Kaktusfeige (Opuntia ficus-indica) und Dattelpalme (Phoenix dactylifera) als unkonventionelle Ölquellen mit Anbaumöglichkeiten in der marokkanischen Wüste charakterisiert. Der, mit Fokus liegt dabei auf deren Untersuchung der chemischenr Zusammensetzung der Öle und den Möglichkeiten, diese für die Bewertung der Authentizität zu nutzensbewertung.

Das erste Ziel der Arbeit war die chemische Identifizierung Charakterisierung dieser Öle durch die Analysehinsichtlich ihrer wichtigsten LipidbestandteileHauptlipidgruppen: Fettsäuren (FA), Triacylglycerine Triacylglyceride (TAGs), Tocochromanole und Phytosterole. In Balaniteskernöl (BKO) und Kaktussamenöl (CO) dominierte Linolsäure, gefolgt von Ölsäure, Palmitinsäure und Stearinsäure. Charakteristisch für CO war der vergleichsweise hohe Gehalt an Vaccensäure (~5 %), einem seltenen Isomer der Ölsäure. Dattelöl (DSO) zeigte ein deutlich vielfältigeres FA-Profil, einschließlich Ölsäure, Laurinsäure, Myristinsäure, Palmitinsäure, Linolsäure und Stearinsäure. Der hohe Anteil an einfach ungesättigten und gesättigten Fettsäuren kann teilweise , was teilweise die hohe oxidative Stabilität dieses Öls erklärenerklärt. Die Analyse der TAG Zusammensetzung-Analyse bestätigte die FA-Ergebnisse, wobei DSO aufgrund seiner verschiedenen Fettsäuren ein breites Spektrum an TAGdas diverseste Profil (34 TAGs) aufwies. Bezüglich der Tocochromanole war BKO reich an α-Tocopherol, CO an γ-Tocopherol, während DSO durch einen hohen Tocotrienolgehalt, insbesondere an α-Tocotrienol, charakterisiert war. Alle drei Öle wiesen eine ähnliche Phytosterolzusammensetzung auf, dominiert von β-Sitosterol, Campesterol und Δ5-Avenasterol, wobei CO den höchsten Gesamtgehalt zeigte.

In einem zweiten Ansatz wurde die Hypothese aufgestellt, dass die Ölzusammensetzung durch dieHinsichtlich der geographischen Herkunft der Ölsamen beeinflusst wirdwurde die Hypothese aufgestellt, dass sich die Ölzusammensetzung je nach geographischer Herkunft unterscheidet und Unterschiede in der Zusammensetzung auf klimatische Bedingungen wie Wasserdefizit und/oder Temperatur zurückzuführen sein könnten. Trotz der begrenzter begrenzten Probenanzahl pro Region konnten durch multivariate statistische Analysen herkunftsbezogeneein Einfluss der geographischen HerkunftTrends auf die Zusammensetzung der Öle identifiziert werden. Der Grad und die Art der Variation hingen von Region und Öltyp ab. Beispielsweise unterschieden sich marokkanische BKO-Proben deutlich von jenen aus dem Sudan und Mauretanien, hauptsächlich aufgrund der Hauptverbindungen aller vier chemischen Klassen (FA, TAGs, Tocochromanole und Phytosterole). Im Fall von DSO, bBei Betrachtung kleinerer geographischer Maßstäbe Entfernungen (drei marokkanische Palmenhaine), spielten im Fall von DSO, insbesondere minore TAGs eine entscheidende Rolle bei der

Unterscheidung der Herkunftsunterscheidung. Diese Ergebnisse wurden im Zusammenhang mit bekannten biosynthetischen Pfaden der Biosynthese dieser Verbindungen und dem Auftreten von stressbedingten Enzymreaktionen diskutiert.

Immer noch im Rahmen der Herkunftsuntersuchung Untersuchungen zum Herkunftsnachweis wurde für DSO ein ungezielter ungerichteter metabolomischer Ansatz angewandt, um zu prüfen, ob das polare Profil der Metabolitenprofil durch die Aufdeckung von Clusterbildungen und die Identifizierung herkunftsspezifischen herkunftsspezifischer Markern geographische Unterschiede widerspiegeln kann, die durch klassische Lipidanalyse nicht erfasst werden. Proben aus drei marokkanischen Palmenhainen wurden mittels UHPLC-ESI-QTOF-MS in positiver positivem und negativem er Ionisationsmodeus analysiert. Eine PCA zeigte eine ähnliche Clusterbildung wie bei der Lipidzusammensetzung, wobei sich der Standort Allougoum deutlich von den Standorten Alnif und Errachidia unterschied. Basierend auf diesen Ergebnissen wurde ein OPLS-DA-Modell verwendet, um diskriminierende Merkmale zu identifizieren. Unter den Top 50 wurden 25 Metaboliten aus verschiedenen chemischen Klassen tentativ identifiziert, wobei Hydroxyfettsäuren am häufigsten vertreten waren. Diese in DSO erstmals beschriebenen Verbindungen erweitern das Verständnis seines chemischen Profils. Der verwendete Workflow für die Datenverarbeitungsworkflow basierte vollständig auf Open-Source-Tools und kann ohne Weiteresauch auf andere Öle wie BKO und CO angewendet werden.

Für den zweiten Aspekt dieser Arbeit, der Authentizität, die Erkennung von Verfälschungen, wurde ein mittels maschinellems Lernen ein Mmodell entwickelt. , Statt der üblichen Verwendung einer großen Zahl von Ölmischungen für die Erstellung des Modellsumfangreicher physikalischer Mischungen wurden Simulationsmethoden eingesetzt, um ausgehend von einer geringen Zahl von Daten zur Zusammensetzung der Öle, synthetisch Daten für das Training und den Test der Modelle zu generieren. Die Hypothese war, dass die Kombination analytischer Daten mit Simulation und maschinellem Lernen eine zuverlässige Erkennung von Verfälschungen ermöglicht. CO wurde als Fallstudie verwendet, wobei raffiniertes Sonnenblumenöl (SO) aufgrund seiner ähnlichen Fettsäurenzusammensetzung als zur Verfälschungsmittel diente. Nach Analyse der FA-, TAG- und Tocochromanol-Profile von reinem CO und SO wurden zwei Simulationsmethoden getestet: Monte Carlo (MC) und Conditional Tabular Generative Adversarial Network (CTGAN). MC lieferte auch bei kleinen Datensätzen konsistent gute Ergebnisse, während CTGAN weniger effektiv war. Auf Basis der MC-Simulation wurden mit einer gewichteten Summenformel unterschiedliche Verfälschungsgrade generiert. Diese simulierten Mischungen dienten zum Training von Random Forest (RF) und Neural Network (NN)-Modellen. RF zeigte die beste Leistung bei der Zuordnung der mit 94 % Genauigkeit bei simulierten und 90 % bei realen Proben. Zudem war RF interpretierbarer und rechenökonomischer als NN. Die Kombination aus MC-Simulation und RF wird daher als robuste Methode zur Ölverfälschungserkennung vorgeschlagen. Die gesamte Methodik wurde in Python implementiert und als Open-Source-Code veröffentlicht, wodurch sie leicht auf andere Öle übertragbar ist.

Insgesamt präsentiert diese Dissertation einen Multi-Ansatz-Rahmen zur Bewertung der Authentizität marokkanischer Öle, der Lipidprofiling, untargeted einen ungerichteten Ansatz des polaren Metabolomics und sowie maschinelles Lernen integriert. Sie erweitert das Verständnis der zur chemischen Zusammensetzung dieser Öle, demonstriert die Möglichkeiten zur Rückverfolgbarkeit des Ursprungs und liefert eine robuste Strategie zur Detektion von Verfälschungen mit preiswerteren Ölen.

Table of content

List of figures

List of abbreviations

ANOVA	Analysis of variance
BKO	Balanites kernel oil
Ca	Capric acid
CO	Cactus seed oil
CTGAN	Conditional tabular generative adversarial network
Cy	Caprylic acid
DPPH	2,2-Diphenyl-1-picrylhydrazyl
DGF	Deutsche Gesellschaft für Fettwissenschaft e.V.
DSO	Date seed oil
EAG	Equivalent of gallic acid
ESI	Electrospray ionization
FA	Fatty acid
FAME	Fatty acid methyl esters
FFA	Free fatty acids
FLD	Fluorescence detector
FID	Flame ionization detector
GC	Gas chromatography
HCA	Hierarchical cluster analysis
HMGR	3-Hydroxy-3-methyl glutaryl CoA reductase
HPLC	High performance liquid chromatography
HPPD	p-hydroxyphenylpyruvate dioxygenase
IC50	Half maximal inhibitory concentration
IOC	International olive council
ISO	International organization for standardization
KLD	Kullback-Leibler divergence
La	Lauric acid
Li	Linoleic acid
MC	Monte Carlo
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
My	Myristic acid
NN	Neural network
OIT	Oxidation induction time

Ol	Oleic acid
OPLS-DA	Orthogonal partial least squares discriminant analysis
Pa	Palmitic acid
PCA	Principal component analysis
PLS-DA	Partial least squares discriminant analysis
PV	Peroxide value
RF	Random forest
SD	Standard deviation
SFA	Statured fatty acids
SO	Sunflower oil
St	Stearic acid
TAG	Triacylglycerol
TPC	Total polyphenols content
UFA	Unsaturated fatty acid
UHPLC	Ultra-high-performance liquid chromatography
UV	Ultraviolet
VIP	Variable importance projection

List of publications

This dissertation is based on the following manuscripts (I – IV):

Manuscript I
El Harkaoui, S., El Kaourat, A., El Monfalouti, H., Kartah, B.E., Mariod, A.A., Charrouf, Z., Rohn, S., Drusch, S. and Matthäus, B., 2024. Chemical Composition and Geographic Variation of Cold Pressed Balanites aegyptiaca Kernel Oil. Foods, 13(7), p.1135. https://doi.org/10.3390/foods13071135.

Manuscript II
Nounah, I., **El Harkaoui, S**., Hajib, A., Gharby, S., Harhar, H., Bouyahya, A., Caprioli, G., Maggi, F., Matthäus, B. and Charrouf, Z., 2024. Effect of seed's geographical origin on cactus oil physico-chemical characteristics, oxidative stability, and antioxidant activity. Food Chemistry: X, p.101445. https://doi.org/10.1016/j.fochx.2024.101445.

Manuscript III
El Harkaoui, S., N'Diaye, K., Gharby, S., Al-Hilal, M., Charrouf, Z., Rohn, S. and Matthäus, B., 2024. Insights into date seed oil composition: Geographical variability and potential applications. European Journal of Lipid Science and Technology, p.2400061. https://doi.org/10.1002/ejlt.202400061.

Manuscript IV
El Harkaoui, S., N'Diaye, K., Charrouf, Z., Rohn, S., Drusch, S. and Matthäus, B., 2025. Untargeted metabolomic analysis of date seed oil (Phoenix dactylifera L.) using UHPLC-ESI-QTOF-MS evaluation of the geographical origin effect. Food Chemistry: X. p.103162. https://doi.org/10.1016/j.fochx.2025.103162

Manuscript V
El Harkaoui, S., Cruz, C.O., Roggenland, A., Schneider, M., Rohn, S., Drusch, S. and Matthäus, B., 2025. Adulteration detection in cactus seed oil: Integrating analytical chemistry and machine learning approaches. Current Research in Food Science, p.100986. https://doi.org/10.1016/j.crfs.2025.100986.

1- Introduction

Investigating new sources of vegetable oils has gained importance due to the growing population and diverse industrial applications. Particularly desirable are oil-producing plants that thrive in environments where conventional oilseed plants struggle, such as areas with water deficits and high temperatures (Cheng et al., 2016; Tahir et al., 2022). This need is especially pressing in countries like Morocco, where the arid and semi-arid climate poses challenges to agricultural sustainability and food security (Gumus et al., 2024). Morocco's agricultural sector is vital to its economy, contributing to 14-20% of the GDP and employing about 40% of the population. However, with 90% rainfed land, the sector is highly vulnerable to climate change (El Mekkaoui et al., 2021; Lago-Olveira et al., 2024; Moussadek et al., 2024). Local conditions of drought and heat, intensified by climate change, limit the production potential of traditional oilseeds like rapeseeds and sunflower seeds. Consequently, domestic production of vegetable oils, including olive oil, meets only about 20% of the national demand, and is expected to suffer yield reductions of up to 30% by 2050 due to climate change (Kettani et al., 2018). This shortfall has led to a reliance on imports of oilseeds and crude oils, which is not always reliable, especially during global crises such as the COVID-19 pandemic or the Russia–Ukraine war, which have disrupted supply chains and driven up prices (Mengoub et al., 2022).

One promising approach to address these challenges is the exploration of unconventional oils derived from resilient plant species native to the Moroccan desert. Plants such as the cactus (*Opuntia ficus-indica*), date palm (*Phoenix dactylifera*), and balanites (*Balanites aegyptiaca*) are remarkably adapted to arid conditions. The valorization and promotion of these robust plants as a source of vegetable oil usually starts in local cooperatives made up of the native population of the areas where these plants grow. For example, the argan oil production model has shown how cooperative-based oil production can be successful, with 1,014 cooperatives supporting 10,834 people and producing 5640 tons of argan oil in 2019, 25% of which was exported (ODCO, 2024; Ministère de l'agriculture, 2024). This cooperative-based model should also be followed for other plant materials like cactus, date palm, and balanites. The involvement of local populations in these cooperatives, ensures that the knowledge, skills, and benefits of oil production are rooted in the communities themselves, fostering sustainable development and resilience against economic and environmental challenges. Additionally, cooperative-based oil production is technically improved compared to household production, with better adherence to good production practices, specially hygienic standards and appropriate packaging (El Kabous & Ouhssine, 2024; Faouzi, 2012). This fundamental quality control allows a focus on other critical aspects such as authenticity to promote these unconventional oils.

Ensuring the authenticity of edible oils is critical for market acceptance, consumer trust, and public health. Authenticity as defined by the International Food Authenticity Assurance Organization is "the process of irrefutably proving that a food or food ingredient is in its original, genuine, verifiable, and intended form as declared and represented". The fight against food fraud remains a pressing concern, and with the

introduction of Regulation (EU) 2017/625 on March 15, 2017, there has been an increased emphasis on legal measures to address this issue. This regulation expands risk-oriented control beyond food safety to include the authenticity of food (Willenberg & Matthäus, 2019). Among food products, vegetable oils are particularly vulnerable to fraud. In fact, "fats and oils" ranked fifth among the most frequently reported fraudulent product categories in the EU, accounting for 6.8% of all notifications (European Commission, 2023). Vegetable oil fraud typically involves adulteration or false origin claims (Moore et al., 2012; Sudhakar et al., 2023; Wang et al., 2024; Yuan et al., 2020). Adulteration involves practices such as substituting more expensive oils with cheaper ones or blending cold-pressed oils with refined ones. False geographical origin claims mislead consumers about the provenance of the oil, which is often an important quality indicator, as seen with Italian extra virgin olive oil (Bimbo et al., 2020; Casadei et al., 2021). The concept of "geographical origin" as related to authenticity involves a range of environmental factors that influence oil composition. Factors such as water availability, temperature, soil quality, and salinity can vary significantly between locations. Studies, especially those conducted under monitored growth conditions, have demonstrated strong correlations between these environmental variables and oil composition. Enzymatic activity within key metabolic pathways plays a crucial role in this variability (Chaudhary & Khurana, 2009; Ghaffari et al., 2023; Kumar et al., 2018; Porokhovinova et al., 2022). Fraud practices in both aspects, adulteration and geographical traceability, not only violate the interests of consumers, but also pose severe risks to public health (Othman et al., 2023). Numerous incidents of vegetable oil fraud have been reported worldwide, with some leading to severe health consequences, such as the olive oil toxic syndrome in Spain (Sudhakar et al., 2023). These well-known cases illustrate the dangers of oil fraud, but many other instances likely go unnoticed. The issue of authenticity is especially critical for unconventional oils, which, despite their edible nature, are often marketed for their unique flavors and potential health or cosmetic applications but have historically received less regulatory attention. Unlike conventional oils with established standards and regulations, such as those set by CODEX or the International Olive Council (Codex, 1999; International Olive Council (IOC), 2024), unconventional oils require rigorous scientific support to ensure their quality and authenticity.

In general, there are two main analytical strategies used for authenticity investigation: targeted analysis and untargeted analysis (Amaral, 2021; Ballin & Laursen, 2019). Targeted analysis focuses on detecting a specific component or target substance with a high degree of precision and specificity. While this approach is effective, it may miss unknown components, and in such cases, we usually turn to untargeted analysis. Compared to targeted analysis, untargeted analysis is a more comprehensive approach that aims to capture all possible components in analytical samples, including unknown components. However, this comprehensive approach generates large and complex datasets, making data processing and interpretation a key challenge (Amaral, 2021; Ballin & Laursen, 2019; Gao et al., 2019; Wang et al., 2024). Following these two analytical strategies, up-to-date, almost all analytical instruments have been used to investigate the authenticity of vegetable oil, essentially involving mass spectrometry (MS)-based studies coupled with chromatography techniques such as GC and HPLC. Other techniques are also often used, including

spectroscopic technologies like near-infrared, mid-infrared, UV-vis, fluorescence spectroscopy, Raman spectroscopy, hyperspectral imaging, and nuclear magnetic resonance. Additionally, sensor technologies, including electronic nose and electronic tongue, have been used for oil authenticity (Deng et al., 2024; Esteki et al., 2018; Medina et al., 2019; Osorio et al., 2014; Sudhakar et al., 2023). While spectroscopy methods are fast, non-destructive, and use fewer non-friendly chemicals, chromatography techniques are superior in terms of robustness, sensitivity, precision, and accuracy, which are key parameters needed especially for authenticity studies (Esteki et al., 2018; Ilić et al., 2022; Osorio et al., 2014; Sudhakar et al., 2023). GC and HPLC, often coupled with detectors such as MS, UV, FID, and FLD, provide detailed chemical fingerprints and are standard tools in most quality control and authenticity laboratories.

Supporting analytical approaches with chemometrics has proven to be a powerful tool for detecting adulteration in vegetable oils (Esteki et al., 2018; Willenberg & Matthäus, 2019). Chemometrics involves the application of mathematical and statistical methods to interpret complex analytical data. This is particularly valuable in authenticity studies where relying on a single chemical parameter, such as an individual fatty acid or tocopherol, may not be sufficient. Instead, chemometric tools allow the combination of multiple variables and facilitate the extraction of meaningful patterns and correlations from complex datasets (Kamal & Karoui, 2015; Sudhakar et al., 2023). Various chemometric models in combination with chromatographical techniques have been applied to detect adulteration or to trace the geographical origin of oils, each demonstrating its performance in different scenarios (Deng et al., 2024; Medina et al., 2019; Sudhakar et al., 2023). Those models are either unsupervised or supervised. The most popular unsupervised model is PCA, used mainly for reducing the dimensionality of the data, allowing visibility and patterns investigating in complex datasets, such as the effect of geographical origin. HCA is another unsupervised method often used to support PCA results by grouping similar samples based on their chemical profiles. On the other hand, supervised methods such as PLS-DA and its orthogonal form OPLS-DA are more commonly used for marker and pattern selection (Mattoli et al., 2022; Sudhakar et al., 2023; Wang et al., 2024; Yuan et al., 2020). These methods are particularly effective in handling collinear and noisy data, which is often encountered in complex chemical datasets. PLS-DA works by finding the directions in the predictor space that best separate the classes of samples, making it a powerful tool for classification tasks. Additionally, OPLS-DA enhances this by removing orthogonal variations that are unrelated to class separation, thereby improving model interpretability. Advanced statistical techniques, such as machine learning and deep learning algorithms, are also employed to execute more complex tasks. Methods like Random Forests and Neural Networks have shown promise in accurately classifying and predicting the authenticity of oils (Sudhakar et al., 2023; Wang et al., 2024; Yuan et al., 2020). These algorithms can handle large datasets and capture intricate patterns that simpler statistical models might miss.

In this PhD work, cactus (*Opuntia ficus-indica*), date palm (*Phoenix dactylifera*), and balanites (*Balanites aegyptiaca*) were considered as an unconventional source of vegetable oil. With the term “unconventional oils” we refer to oils that are not widely commercialized or standardized, despite being used at the local or

regional level. The focus was about investigating the authenticity of the oils, either regarding geographical origin or adulteration.

Balanites Kernel Oil (BKO): *Balanites aegyptiaca*, a member of the *Zygophyllaceae* family, is widely grown in arid and semi-arid regions of Africa, the Arabian Peninsula, and South Asia (Ahmed et al., 2020). In Morocco, it thrives in the Moroccan Sahara, particularly in regions such as Assa Zag, Ait Ouabelli, and Tata, along the Oued Draat (Ozenda, 1991). Each tree can yield 100-150 kg of fruit annually, with the dark fleshy pulp being the edible part (Saini et al., 2021). The pulp is rich in protein (1.2-1.5%) and sugars (35-42%) and contains essential minerals such as calcium, magnesium, phosphorus, potassium, and sodium (Mariod, 2019; Murthy et al., 2020). The kernel, making up 8-12% of the fruit's weight and containing 40-51% oil, is the source of BKO (Chapagain et al., 2009; Mohamed et al., 2002). The extraction and utilization of BKO in Morocco are limited to a few traditional medicinal applications (Abouri et al., 2012; Bellakhdar, 1997), with no commercial extraction or utilization till now. However, in many Sahel countries like Mauritania, Sudan, and Burkina Faso, the oil is traditionally used as an edible oil and sold in local markets (Lykke et al., 2021; Ouédraogo et al., 2023). Indeed, in Mauritania, the local population has started creating small production units (similar to the concept of cooperatives in Morocco) for BKO production with more appropriate processes, and the extracted oil is sold for both edible and cosmetic purposes.

Date Seed Oil (DSO): the date palm tree (*Phoenix dactylifera*), a member of the palm family *Arecaceae*, is widely spread in the hot desert regions of South-West Asia and North Africa (Echegaray et al., 2023). Date palm is the most important arboricultural plant in the arid area of Morocco; it covers 61,332 hectares, mainly in the Draâ-Tafilalet province (77%), with smaller concentrations in Sous Massa (15%), Oriental (5%), and Guelmim Oued Noun (4%) (Bouhlali et al., 2017; Elhoumaizi et al., 2023). Date production in Morocco was estimated at 143,160 tons in 2020, up from 101,351 tons in 2010 (FAO 2023). The fruit of the date palm is known for its high nutritional value, containing complex sugars, proteins, and antioxidant compounds (Bouhlali et al., 2017). Despite their potential, the seeds, making up about 10-15% of the fruit, are often considered waste and used primarily as animal feed. However, these seeds can yield 6-12% oil, which exhibits various biological activities and has demonstrated pharmacological effectiveness against certain diseases (Maqsood et al., 2020; Mrabet et al., 2020). It also holds potential applications in many fields, including the non-food sectors (Alkhoori et al., 2022; Farag et al., 2023; Maqsood et al., 2020; Mrabet et al., 2020). In Morocco, the extraction and commercialization of DSO just started as an approach for the valorization of date seeds, and it is carried out by very few cooperatives in areas like *Errachidia*, where date fruit processing is prevalent.

Cactus Seed Oil (CO): the cactus pear (*Opuntia ficus-indica*), or the prickly pear, belongs to the *Cactaceae* family and thrives in arid and semi-arid regions worldwide, including Africa, Mediterranean countries, the Southwestern United States, and Northern Mexico (Silva et al., 2021). In Morocco, *Opuntia ficus-indica* grows extensively, particularly in the Southwestern regions (Sabbahi & Hock, 2022). Recognizing its

significant agronomic value within the Moroccan context, the Moroccan Ministry of Agriculture, as part of the Green Morocco Plan, initiated an ambitious program aimed at expanding cactus plantations to 160,000 hectares by 2020 (Akroud et al., 2021; Sabbahi & Hock, 2022). Unfortunately, this promising trajectory was affected by the invasion of *Dactylopius opuntiae*, a pest that damaged many of the cactus-growing areas. The Moroccan authorities are actively working to combat the infestation with the expectations for recovery and improved yields (Alkhoori et al., 2022; Sabbahi & Hock, 2022). The edible part of the cactus is the fruit, containing up to 300 small seeds, which can yield 5-15% oil (Chbani et al., 2020; Chbani et al., 2023). CO shows diverse biological activities, including antioxidant, antimicrobial, antidiabetic, lipid-lowering, anticancer, anti-inflammatory, and anti-ulcer properties as highlighted in our recent review and other articles (Al-Naqeb et al., 2021; Barba et al., 2017; Chbani et al., 2023; Ramadan et al., 2021). Despite its edible nature, the primary application of CO currently lies in the cosmetic industry, where it is prized for its moisturizing, anti-aging, and skin-nourishing properties (Chbani et al., 2023). In Morocco, the production and commercialization of CO is more established compared to BKO and DSO. It is mainly produced and marketed by cooperatives in the regions where the cactus plant is abundant in Morocco. The oil is sought in both national and international markets, particularly in the cosmetic sector. Although there are no official statistics on the amount of oil produced and its turnover, cooperatives involved in this PhD project report a thriving production, with the oil available in supermarkets and national beauty centers. The cooperatives also reported that Japan is one of the primary export destinations for Moroccan CO. Despite its market presence, the CO sector remains relatively small compared to the well-established argan oil market.

At the start of this PhD project, it was clear that local cooperatives had recently begun producing and marketing these oils, reflecting growing interest in their nutritional and economic potential. However, the scientific understanding of their chemical composition and authenticity remained limited, especially in terms of internationally recognized standards. In contrast, conventional oils like rapeseed, sunflower, or soybean are well characterized and regulated by frameworks such as the Codex Alimentarius (Codex, 1999), which define benchmark values for key lipid constituents including FA, tocochromanol, and phytosterols. For the three Moroccan oils studied here, no such reference standards exist. Basic compositional data were either missing or very limited, and existing studies focused mainly on selected components, such as major fatty acids, offering only a fragmented view. For BKO, no compositional data from Moroccan sources had been published before this work. Chemical data for DSO are scarce, despite Morocco's rich biodiversity, including over 223 date varieties and many unexplored palm groves. CO, though gaining more attention recently, also remains at an early stage of scientific study.

Thus, determining the composition of these oils according to Codex-defined lipid classes and comparing them with known oils is a crucial first step toward their nutritional and commercial validation. This chemical information also supports the investigation of geographical variation, which has been documented in many vegetable oils (Lucini et al., 2020). Based on this, we hypothesize that the composition of BKO,

CO, and DSO may vary depending on regions, potentially due to known biosynthetic mechanisms. By generating detailed chemical profiles and applying statistical tools, this variation can be assessed. However, major lipid classes alone, while an essential starting point, do not capture the full chemical complexity of these oils. To gain a more comprehensive understanding, it is also important to consider the polar fraction, which, although smaller in proportion, may contain unique metabolites valuable for origin differentiation. Untargeted metabolomics using UHPLC-ESI-QTOF-MS can reveal such compounds, especially when combined with multivariate statistics. Another core dimension of authenticity addressed in this thesis is the detection of adulteration. It is crucial to develop robust, scalable, and cost-effective strategies for detecting adulteration, particularly with cheap refined oils. We hypothesize that analytical data, when integrated with simulation models and machine learning, can be leveraged for this purpose.

In light of these gaps, hypotheses, and challenges, this dissertation seeks to answer the following key research questions:

- What is the chemical composition of BKO, CO, and DSO in terms of FA, TAG, tocochromanol, and phytosterol? how does this compare to standard oils? and how does it vary with geographical origin?
- Can untargeted metabolomics reveal additional compounds for geographical differentiation?
- Can lipid profiling data, combined with simulation and machine learning, be used to detect adulteration?

Together, these questions aim to address critical gaps in the compositional knowledge and authenticity assessment of Moroccan unconventional oils, contributing to their scientific validation, quality assurance, and ultimately, their sustainable development and commercial success.

2- Cumulative part of the thesis

This dissertation characterizes cactus (*Opuntia ficus-indica*), date palm (*Phoenix dactylifera*), and balanites (*Balanites aegyptiaca*) as unconventional source of vegetable oil in the Moroccan desert with a special focus on the investigation of their authenticity. Each of the five presented manuscripts contains a thorough introduction, leading to a more profound understanding of the individual study objectives and their background. The manuscripts presented in this dissertation (**manuscript I, II, III, IV, and V**) have been published in peer-reviewed journals: Foods (**manuscript I**), Food Chemistry: X (**manuscript II**), European Journal of Lipid Science and Technology (**manuscript III**), Food Chemistry: X (**manuscript IV**), and Current Research in Food Science (**manuscript V**). Supporting information for these manuscripts is included in the appendix. Each article is reprinted with permission from the respective copyright holders and co-authors. All individuals involved in the research are listed as co-authors, and the specific contributions of the author of this thesis are indicated.

The first part of the thesis introduces those three plants as potential oil sources in arid regions with a focus on the chemical composition of the main lipid suggested by international standardization like CODEX. The analytical measurements were done using GC and HPLC coupled with different detectors, FID, FLD, and MS. Since the chemical composition of vegetable oils is known to vary with geographical origin, this hypothesis was tested for the investigated oils. Multivariate analysis was applied to identify patterns of variation and correlation with geographical origin, aiming to determine whether such variation also exists in DSO, BKO, and CO. Where possible, attempts were made to interpret these variations in relation to known biosynthetic pathways of the measured compounds (**manuscripts I-III**).

Once this foundational work was established, the research moved toward acquiring a deeper understanding of the oils chemical profile by implementing an untargeted metabolomics approach. Using UHPLC-ESI-QTOF-MS, untargeted metabolomics was applied to DSO as a case study. The resulting data were pre-processed, filtered, and analyzed using multivariate statistical methods. The significant features which were most likely to be responsible of the geographical origin discrimination were identified and discussed. The results are reported in **manuscript IV**. This analytical framework can be transferable to the other oils (BKO and CO), offering a versatile tool for future in-depth studies.

A third and equally important aspect was to address the authenticity of oils from the perspective of adulteration, a growing concern given the economic value of these unconventional oils. The objective was to develop a machine learning-based model for adulteration detection that does not rely on large sets of physical mixtures. Instead, it combines real analytical measurements with simulation algorithms to generate enough data needed to build robust classification models. CO was selected as a challenging case study, using refined sunflower oil as the adulterant. The model was trained to detect varying levels of adulteration and to determine the minimum detectable adulterant concentration. The whole methodology is coded in python so it can be replicated and all the data used is published in the form of data publications available

open access. This methodology, detailed in **manuscript V**, was designed to be generalizable and could also be adapted to assess the adulteration of BKO and DSO. The ultimate goal is to provide Moroccan quality control authorities and industry stakeholders with practical, scalable tools for ensuring oil integrity and protecting both producers and consumers.

Manuscript I

Chemical composition and geographic variation of cold pressed *Balanites aegyptiaca* kernel oil

Said El Harkaoui, Asma El Kaourat, Hanae El Monfalouti, Badr Eddine Kartah, Abdalbasit Adam Mariod, Zoubida Charrouf, Sascha Rohn, Stephan Drusch, Bertrand Matthäus

Foods 2024, 13, 1135

Accepted manuscript available via https://doi.org/10.3390/foods13071135

foods

MDPI

Article

Chemical Composition and Geographic Variation of Cold Pressed *Balanites aegyptiaca* Kernel Oil

Said El Harkaoui [1,2,3], Asma El Kaourat [4], Hanae El Monfalouti [4], Badr Eddine Kartah [4], Abdalbasit Adam Mariod [5,6], Zoubida Charrouf [4], Sascha Rohn [2], Stephan Drusch [3] and Bertrand Matthäus [1,*]

1 Department for Safety and Quality of Cereals, Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Schützenberg 12, 32756 Detmold, Germany; said.elharkaoui@mri.bund.de
2 Department of Food Chemistry and Analysis, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany; rohn@tu-berlin.de
3 Department of Food Technology and Food Material Science, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Königin-Luise-Str. 22, 14195 Berlin, Germany; stephan.drusch@tu-berlin.de
4 Laboratory of Plant Chemistry and Organic and Bio-Organic Synthesis, Faculty of Sciences, Mohammed V University in Rabat, 4 Avenue Ibn Battouta B.P., Rabat RP 1014, Morocco; asma.elkaourat@gmail.com (A.E.K.); h.elmonfalouti@um5r.ac.ma (H.E.M.); b.kartah@um5r.ac.ma (B.E.K.); zcharrouf@yahoo.fr (Z.C.)
5 Department of Biological Science, College of Science, University of Jeddah, Jeddah 21931, Saudi Arabia; basitmariod58@gmail.com
6 Indigenous Knowledge and Heritage Center, Ghibaish College of Science & Technology, Ghibaish P.O. Box 100, Sudan
* Correspondence: bertrand.matthaeus@mri.bund.de

Abstract: With the increasing impacts of climate change, establishing more sustainable and robust plants such as desert dates (*Balanites aegyptiaca*) seems to be necessary. Known for its resilience in arid conditions, this tree has the potential to become a more important food source, particularly for its potential to yield edible oil. This study characterized *Balanites* kernel oil (BKO) as a promising oil source in arid regions, studying the influence of geographical origin and environmental factors. Moroccan and Sudanese BKO samples were analyzed and compared with Mauritanian BKO. In the fatty acid profile, unsaturated fatty acids constituted over 70% of the BKO profile, with a predominance of linoleic acid (Li), oleic acid (Ol), palmitic acid (Pa), and stearic acid (St). Consequently, the predominant triacylglycerols were PaLiLi, PaLiOl, LiLiOl, OlLiOl, and StLiOl. α-Tocopherol dominated the tocochromanol composition (324 to 607 mg/kg), followed by γ-tocopherol (120 to 226 mg/kg), constituting 90% of the total tocochromanols. The total phytosterol content in BKO ranged from 871 to 2218 mg/kg oil, with β-sitosterol dominating (58% to 74%). Principal Component Analysis revealed that the geographical origin significantly influences BKO composition, emphasizing environmental factors, particularly water deficit and/or temperatures. Notably, Moroccan BKO collected from an area characterized by high aridity and relatively low winter temperatures, showcased a unique profile in fatty acid, phytosterols, and tocochromanols. The valorization of BKO presents an opportunity for local agricultural development in arid regions and a role model for plant development and agricultural practices in other parts of the world.

Keywords: *Balanites aegyptiaca*; date oil; fatty acid; triacylglycerol; tocochromanol; phytosterol; water deficit; temperature; multivariate analysis; sustainable utilization

Citation: El Harkaoui, S.; El Kaourat, A.; El Monfalouti, H.; Kartah, B.E.; Mariod, A.A.; Charrouf, Z.; Rohn, S.; Drusch, S.; Matthäus, B. Chemical Composition and Geographic Variation of Cold Pressed *Balanites aegyptiaca* Kernel Oil. *Foods* **2024**, *13*, 1135. https://doi.org/10.3390/foods13071135

Academic Editor: Antonio Piga

Received: 18 March 2024
Revised: 2 April 2024
Accepted: 5 April 2024
Published: 8 April 2024

1. Introduction

Establishing more sustainable and robust plants such as desert dates (*Balanites aegyptiaca* (L.) Delile.) seems to be necessary, because of the increasing impacts of climate change. This tree holds great esteem across the Sahelian and Saharan regions for its diverse traditionally uses, including its wood suitable for lighting fires for cooking, its edible fruits

diverse traditionally uses, including its wood suitable for lighting fires for cooking, its edible fruits with a sweet mesocarp and the oil made from the kernels, its animal feed value, and its numerous ethnomedicinal applications [1,2].

Balanites aegyptiaca is a multibranched, spiny, evergreen wild shrub or tree that grows up to a height of 10 m [3].. It is a widely grown desert plant, native to arid and semi-arid areas of Africa, the Arabian Peninsula, and South Asia [3]. The plant is known by various names, such as the desert date tree in the English language or *Heglig*, *Zaqqum*, and *Lalob* in the Arabic language, and it is also referred to as *Tugga*, *Hiliaje*, or *Taberkat* in some parts of Mauritania and in the Moroccan Desert [4–6]. The annual fruit yield per tree can reach 100–150 kg [7]. The fruit, a plum-sized and faintly five-grooved drupe, consists of an epicarp (5–9%), the mesocarp (28–33%), the endocarp (49–54%), and the kernel (8–12%) [8]. The fruit is green in the first stages of growth and turns yellow to brownish when ripe; it weighs on average 5.7 g and the kernel is about 0.7 g. Different parts of the *Balanites* fruit are illustrated in Figure 1. The dark fleshy pulp, the edible part of the fruit, contains 1.2–1.5% protein and a high sugar content (35–42%), predominantly present as reducing sugars (81.3–91.1%) [4]. Moreover, the fruits are also rich in minerals, including calcium, magnesium, phosphorus, potassium [9], and sodium [10]. Particularly interesting is the kernel's substantial oil content (40–51%), comprised of diverse fatty acids, including saturated, mainly palmitic and stearic acid, and unsaturated, mainly oleic and linoleic acid, making it a potential source for various applications [1,10,11]. Traditionally used in African and Indian pharmacopeias for treating ailments such as jaundice, epilepsy, and rheumatism, BKO has particularly gained attention in biodiesel production, and many publications support this application [8,12–16]. However, the focus should be extended beyond the prevailing trend in biodiesel production, aiming to redirect research towards a promoting broader application of BKO particularly in the food or cosmetic sectors, positioning BKO as a promising source of vegetable oil, especially in arid regions, but also as a role model for plant development and agricultural practices in other parts of the world.

Figure 1. Different parts of *Balanites aegyptiaca* fruit and *Balanites aegyptiaca* kernel oil. I: mature fruit, II: pulp, III: nut, IV: kernel, V: nut shell, VI: milled kernel, VII: kernel oil.

In recent years, noteworthy efforts have emerged, particularly in countries like Mauritania, where local cooperatives have taken the initiative to produce and commercialize BKO as both an edible oil (extracted from roasted kernels) and a cosmetic oil (extracted from unroasted kernels), drawing parallels with the production of argan oil in Morocco. This approach serves as a model that warrants encouragement and promotion in other regions where the *Balanites* tree exists. While local initiatives mark a positive stride, there exists a critical gap in comprehensive knowledge regarding the limitations in knowledge of its chemical composition, especially, how various factors like geographic origin and climate conditions, may impact composition.

The aim of this study was to provide additional insights into the lipid composition of BKO focusing on fatty acids, tocochromanols, phytosterols, and triacylglycerols and their variation according to geographical origin. This will contribute to introduce the *Balanites* tree to the scientific community as a potential source of edible vegetable oil.

For the characterization, samples from Morocco and Sudan were included in this study and compared with BKO produced by a Mauritanian cooperative serving as a reference. Despite the relatively small number of samples, multivariate statistics were used to test whether there were differences in the composition of the oils with respect to the location of cultivation.

2. Materials and Methods

2.1. Material

Balanites kernel oil, labeled as "alimentary" and "cosmetic", was obtained from a local cooperative in *Guidimakha*, Mauritania, and used as a reference. Additionally, *Balanites* fruits were harvested at the maturity stage from Morocco (three samples collected from the *Tata* area over three successive harvest years: 2020, 2021, and 2022) and Sudan (collected from three areas: *El Fulah*, *Al Fashir*, and *Al 'Abbasiyah*). The fruit processing steps included removing the pulp by soaking the entire fruit in warm water, washing it, drying the nuts in the sun, and finally cracking them to collect the kernels. These processing steps mirror those of the cooperatives in Mauritania. Details about the sample set, including corresponding temperatures and rainfall in the collection areas, can be found in Table 1. Additional information is available in Table S1.

Table 1. Harvest location, temperatures and rainfall. The rainfall and temperature data for the corresponding harvest year and location were obtained from Google Climate Engine [17].

Sample Code	Harvest Location	Temperature Min–Max (°C)	Rainfall (mm/Year)
Mo1	*Tata*–Morocco	3.98–41.04	69.7
Mo2	*Tata*–Morocco	6.08–41.54	64.0
Mo3	*Tata*–Morocco	6.02–41.49	205.4
Mau1	*Guidimakha*–Mauritania	18.03–43.49	352.5
Mau2	*Guidimakha*–Mauritania	18.03–43.49	352.5
Su1	*El Fulah*–Sudan	11.03–39.90	533.2
Su2	*Al Fashir*–Sudan	16.76–40.67	466.6
Su3	*Al 'Abbasiyah*–Sudan	18.91–38.12	591.8

2.2. Cold Pressing Process

Mechanical pressing of *B. aegyptiaca* kernels was carried out using a Komet single press (IBG Monforts Oekotec GmbH & Co. KG, Mönchengladbach, Germany). Due to the soft nature of kernels, small pieces of their shells were added in a 2:1 ratio (kernel: shell) to induce a friction effect during pressing. The pressing was performed using an 8 mm nozzle at temperatures kept below 60 °C, while the screw rotated at a speed of 35 rpm. The temperature of the resultant oil, remained below 40 °C. The extract obtained was centrifuged at 1550× *g* for 15 min to separate the oil from the sediment. All the oil samples were carefully preserved in sealed dark glass bottles and stored in a freezer (−18 °C) until later analysis.

2.3. Determination of the Fatty Acid Composition

The fatty acid composition was assessed through gas chromatography, employing the standard methodologies DGF C-VI 10a (00) and C-VI 11d (19) [18]. Initially, a droplet of oil was dissolved in 1 mL of *n*-heptane (LiChrosolv®, Merck KGaA, Darmstadt, Germany) for liquid chromatography. Subsequently, 50 µL of sodium methylate (30% solution in methanol; Merck KGaA, Darmstadt, Germany) were added and mixed with the oil. After brief agitation at room temperature for 1 min, 100 µL of Millipore water were introduced. The mixture was centrifuged at 1550× *g* for 5 min, and the lower aqueous phase was removed. Following this, 50 µL 1 M hydrochloric acid (min. 25%, p.a; Chemsolute®, Th.Geyer GmbH & Co. KG, Renningen, Germany) and a few drops of methyl orange indicator (ACS reagent, dye content 85%; Merck KGaA, Darmstadt, Germany) were added.

After a brief mix, the lower aqueous phase was removed. Twenty milligrams of sodium hydrogen sulfate (EMSURE® ACS, ISO, Reag. Ph Eur, Merck KGaA, Darmstadt, Germany) were added. After centrifugation at 1550× *g* for 5 min, the upper *n*-heptane phase was injected into an HP5890 gas chromatograph (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The chromatograph was equipped with a CP-Sil 88 capillary column (100 m × 0.25 mm × 0.25 µm; Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The temperature was gradually increased from 150 °C to 250 °C at a rate of 1.5 °C/min and held at 250 °C for 5 min. The injector and detector temperatures were set to 260 °C, and the carrier gas (H_2) flow rate was 1.7 mL/min with a 1:50 split ratio. The fatty acid methyl esters (FAME) were identified by comparing their retention times with a standard mix (Supelco®37 Component FAME Mix; Merck KGaA, Darmstadt, Germany), and their composition was quantified as a percentage of the total fatty acids (FA).

2.4. Determination of the Triacylglecerol Composition

The triacylglycerol (TAG) composition was evaluated using gas chromatography following the standard method DGF C-VI 14 (08) [18]. The analysis was conducted on an Agilent 6890 gas chromatograph equipped with a flame ionization detector (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). Fifty milligrams of oil were dissolved in 10 mL isooctane (EMSURE®, ACS, Reag. Ph Eur; Merck KGaA, Darmstadt, Germany), and 1 µL of this solution was injected into a Restek™ RTX®65TG column (30 m × 0.25 mm × 0.1 µm; Restek Corp., Bellefonte, PA, USA). The oven temperature was maintained at 300 °C for 1 min, then ramped from 300 °C to 360 °C at a rate of 2 °C/min and held at 360 °C for 10 min. The injector and detector were set to 370 °C, and the carrier gas (H_2) flow rate was 1 mL/min with a 1:40 split ratio. The detector operated using 40 mL/min hydrogen, 450 mL/min air, and 45 mL/min nitrogen. According to the DGF method, TAGs were identified by comparing their retention times with those of sesame oil, which were previously established for TAG composition analysis. The composition of the individual triacylglycerols was calculated as percentage of the individual peak areas from the total peak areas of all peaks.

2.5. Determination of the Tocochromanol Composition

To determine the tocochromanol content and composition, the standard method DGF F-II 4a (00) [18] was used. Initially, 150 mg of oil was dissolved in 1 mL of n-heptane (for liquid chromatography, LiChrosolv®, Merck KGaA, Darmstadt, Germany) and underwent filtration using a 1.0 µm PTFE filter (Whatmann™, Cytiva, Buckinghamshire, UK) followed by a 0.45 µm PTFE filter (Restek Corp., Bellefonte, PA, USA). The resulting filtered solution was then transferred into a vial and subsequently injected into the HPLC-FLD. The HPLC system comprised a pump (L-7100 LaChrom Elite®, Merck KGaA, Darmstadt, Germany), an autosampler (Spark Holland BV, Emmen, The Netherlands), a fluorescence detector (L-2485 LaChrom Elite®, Merck KGaA, Darmstadt, Germany), and the interface 35900E (software CAG BootP Server, Agilent Technologies Inc., Santa Clara, CA, USA). A diol phase column (25 cm × 4 mm × 5 µm, LiChroCART® 250-4, Merck KGaA, Darmstadt, Germany) was used for isocratic separation. The mobile phase consisted of n-heptane/tert.butyl methyl ether (96:4 (*v*/*v*), for liquid chromatography, LiChrosolv®, Merck KGaA, Darmstadt, Germany) at a flow rate of 1.3 mL/min. For all samples, the injection volume was 20 µL, and the analysis was conducted over 66 min. The fluorescence detector settings were an excitation wavelength of 295 nm and an emission wavelength of 330 nm. Identification of tocochromanols was done using α-, β-, δ-, and γ-tocopherol reference standards (chromatographic purity 97.6–99.6%, Merck KGaA, Darmstadt, Germany), and quantification was achieved through external calibration using standard solutions (0.25–40 µg/mL).

2.6. Determination of the Phytosterol Composition

The phytosterol composition was determined following the standard method DGF F-III 1 (98) [18]. An internal standard, betulin (98%; Merck KGaA, Darmstadt, Germany), was

dissolved in diisopropyl ether (EMSURE® ACS, Reag. Ph Eur; Merck KGaA, Darmstadt, Germany) at 1 mg/mL concentration. Subsequently, 250 mg of oil underwent refluxing in 5 mL ethanolic KOH solution for 15 min after adding the internal standard. The unsaponifiable components were isolated using solid phase extraction on an aluminum oxide column (Aluminum Oxide 90, Merck KGaA, Darmstadt, Germany). A thin-layer chromatography (TLC silica gel 60 plates, 20 cm × 20 cm; Merck KGaA, Darmstadt, Germany) was then used to separate the sterol fraction from the unsaponifiable matter, employing a mobile phase of *n*-heptane (LiChrosolv®, Merck KGaA, Darmstadt, Germany) and distilled diethyl ether (min. 99.5%, stabilized with BHT, Chemsolute®, Th.Geyer, Renningen, Germany) in a 50:50 v/v ratio. After separation, the sterol fraction was converted into silylated derivatives (TMS) using a silylating agent (*N*-Methyl-*N*-trimethylsilyl-heptafluorobutyramid, Macherey-Nagel GmbH & Co. KG, Düren, Germany). Gas chromatography (HP6890N, Agilent Technologies Deutschland GmbH, Waldbronn, Germany) was employed to determine the sterol composition. Compounds were separated on an SE 54 CB column (50 mm × 0.25 mm × 0.1 µm; Macherey-Nagel GmbH & Co. KG, Düren, Germany) with the oven temperature increasing from 245 °C to 265 °C at a rate of 5 °C/min and maintained at 265 °C for 40 min. The injector and detector temperatures were set to 320 °C. Identification of substances was based on relative retention times compared to betulin. Additionally, standard compounds (cholesterol, 99%; campesterol, 65%; stigmasterol, 95%; β-sitosterol, 95%; Merck KGaA, Darmstadt, Germany) were applied in the identification process.

2.7. Statistical Analysis

Chemical analyses were conducted in triplicates for each sample, with the mean calculated from these triplicates for subsequent analysis. Given the study's nature and the limited sample size, assumptions regarding normality were considered. To assess statistical significance, a one-way ANOVA followed by a Tukey-Kramer HSD test ($p < 0.05$) was conducted using JMP software (JMP 14.3.0, SAS Institute Inc., Cary, NC, USA). Additionally, principal component analysis (PCA) was employed. Initially, the data were autoscaled, followed by PCA using JMP software for further insights into the relationships among variables.

3. Results and Discussion

3.1. Fatty Acid Composition

The fatty acid (FA) composition of BKO samples was determined and the main FA are reported in Figure 2 (the full list of identified FA is given in Table S2). The unsaturated FA constitute over 70% of the FA composition with a predominance of linoleic acid (Li) at 38.0% to 48.5%, followed by oleic acid (Ol) at 23.4% to 35.6%. Palmitic (Pl) and Stearic (St) acids are present in comparable proportions, ranging from 11.1% to 14.5% and 9.7% to 13.0%, respectively. Additionally, minor amounts of α-linoleic acid, arachidic acid, vaccenic acid, and behenic acid were detected in the analyzed BKO samples. The abundance of linoleic acid, an essential omega-6 fatty acid, and oleic acid, a heart-healthy monounsaturated fatty acid, may promote BKO in food applications. In comparison to well established oils, the FA composition of BKO resembles that of sesame oil in terms of linoleic and oleic FA content [19,20]. FA composition is a critical factor in determining the oxidative stability and shelf life of vegetable oil, and vegetable oil with a low oleic/linolenic acid ratio, such as BKO, tends to have lower stability. Therefore, careful storage practices, including dark storage, nitrogen atmosphere, and refrigeration are recommended.

The FA composition of BKO has already been described partly in the literature with linoleic, oleic, palmitic, and stearic acids identified as the main FA. However, there are conflicting results regarding the predominant FA, with some studies agreeing with the present findings and indicating linoleic acid as the major FA, followed by oleic acid [3,8,14,21–24]. In contrast, other studies claimed that oleic acid is the most important FA in BKO, followed by linoleic acid [10–12,25]. Assuming consistent fruit maturity across studies, differences may be attributed to ecophysiological factors, partly influenced by geographic origin as

well. On the other hand, even within the same geographical location, a notable variability in FA composition is evident among *Balanites aegyptiaca* trees, particularly in the concentrations of oleic and linoleic FA [3]. These authors correlated the variance in FA composition with fruit morphology, highlighting that the highest oleic/linoleic ratio is observed in fruits characterized by the lowest pulp percentage [3]. However, the authors further highlighted that linoleic acid consistently retains its status as the primary FA, while the oleic/linoleic ratio may shift in response to fruit morphology. Based on the findings described by Mehanni et al. [22], Mohamed et al. [11], and Elbadawi et al. [21], extraction methods, including enzymatic extraction, solvent extraction, and different pre-processing practices like a prior roasting or boiling of the kernels, have been shown not to significantly affect the FA composition of BKO. Generally, appropriate roasting seems to have only a minimal effect on the fatty acid composition of oils derived from roasted kernels [26]. However, suboptimal roasting conditions may lead to a slight decrease in the percentage of polyunsaturated fatty acids, potentially due to thermal degradation [26]. In addition, the influence of various extraction techniques on the fatty acid composition appears to be minor [27,28]. Nevertheless, it remains essential to assess the effects of the mentioned pre-processing practices in terms of storability and oxidative stability considerations.

ratio, such as BKO, tends to have lower stability. Therefore, careful storage practices, including dark storage, nitrogen atmosphere, and refrigeration are recommended.

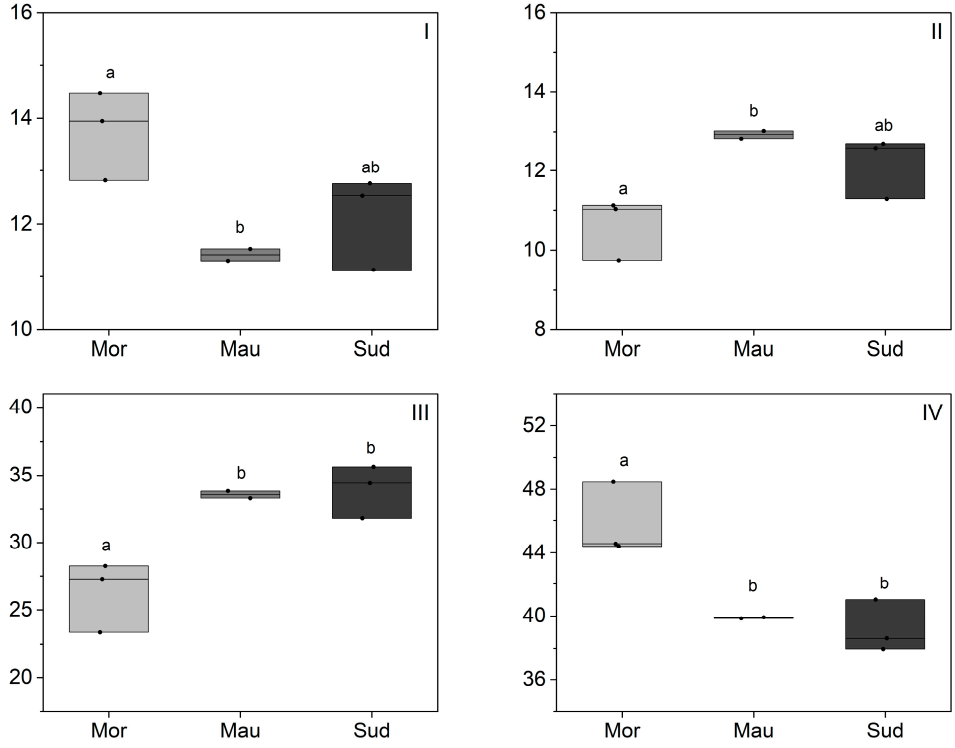


Figure 2. Fatty acid composition of BKO depending on geographic origin (%): Morocco (Mor), Mauritania (Mau), Sudan (Sud). (**I**) palmitic acid, (**II**) stearic acid, (**III**) oleic acid, (**IV**) linoleic acid. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with descriptive 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

The FA composition of BKO has already been described partly in the literature with linoleic, oleic, palmitic, and stearic acids identified as the main FA. However, there are conflicting results regarding the predominant FA, with some studies agreeing with the present findings and indicating linoleic acid as the major FA, followed by oleic acid [3,8,14,21–24]. In contrast, other studies claimed that oleic acid is the most important FA in BKO, followed by linoleic acid [10–12,25]. Assuming consistent fruit maturity across studies, differences may be attributed to ecophysiological factors, partly influenced by geographic origin as well. On the other hand, even within the same geographical location, a notable variability in FA composition is evident among *Balanites aegyptiaca* trees, particularly in the concentrations of oleic and linoleic FA [3]. These authors correlated the variance in FA composition with fruit morphology, highlighting that the highest oleic/linoleic ratio is observed in fruits characterized by the lowest pulp percentage [3]. However, the authors further highlighted that linoleic acid consistently retains its status as the primary

Regarding the geographical origin effects, significant differences in FA composition were observed among the analyzed BKO samples from Morocco, Sudan, and Mauritania, particularly in the main fatty acids, linoleic, oleic, palmitic, and stearic acid (Figure 2). Moroccan samples exhibited a distinctive profile characterized by higher levels of palmitic and linoleic acids and lower levels of stearic and oleic acids compared to samples from Mauritania and Sudan. This unique BKO profile in Morocco is likely attributed to the climate, marked by lower rainfall (water deficit) and a broader temperature range compared

to Sudan and Mauritania (Table 1), suggesting an adaptive response to water deficit and/or temperature. The adaptability to water deficit and/or temperature changes associated with alterations in FA composition, is well-documented for several plant species and has often been linked to changes in enzymatic activities involved in FA biosynthesis [29–31]. The proposed biosynthetic pathway of FA underscores a reciprocal quantitative relationship among different FA, monitored by various enzymes responsible for transitions between FA, primarily ketoacyl synthase and fatty acid desaturase [30]. Within this biosynthesis pathway, lower temperatures may increase the activity of oleate desaturase, favoring the synthesis of linoleic acid [30,32], as reported for cottonseed, sunflower, and camelina oil [30]. This mechanism could also explain the higher levels of linoleic acid observed in Moroccan BKO samples, and with larger samples set more concrete assumptions should be made. However, understanding the complete behavior of the plant under wild conditions remains challenging due to the multitude of interacting factors such as soil type, salinity, day length, and solar radiation levels.

3.2. Triacylglycerol Composition

A total of 14 triacylglycerols (TAG) were determined in BKO (Figure 3 and Table S3) with carbon numbers (CN) of 50, 52, and 54. The main TAG in the present BKO samples were PaLiLi (8.6–16.5%), PaLiOl (13.7–15.4%), LiLiOl (11.0–13.2%), OlLiOl (7.2–11.9%), and StLiOl (8–11.1%).

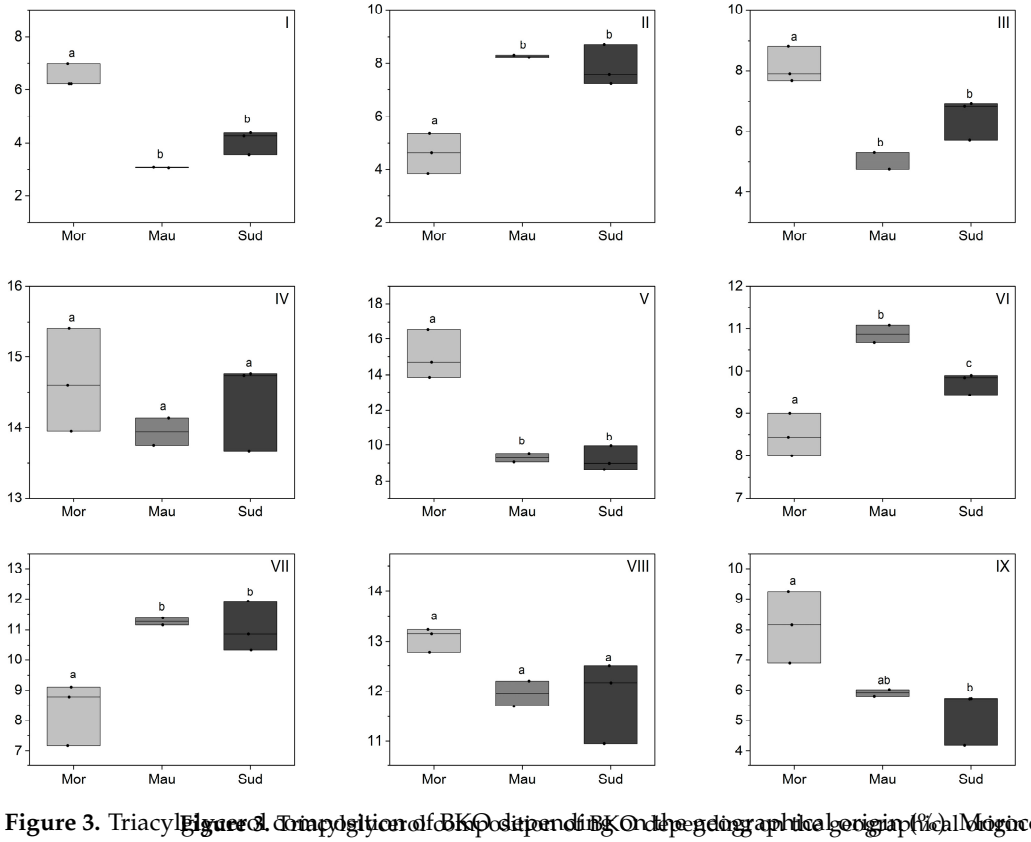


Figure 3. Triacylglycerol composition of BKO depending on the geographical origin (%). Morocco (Mor), Mauritania (Mau), Sudan (Sud). (**I**) PaLiLi, (**II**) PaLiOl, (**III**) PaLiSt, (**IV**) PaLiO, (**V**) PaLiL, (**VI**) StLiOl, (**VII**) OlLiOl, (**VIII**) LiLiOl, (**IX**) LiLiLi. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

The TAG composition was composed of a mixture of different fatty acids, reflecting the relative concentrations of oleic, linoleic, palmitic, and stearic acids as the main fatty acids. In terms of predominance according to the carbon number in BKO TAG, the predominant

3.3. Tocochromanol Composition

The tocochromanol composition of BKO was determined and showed total conten between 552 mg/kg and 828 mg/kg (Figure 4 and Table S4). Notably, α-tocopherol (α-to predominated in all analyzed samples, ranging from 324 to 607 mg/kg, followed by tocopherol (γ-toc) ranging from 120 to 226 mg/kg, both constituting 90% of the total t

ones had a carbon number of 54 (LiLiOl, OlLiOl, StLiOl, LiLiLi, OlOlOl, StOlOl, StOlSt), followed by 52 (PaLiLi, PaLiOl, PaLiSt, PaOlOl), and then 50 (PaLiPa, PaLiPa, PaOlPa). The study described by Diedhiou et al. [24] reported only three TAG in BKO (PaPlLi, PaOlLi, and StOlLi). BKO contains the most commonly represented TAG among the various oils reported in the literature. PaLiPa, PaLiLi, PaOlOl; LiLiLi, OlLiLi were identified in over 90% of all edible vegetable oils reported in the literature [33,34].

The variation of the TAG composition according to geographical origin followed a trend similar to what was observed for the fatty acid composition. Thus, TAG containing linoleate or palmitate (PaLiLi, LiLiOl, PaLiPa, LiLiLi) were significantly higher in Moroccan samples, whereas TAG containing oleate and stearate (StOlOl, StLiOl, PaOlOl, PaOlSt) were lower in Moroccan BKO samples compared with the samples from Sudan and Mauritania. Moroccan samples were characterized by a high level of TAG with CN 52 for TAG assembly. Thus, water deficit and/or temperature could have resulted in a different balance of isoforms for the lysophosphatidate acyltransferase or diacylglycerol acyltransferase reactions used in TAG accumulation. This hypothesis was suggested for differences in TAG for olive oil; however, no definite enzymatic evidence on these possibilities was suggested [35].

The presence of common TAG in BKO, challenges the establishment of a distinct TAG fingerprint; however, quantitative differences compared with other oils may serve as indicators.

3.3. Tocochromanol Composition

The tocochromanol composition of BKO was determined and showed total contents between 552 mg/kg and 828 mg/kg (Figure 4 and Table S4). Notably, α-tocopherol (α-toc) predominated in all analyzed samples, ranging from 324 to 607 mg/kg, followed by γ-tocopherol (γ-toc) ranging from 120 to 226 mg/kg, both constituting 90% of the total tocochromanol composition. Only two studies have been reported so far on the tocochromanol composition of BKO, indicating total tocochromanol contents of 510 mg/kg for BKO from Senegal [24] and 420 mg/kg for BKO from Sudan [21]. Both studies reported α-toc and γ-toc as the primary tocochromanols in BKO samples. The total tocochromanol content from these published studies was comparable to the content found in the present BKO from Mauritania and Sudan but lower than that found in the Moroccan BKO. In a study described by Elbadawi et al. [21], they also evaluated the effect of roasting or boiling *Balanites* kernels before oil extraction (part of the household extraction of BKO in some areas) on tocochromanol composition and reported no significant differences. In comparison to other oils, BKO is characterized by a higher tocochromanol content than olive oil and comparable values to argan and cactus oil; the particularity of BKO lies in its high content of α-toc, unlike argan and cactus oil, which are rich in γ-toc [27,36–38]. The high content of α-toc in BKO can be regarded as advantageous as it is considered the most biologically active form of vitamin E, which is preferentially utilized by the human body [39].

Regarding the geographical origin effect on the tocochromanol composition of BKO, significant differences were found. Moroccan samples exhibited a higher amount of total tocochromanol (mean value: 819 mg/kg) compared to Mauritania (mean 582 mg/kg) and Sudan (mean 575 mg/kg). Across all geographical origins, α-toc was the primary isomer. Moroccan samples displayed a high α-toc content (mean 577 mg/kg) in contrast to Mauritania (mean 435 mg/kg) and Sudan (mean 364 mg/kg). The elevated tocochromanol levels in BKO from Morocco could be attributed to the effects of water deficit, aligning with findings in other oils such as almond, olive, and soybean [40–42]. Additionally, cold stress may also contribute to the elevated tocochromanol levels in Moroccan BKO, as reported for rice [43]. The authors conducted expression profiling studies, demonstrating that cold stress enhances the expression of p-hydroxyphenylpyruvate dioxygenase (HPPD), leading to an increased supply of precursors for the biosynthesis of tocochromanols [43]. This first aspect of variation underscores the influence of ecophysiological factors on the overall tocochromanol profile. For a second aspect, the percentage of individual tocochromanol,

it was noted that the percentage of each isomer to the total tocochromanol content varies based on geographical origin. The percentage of α-toc in Moroccan and Mauritanian samples was 70% (mean value) and 75% (mean value), respectively, while in Sudan, it represented only 63% (mean value). Conversely, the percentage of γ-toc in Moroccan and Mauritanian samples was 25% (mean value) and 21% (mean value), respectively, whereas in Sudan, γ-toc represented 35%. This variation in α- and γ-toc between analyzed BKO samples may be explained by the biosynthesis pathway influenced by water deficit and/or temperature, which differed in the collection areas (Table 1). The biosynthesis pathway of tocopherol (shikimate pathway), starting with p-hydroxyphenylpyruvic (HPP), shows that there is a relationship between α- and γ-toc, as α-toc is generated from the methylation of the aromatic ring of γ-toc, and this reaction is catalyzed by γ-toc methyl-transferase (γ-TMT) [44]. The higher α-toc/γ-toc in Moroccan and Mauritanian BKOs may be due to a water deficit and/or temperature, catalytic effect on γ-TMT. A similar trend of variation was also found in other plant species [40,41,43].

Foods **2024**, *13*, x FOR PEER REVIEW 9 of 16

α-toc in BKO can be regarded as advantageous as it is considered the most biologically active form of vitamin E which is preferentially utilized by the human body [39].

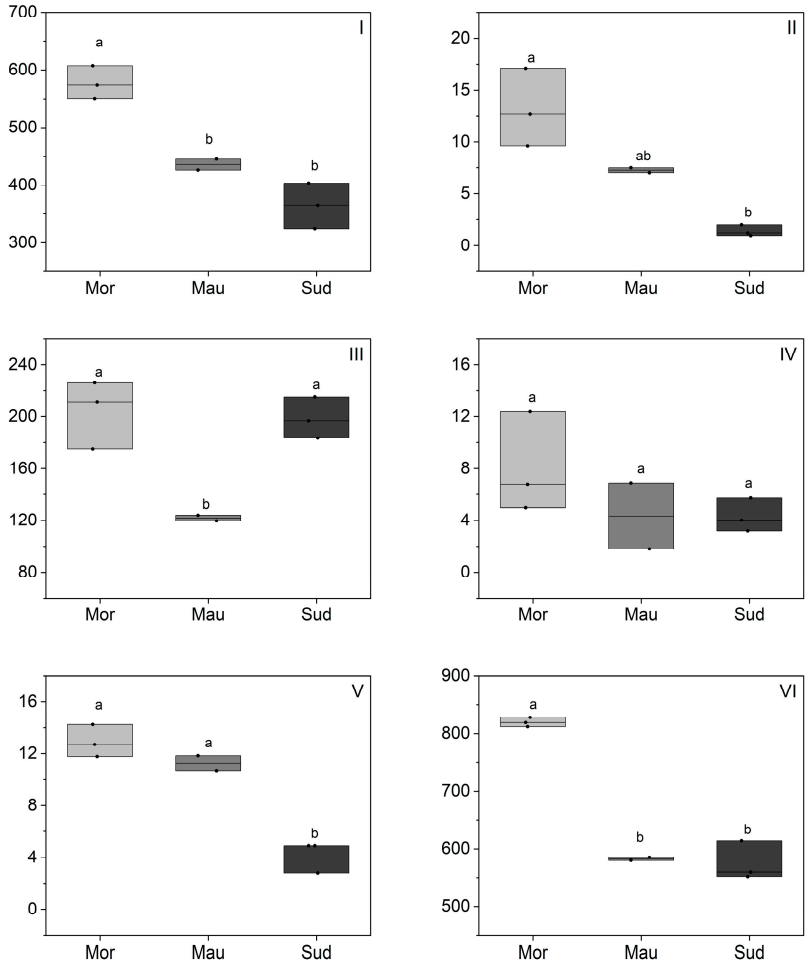


Figure 4. Tocochromanol composition of BKO depending on geographical origin (mg/kg). Morocco (Mor), Mauritania (Mau), Sudan (Sud). (**I**) α-tocopherol, (**II**) β-tocopherol, (**III**) γ-tocopherol, (**IV**) plastochromanol-8, (**V**) δ-tocopherol, (**VI**) total tocochromanols. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

Figure 4. Tocochromanol composition of BKO depending on geographical origin (mg/kg). Morocco (Mor), Mauritania (Mau), Sudan (Sud). (**I**) α-tocopherol, (**II**) β-tocopherol, (**III**) γ-tocopherol, (**IV**) plastochromanol-8, (**V**) δ-tocopherol, (**VI**) total tocochromanols. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

Regarding the geographical origin effect on the tocochromanol composition of BKO, significant differences were found. Moroccan samples exhibited a higher amount of total tocochromanol (mean value: 819 mg/kg) compared to Mauritania (mean 582 mg/kg) and Sudan (mean 575 mg/kg). Across all geographical origins, α-toc was the primary isomer. Moroccan samples displayed a high α-toc content (mean 577 mg/kg) in contrast to Mauritania (mean 435 mg/kg) and Sudan (mean 364 mg/kg). The elevated tocochromanol levels in BKO from Morocco could be attributed to the effects of water deficit, aligning with findings in other oils such as almond, olive, and soybean [40–42]. Additionally, cold stress may also contribute to the elevated tocochromanol levels in Moroccan BKO, as reported for rice [43]. The authors conducted expression profiling studies, demonstrating that cold

3.4. Phytosterol Composition

The phytosterol composition of BKO is detailed in Figure 5 and Table S5, revealing a comprehensive profile. The total phytosterol content in BKO ranged from 871 to

2218 mg/kg oil, encompassing a diverse mixture of phytosterols. Notably, β-sitosterol dominates across all analyzed samples, with concentrations ranging from 570 to 1295 mg/kg oil, constituting 58% to 74% of the total phytosterol content in BKO. The results align with comparative studies in Senegal, where the reported total phytosterol content is 2110 mg/kg oil, and Nigeria, where it is 1102 mg/kg, emphasizing also the prevalence of β-sitosterol in BKO [24,46]. However, these were the only studies reporting a phytosterol composition of BKO. In comparison with other edible oils from arid regions like argan kernel oil or cactus seed oil, BKO exhibits comparable phytosterol content to argan oil and lower content in comparison with cactus oil [27,36,37]. The ubiquitous dominance of β-sitosterol in various vegetable oils, including soybean, corn, sesame, and palm oils is well-documented [19]. β-Sitosterol has multiple applications in various fields such as medicine and the global food industry [47] and this may promote BKO in food applications as well. In addition to β-sitosterol, BKO showcases other common phytosterols such as campesterol, Δ5-avenasterol, and stigmasterol. Cholesterol, previously detected in other vegetable oils such as camelina oil [48], was also found in the present BKO.

effect result only in total phytosterol content variation, with very small changes in the percentages of each individual phytosterol [35,52]. In contrast, other studies reported changes in total phytosterols and the percentages of each individual phytosterol [49,53]. For instance, the total phytosterol content in soybean seeds was affected by the temperature in the growing location, while the phytosterol composition was almost the same in all soybean seed samples, and it was not affected by growing location temperature [52]. Similarly, in the case of olive oil, irrigation caused a significant decrease in total phytosterol content as described above, while only small changes in the percentages of each individual phytosterol were noted [35]. Conversely, in sunflower seeds, a delay in sowing, leading to higher temperatures during seed formation, induced a general increase in total sterol concentration by up to 35%, as well as a significant variation in the β-sitosterol/campesterol ratio [53].

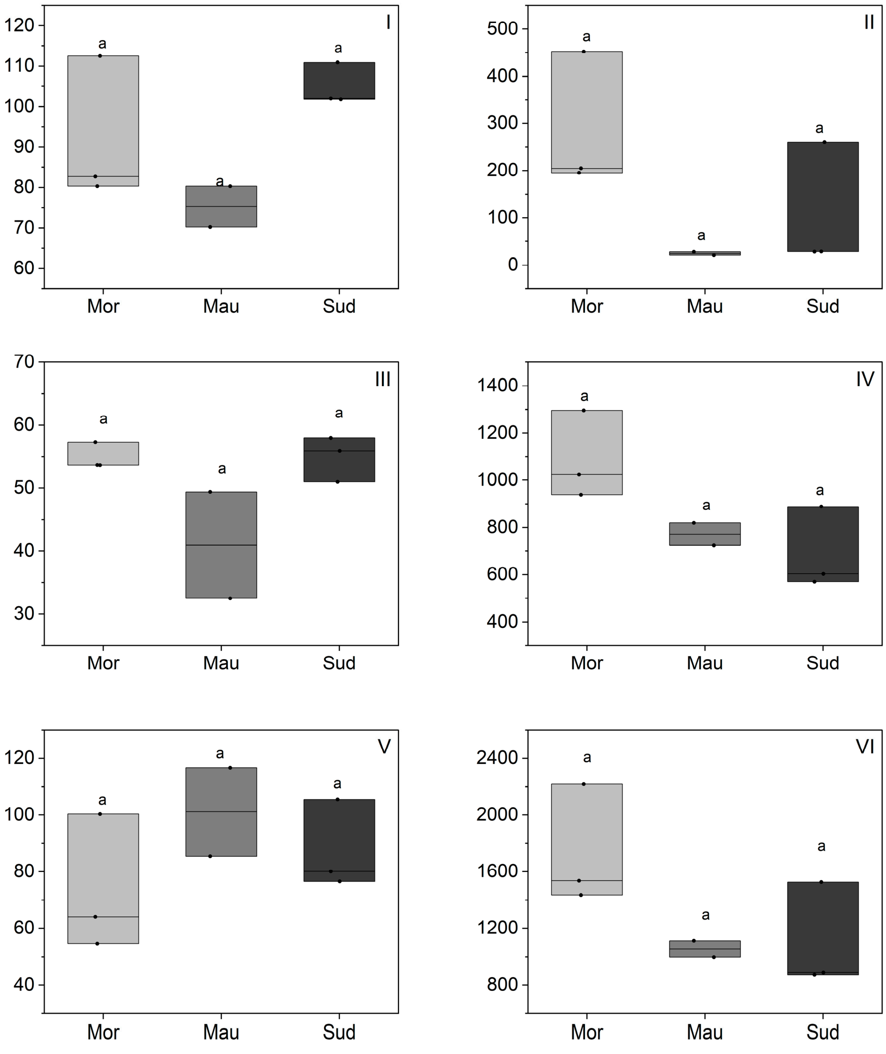


Figure 5. Phytosterol composition of BKO depending on geographic origin (mg/kg). Morocco (Mor), Mauritania (Mau), Sudan (Sud); (**I**) cholesterol, (**II**) campesterol, (**III**) stigmasterol, (**IV**) β-sitosterol, (**V**) Δ5-avenasterol, (**VI**) total phytosterol. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

Figure 5. Phytosterol composition of BKO depending on geographic origin (mg/kg). Morocco (Mor), Mauritania (Mau), Sudan (Sud); (**I**) cholesterol, (**II**) campesterol, (**III**) stigmasterol, (**IV**) β-sitosterol, (**V**) Δ5-avenasterol, (**VI**) total phytosterol. Statistical differences were identified using a one-way ANOVA test, which was then followed by a Tukey post-hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. Individual lines in box plots represent minimum, mean and maximum values.

Regarding the geographical origin effect on the total phytosterol composition, the results revealed notable differences, with Moroccan samples exhibiting the highest total phytosterol content (mean 1730 mg/kg oil) compared to Mauritania (mean 1054 mg/kg oil) and Sudan (mean 1095 mg/kg oil). This discrepancy aligns with ecophysiological factors, particularly the severity of the water deficit in Moroccan regions (Table 1), offering a plausible explanation for the elevated phytosterol content in Moroccan samples. The trends observed are in line with findings reported on other oils. For instance, an investigation into olive oil documented a substantial decrease in total phytosterol content following irrigation-induced alterations, dropping from 1604 to 1005 mg/kg [35]. Similarly, a study describing the effect of water deficit on the total phytosterol composition of sunflower oil under managed and controlled field conditions reported an increase in total phytosterol content under water deficit conditions [49]. A parallel trend was also reported for rice [50], where higher total phytosterol contents were noted in response to water deficit, with contents rising proportionally to the number of days of water stress. The authors attributed this to the increasing expression of HMG-CoA reductase, the limiting enzyme in the biosynthesis of plant sterols [50]. On the other hand, cold stress was also reported to increase the level of phytosterols in different plants [51]. These insights could be extended to other plant species, including *Balanites*, although further studies are needed for confirmation. Regarding the variation in the proportions of individual phytosterols, no clear trend emerged for the present BKO samples. Literature shows mixed findings, with some studies indicating that phytosterol variations in response to water deficit and/or temperature effect result only in total phytosterol content variation, with very small changes in the percentages of each individual phytosterol [35,52]. In contrast, other studies reported changes in total phytosterols and the percentages of each individual phytosterol [49,53]. For instance, the total phytosterol content in soybean seeds was affected by the temperature in the growing location, while the phytosterol composition was almost the same in all soybean seed samples, and it was not affected by growing location temperature [52]. Similarly, in the case of olive oil, irrigation caused a significant decrease in total phytosterol content as described above, while only small changes in the percentages of each individual phytosterol were noted [35]. Conversely, in sunflower seeds, a delay in sowing, leading to higher temperatures during seed formation, induced a general increase in total sterol concentration by up to 35%, as well as a significant variation in the β-sitosterol/campesterol ratio [53].

3.5. Multivariate Analysis

Principal component analysis (PCA) was employed to investigate the relationship between geographic location and chemical composition, providing initial insights into how the geographical origin influences the composition of BKO. This approach aimed to discern trends or correlations associated with specific origins by utilizing PCA, which was executed on a fused matrix comprising 8 rows corresponding to BKO samples and 44 columns representing the compositions of fatty acids, triacylglycerols, tocochromanols, and phytosterols. The combination of all variables into a single matrix allowed for the assessment of the contribution of both minor and major variables on a normalized scale.

As shown in Figure 6 (left, PCA Score Plot), the first two principal components (PC) explained 70% of the total variance (PC1: 52.7%; PC2: 17.2%). The score plot revealed a distinct distribution of oils based on geographical origin, with Moroccan samples notably separated along PC1, and Mauritanian and Sudanese samples distinguished along PC2 (refer to Figure 6 on the left). To better understand and visualize the correlation between chemical composition and geographical origin, a biplot was generated (Figure 6, right, biplot). The biplot visually illustrates the correlation between variables and geographical origin, offering valuable insights into the chemical composition of BKO associated with different regions. The correlated variables with PCs encompassed various chemical classes, including different FA, TAG, tocochromanols, and phytosterols, signifying the influence of geographical origin on the entire analyzed chemical profile. Moroccan samples, situated in

the positive part of PC1 (which exhibited maximum variance), displayed a distinct profile compared to Sudanese and Mauritanian samples. They exhibited positive correlations with Li and Pa fatty acids and corresponding TAG such as PaLiLi, LiLiLi, and PaLiPa. Additionally, positive correlations with α-tocopherol and β-sitosterol were also observed. The discrimination of Sudanese and Mauritanian samples occurred along the PC2 direction, with Sudanese samples in the positive part and Mauritanian samples in the negative part. The contribution of each variable is detailed in Figure S1, representing different loadings of the constructed PCA model.

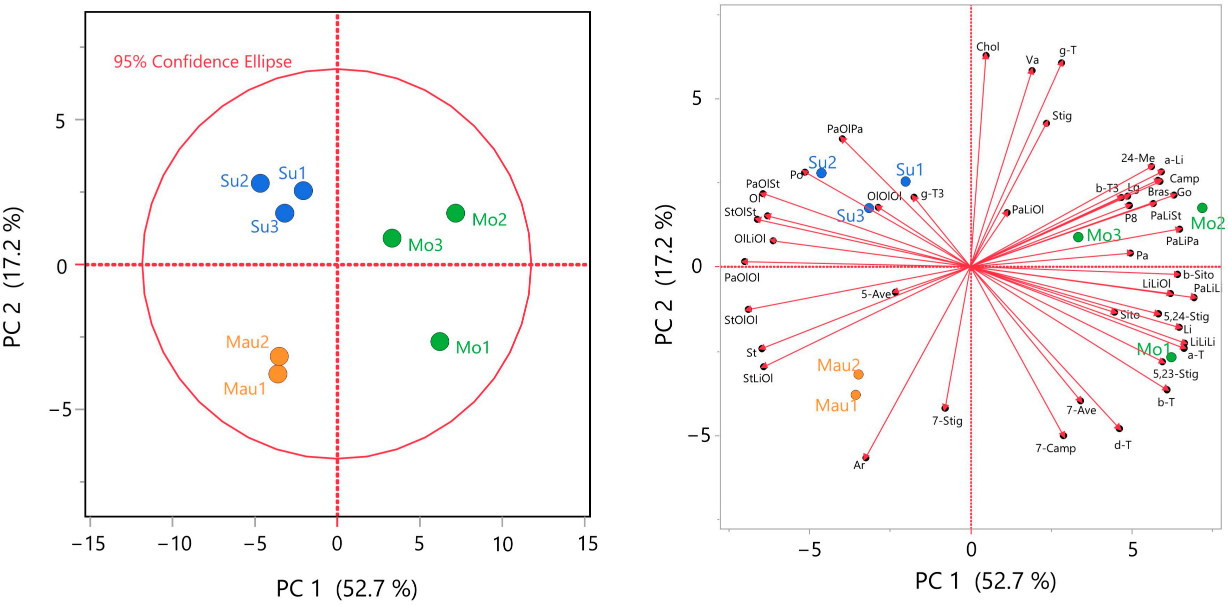


Figure 6. Score plot (**left**) and Biplot (**right**) resulting from Principal Component Analysis (PCA) based on the chemical composition of the examined BKOs. Variables < 0.1 were not included in the multivariate analysis, behenic acid also removed since it has constant values. Morocco (Mor), Mauritania (Mau), Sudan (Sud), Pa: palmitic acid, Po: palmitoleic acid, St: stearic acid, Ol: oleic acid, Va: vaccenic acid, Li: linoleic acid, a-Li: α-linolenic acid, Ar: arachidic acid, Go: gondoic acid, Lg: lignoceric acid, a-T: α-tocopherol, b-T: β-tocopherol, g-T: γ tocopherol, b-T3: β-tocotrienol, P8: plastochromanol-8, g-T3: γ-tocotrienol, d-T: δ-tocopherol, Chol: cholesterol, Bras: brassicasterol, 24-Me: 24-methylenecholesterol, Camp: campesterol, Stig: stigmasterol, 7-Camp: Δ7-campesterol, 5,23-Stig: Δ5,23-stigmastadienol, b-Sito: β-sitosterol, Sito: sitostanol, 5-Ave: Δ5-avenasterol, 5,24-Stig: Δ5,24-stigmastadienol, 7-Stig: Δ7-stigmastenol, 7-Ave: Δ7-avenasterol. The full sample names are available in Table S1 for reference.

The PCA results confirm the discussion in the preceding sections, suggesting that specific climate conditions, particularly water deficit and/or temperature, may explain the observed differences in the chemical composition of BKO samples. However, understanding the complete behavior of the plant under wild conditions remains challenging due to the multitude of interacting factors such as soil type, salinity, day length, and solar radiation levels.

4. Conclusions

The present study aimed at providing further insights into the lipid composition of BKO, focusing on FA, tocochromanols, phytosterols, and TAG and their variations based on geographical origin. Chemical and principal component analyses revealed significant geographical variation emphasizing the influence of ecosphysiological factors, notably water deficits and/or temperatures. While this study provides valuable insights into BKO's chemical composition and its geographical variations, further research with a larger and more diverse sample set is strongly recommended for a comprehensive understanding.

Future investigations will include the analysis of additional chemical compounds, especially polyphenols, and address the storability and oxidative stability, as well as the sensory properties of BKO. Setting and analyzing quality criteria that match legal aspects and being comparable to parameters that are already applied for other edible oils will help to make BKO a marketable edible oil. Despite its promising oil yield and chemical composition, the practicality of large-scale BKO production presents challenges, specifically with regards to the efficient removal of pulp and the cracking of *Balanites* nuts for kernel extraction. To ensure the high quality of the kernel, the utilization of appropriate machinery in the pulp removal process, discouraging the traditional household practice involving water is suggested. Additionally, for the cracking of nuts, the maintenance of manual cracking processes within local cooperatives is recommended. This approach, not only ensures efficient processing, but also guarantees an income for the local population and corresponds to the successful model of argan oil production, where the fruit is harvested and shelled manually while the subsequent steps are automated. The valorization of BKO might serve to motivate local communities to cultivate more resilient *Balanites* trees, especially in arid regions, thereby, promoting sustainability and fostering local agricultural development.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/foods13071135/s1, Table S1: Additional information on sample set, Table S2: Fatty acid composition of *Balanites* kernel oil (%), Table S3: Triacylglycerol composition of *Balanites* kernel oil (%), Table S4: Tocochromanol composition of *Balanites* kernel oil (mg/kg of oil), Table S5: Phytosterol composition of *Balanites* kernel oil (mg/kg of oil), Figure S1: Loadings of the two dimensions principal component analysis.

Author Contributions: Conceptualization, S.E.H., H.E.M., S.R., S.D. and B.M.; Formal analysis, S.E.H.; Funding acquisition, B.M.; Investigation, S.E.H. and A.E.K.; Methodology, S.E.H.; Project administration, B.E.K., Z.C. and B.M.; Resources, H.E.M., A.A.M. and B.M.; Supervision, Z.C., S.R., S.D. and B.M.; Validation, S.R.; Writing—original draft, S.E.H.; Writing—review & editing, A.E.K., H.E.M., B.E.K., A.A.M., Z.C., S.R., S.D. and B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Federal Ministry of Food and Agriculture (BMEL), support program: (FKZ 2819DOKA03).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abdelaziz, S.M.; Lemine, F.M.M.; Tfeil, H.O.; Filali-Maltouf, A.; Boukhary, A.O.M.S. Phytochemicals, Antioxidant Activity and Ethnobotanical Uses of *Balanites aegyptiaca* (L.) Del. Fruits from the Arid Zone of Mauritania, Northwest Africa. *Plants* **2020**, *9*, 401. [CrossRef] [PubMed]
2. Okia, C.A.; Agea, J.G.; Kimondo, J.M.; Refaat Atalla, A.A.; Obua, J.; Teklehaimanot, Z. Harvesting and processing of *Balanites aegyptiaca* leaves and fruits for local consumption by rural communities in Uganda. *J. Food Technol.* **2011**, *9*, 83–90. [CrossRef]
3. Ahmed, A.A.O.; Kita, A.; Nemś, A.; Miedzianka, J.; Foligni, R.; Abdalla, A.M.A.; Mozzon, M. Tree-to-tree variability in fruits and kernels of a *Balanites aegyptiaca* (L.) Del. population grown in Sudan. *Trees* **2020**, *34*, 111–119. [CrossRef]
4. Mariod, A.A. (Ed.) *Wild Fruits: Composition, Nutritional Value and Products*, 1st ed.; Springer: Cham, Switzerland, 2019; ISBN 9783030318840.
5. Bellakhdar, J. *La Pharmacopée Marocaine Traditionnelle: Médecine Arabe Ancienne et Savoirs Populaires*; Ibis Press: Paris, France, 1997; ISBN 9782910728038.
6. Abouri, M.; Mousadik, A.E.; Msanda, F.; Boubaker, H.; Saadi, B.; Cherifi, K. An ethnobotanical survey of medicinal plants used in the Tata Province, Morocco. *Int. J. Med. Plant Res.* **2012**, *1*, 99–123.
7. Saini, M.K.; Prasad, J.; Raju, P.V.S.; Kothari, S.L.; Harish; Shukla, J.K.; Gour, V.S. Morphological Descriptors and Heritability as Markers for Oil Yield in *Balanites aegyptiaca* (L.) Del.: A Potential Biodiesel Xerophyte. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **2021**, *91*, 695–706. [CrossRef]
8. Chapagain, B.P.; Yehoshua, Y.; Wiesman, Z. Desert date (*Balanites aegyptiaca*) as an arid lands sustainable bioresource for biodiesel. *Bioresour. Technol.* **2009**, *100*, 1221–1226. [CrossRef] [PubMed]

9. Murthy, H.N.; Yadav, G.G.; Dewir, Y.H.; Ibrahim, A. Phytochemicals and Biological Activity of Desert Date (*Balanites aegyptiaca* (L.) Delile). *Plants* **2020**, *10*, 32. [CrossRef]
10. Al Ashaal, H.A.; Farghaly, A.A.; Abd El Aziz, M.M.; Ali, M.A. Phytochemical investigation and medicinal evaluation of fixed oil of *Balanites aegyptiaca* fruits (Balantiaceae). *J. Ethnopharmacol.* **2010**, *127*, 495–501. [CrossRef]
11. Mohamed, A.M.; Wolf, W.; Spiess, W.E.L. Physical, morphological and chemical characteristics, oil recovery and fatty acid composition of *Balanites aegyptiaca* Del. kernels. *Plant Foods Hum. Nutr.* **2002**, *57*, 179–189. [CrossRef]
12. Saini, M.K.; Sharma, P.; Prasad, J.; Kothari, S.L.; Gour, V.S. Quality assessment of oil and biodiesel derived from *Balanites aegyptiaca* collected from different regions of Rajasthan. *Biocatal. Agric. Biotechnol.* **2019**, *22*, 101374. [CrossRef]
13. Jauro, A.; Adams, M.H. Production and Biodegradability of Biodiesel from *Balanites aegyptiaca* Seed Oil. *J. Korean Chem. Soc.* **2011**, *55*, 680–684. [CrossRef]
14. Kaoke, D.F.; Siryabe, E.; Iya-Sou, D.; Talla, E.; Kouotou, P.M. Physicochemical Potential of *Balanites aegyptiaca* Seed Kernel Oil from Northern Cameroon for Biodiesel valorization. *J. Energy Res. Rev.* **2021**, *9*, 21–31. [CrossRef]
15. Bambara, L.; Sawadogo, M.; Roy, D.; Anciaux, D.; Blin, J.; Ouiminga, S. Biofuel from *Balanites aegyptiaca*: Optimization of the Feedstock Supply Chain. *Sustainability* **2018**, *10*, 4501. [CrossRef]
16. Jamil, F.; Al-Riyami, M.; Al-Haj, L.; Al-Muhtaseb, A.H.; Myint, M.T.Z.; Baawain, M.; Al-Abri, M. Waste *Balanites aegyptiaca* seed oil as a potential source for biodiesel production in the presence of a novel mixed metallic oxide catalyst. *Int. J. Energy Res.* **2021**, *45*, 17189–17202. [CrossRef]
17. Huntington, J.L.; Hegewisch, K.C.; Daudert, B.; Morton, C.G.; Abatzoglou, J.T.; McEvoy, D.J.; Erickson, T. Climate Engine: Cloud Computing and Visualization of Climate and Remote Sensing Data for Advanced Natural Resource Monitoring and Process Understanding. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2397–2410. [CrossRef]
18. DGF. *Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten, Tensiden und Verwandten Stoffen*; Wissenschaftliche Verlagsgesellschaft: Stuttgart, Germany, 1998.
19. Li, C.; Yao, Y.; Zhao, G.; Cheng, W.; Liu, H.; Liu, C.; Shi, Z.; Chen, Y.; Wang, S. Comparison and analysis of fatty acids, sterols, and tocopherols in eight vegetable oils. *J. Agric. Food Chem.* **2011**, *59*, 12493–12498. [CrossRef] [PubMed]
20. Zhang, L.; Li, P.; Sun, X.; Wang, X.; Xu, B.; Wang, X.; Ma, F.; Zhang, Q.; Ding, X. Classification and adulteration detection of vegetable oils based on fatty acid profiles. *J. Agric. Food Chem.* **2014**, *62*, 8745–8751. [CrossRef]
21. Elbadawi, S.M.; Ahmad, E.E.; Mariod, A.A.; Mathäus, B. Effects of thermal processing on physicochemical properties and oxidative stability of *Balanities aegyptiaca* kernels and extracted oil. *Grasas Aceites* **2017**, *68*, 184. [CrossRef]
22. Mehanni, A.E.S.; El-Reffaei, W.H.M.; Melo, A.; Casal, S.; Ferreira, I.M. Enzymatic Extraction of Oil from *Balanites aegyptiaca* (Desert Date) Kernel and Comparison with Solvent Extracted Oil. *J. Food Biochem.* **2017**, *41*, e12270. [CrossRef]
23. Okia, C. *Balanites aegyptiaca*: A Resource for Improving Nutrition and Income of Dryland Communities in Uganda. Ph.D. Thesis, Bangor University, Bangor, UK, 2010.
24. Diedhiou, D.; Faye, M.; Candy, L.; Vandenbossche, V.; Vilarem, G.E.; Sock, O.; Rigal, L. Composition and balance of the analytical fractionation of desert date (*Balanites aegyptiaca* L.) seeds harvested in Senegal. *Afr. J. Biotechnol.* **2021**, *20*, 150–158.
25. Gardette, J.-L.; Baba, M. FTIR and DSC studies of the thermal and photochemical stability of *Balanites aegyptiaca* oil (Toogga oil). *Chem. Phys. Lipids* **2013**, *170–171*, 1–7. [CrossRef] [PubMed]
26. Zhang, Y.; Li, X.; Lu, X.; Sun, H.; Wang, F. Effect of oilseed roasting on the quality, flavor and safety of oil: A comprehensive review. *Food Res. Int.* **2021**, *150*, 110791. [CrossRef] [PubMed]
27. Chbani, M.; El Harkaoui, S.; Willenberg, I.; Matthäus, B. Review: Analytical Extraction Methods, Physicochemical Properties and Chemical Composition of Cactus (*Opuntia ficus-indica*) Seed Oil and Its Biological Activity. *Food Rev. Int.* **2023**, *39*, 4496–4512. [CrossRef]
28. Hajib, A.; El Harkaoui, S.; Choukri, H.; Khouchlaa, A.; Aourabi, S.; El Menyiy, N.; Bouyahya, A.; Matthaeus, B. Apiaceae Family an Important Source of Petroselinic Fatty Acid: Abundance, Biosynthesis, Chemistry, and Biological Proprieties. *Biomolecules* **2023**, *13*, 1675. [CrossRef] [PubMed]
29. Ghaffari, M.; Gholizadeh, A.; Rauf, S.; Shariati, F. Drought-stress induced changes of fatty acid composition affecting sunflower grain yield and oil quality. *Food Sci. Nutr.* **2023**, *11*, 7718–7731. [CrossRef]
30. Porokhovinova, E.A.; Matveeva, T.V.; Khafizova, G.V.; Bemova, V.D.; Doubovskaya, A.G.; Kishlyan, N.V.; Podolnaya, L.P.; Gavrilova, V.A. Fatty acid composition of oil crops: Genetics and genetic engineering. *Genet. Resour. Crop Evol.* **2022**, *69*, 2029–2045. [CrossRef]
31. Nguyen, Q.-H.; Talou, T.; Evon, P.; Cerny, M.; Merah, O. Fatty acid composition and oil content during coriander fruit development. *Food Chem.* **2020**, *326*, 127034. [CrossRef]
32. Sidorov, R.A.; Tsydendambaev, V.D. Biosynthesis of fatty oils in higher plants. *Russ. J. Plant Physiol.* **2014**, *61*, 1–18. [CrossRef]
33. Wei, W.; Sun, C.; Jiang, W.; Zhang, X.; Hong, Y.; Jin, Q.; Tao, G.; Wang, X.; Yang, Z. Triacylglycerols fingerprint of edible vegetable oils by ultra-performance liquid chromatography-Q-ToF-MS. *LWT* **2019**, *112*, 108261. [CrossRef]
34. Indelicato, S.; Bongiorno, D.; Pitonzo, R.; Di Stefano, V.; Calabrese, V.; Indelicato, S.; Avellone, G. Triacylglycerols in edible oils: Determination, characterization, quantitation, chemometric approach and evaluation of adulterations. *J. Chromatogr. A* **2017**, *1515*, 1–16. [CrossRef]
35. Stefanoudaki, E.; Williams, M.; Chartzoulakis, K.; Harwood, J. Effect of irrigation on quality attributes of olive oil. *J. Agric. Food Chem.* **2009**, *57*, 7048–7055. [CrossRef] [PubMed]

36. El Harkaoui, S.; Gharby, S.; Kartah, B.; El Monfalouti, H.; El-sayed, M.E.; Abdin, M.; Salama, M.A.; Charrouf, Z.; Matthäus, B. Lipid profile, volatile compounds and oxidative stability during the storage of Moroccan Opuntia ficus-indica seed oil. *Grasas Aceites* **2023**, *74*, e486. [CrossRef]
37. Gharby, S.; Charrouf, Z. Argan Oil: Chemical Composition, Extraction Process, and Quality Control. *Front. Nutr.* **2021**, *8*, 804587. [CrossRef] [PubMed]
38. Jimenez-Lopez, C.; Carpena, M.; Lourenço-Lopes, C.; Gallardo-Gomez, M.; Lorenzo, J.M.; Barba, F.J.; Prieto, M.A.; Simal-Gandara, J. Bioactive Compounds and Quality of Extra Virgin Olive Oil. *Foods* **2020**, *9*, 1014. [CrossRef] [PubMed]
39. Brigelius-Flohé, R.; Kelly, F.J.; Salonen, J.T.; Neuzil, J.; Zingg, J.-M.; Azzi, A. The European perspective on vitamin E: Current knowledge and future research. *Am. J. Clin. Nutr.* **2002**, *76*, 703–716. [CrossRef] [PubMed]
40. Kodad, O.; Socias i Company, R.; Alonso, J.M. Genotypic and Environmental Effects on Tocopherol Content in Almond. *Antioxidants* **2018**, *7*, 6. [CrossRef] [PubMed]
41. Carrera, C.S.; Seguin, P. Factors Affecting Tocopherol Concentrations in Soybean Seeds. *J. Agric. Food Chem.* **2016**, *64*, 9465–9474. [CrossRef]
42. Romero, M.P.; Tovar, M.J.; Ramo, T.; Motilva, M.J. Effect of crop season on the composition of virgin olive oil with protected designation of origin "Les garrigues". *J. Am. Oil Chem. Soc.* **2003**, *80*, 423–430. [CrossRef]
43. Chaudhary, N.; Khurana, P. Vitamin E biosynthesis genes in rice: Molecular characterization, expression profiling and comparative phylogenetic analysis. *Plant Sci.* **2009**, *177*, 479–491. [CrossRef]
44. Sattler, S.E.; Gilliland, L.U.; Magallanes-Lundback, M.; Pollard, M.; DellaPenna, D. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell* **2004**, *16*, 1419–1432. [CrossRef]
45. Lushchak, V.I.; Semchuk, N.M. Tocopherol biosynthesis: Chemistry, regulation and effects of environmental factors. *Acta Physiol. Plant* **2012**, *34*, 1607–1628. [CrossRef]
46. Aremu, M.O.; Andrew, C.; Oko, O.J.; Odoh, R.; Zando, C.; Usman, A.; Akpomie, T. Comparative Studies on the Physicochemical Characteristics and Lipid Contents of Desert Date (*Balanites aegyptiaca* (L.) Del) Kernel and Pulp Oils. *Eur. J. Nutr. Food Saf.* **2022**, *14*, 20–30. [CrossRef]
47. Khan, Z.; Nath, N.; Rauf, A.; Emran, T.B.; Mitra, S.; Islam, F.; Chandran, D.; Barua, J.; Khandaker, M.U.; Idris, A.M.; et al. Multifunctional roles and pharmacological potential of β-sitosterol: Emerging evidence toward clinical applications. *Chem. Biol. Interact.* **2022**, *365*, 110117. [CrossRef] [PubMed]
48. Kirkhus, B.; Lundon, A.R.; Haugen, J.-E.; Vogt, G.; Borge, G.I.A.; Henriksen, B.I.F. Effects of environmental factors on edible oil quality of organically grown *Camelina sativa*. *J. Agric. Food Chem.* **2013**, *61*, 3179–3185. [CrossRef]
49. Roche, J.; Bouniols, A.; Mouloungui, Z.; Barranco, T.; Cerny, M. Management of environmental crop conditions to produce useful sunflower oil components. *Eur. J. Lipid Sci. Technol.* **2006**, *108*, 287–297. [CrossRef]
50. Kumar, M.S.S.; Mawlong, I.; Ali, K.; Tyagi, A. Regulation of phytosterol biosynthetic pathway during drought stress in rice. *Plant Physiol. Biochem.* **2018**, *129*, 11–20. [CrossRef] [PubMed]
51. Rogowska, A.; Szakiel, A. The role of sterols in plant response to abiotic stress. *Phytochem. Rev.* **2020**, *19*, 1525–1538. [CrossRef]
52. Yamaya, A.; Endo, Y.; Fujimoto, K.; Kitamura, K. Effects of genetic variability and planting location on the phytosterol content and composition in soybean seeds. *Food Chem.* **2007**, *102*, 1071–1075. [CrossRef]
53. Roche, J.; Alignan, M.; Bouniols, A.; Cerny, M.; Mouloungui, Z.; Vear, F.; Merah, O. Sterol content in sunflower seeds (*Helianthus annuus* L.) as affected by genotypes and environmental conditions. *Food Chem.* **2010**, *121*, 990–995. [CrossRef]

Manuscript II

Effect of seed's geographical origin on cactus oil physico-chemical characteristics, oxidative stability, and antioxidant activity

Issmail Nounah*, Said El Harkaoui*, Ahmed Hajib, Said Gharby, Hicham Harhar, Abdelhakim Bouyahya, Giovanni Caprioli, Filippo Maggi, Bertrand Matthäus, Zoubida Charrouf

Food Chemistry: X 2024, 22,101445

Accepted manuscript available via https://doi.org/10.1016/j.fochx.2024.101445

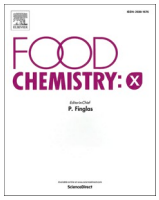
*Equal Contribution

Food Chemistry: X 22 (2024) 101445

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Effect of seed's geographical origin on cactus oil physico-chemical characteristics, oxidative stability, and antioxidant activity

Issmail Nounah [a,1], Said El Harkaoui [b,1], Ahmed Hajib [c], Said Gharby [d], Hicham Harhar [e], Abdelhakim Bouyahya [f,**], Giovanni Caprioli [g,*], Filippo Maggi [g], Bertrand Matthäus [b], Zoubida Charrouf [a]

[a] Department of Chemistry, Faculty of Sciences, Mohammed V University in Rabat, BP 1014 Rabat, Morocco
[b] Max Rubner-Institut, Federal Research Insitute for Nutrition and Food, Department of Safety and Quality of Cereals, Schützenberg 12, D-32756 Detmold, Germany
[c] Higher School of Education and Training (ESEF), Université Ibn Zohr, Agadir, Morocco
[d] Biotechnology Analytical Sciences and Quality Control Team, Polydisciplinary Faculty of Taroudant, Université Ibn Zohr, Morocco
[e] Laboratory of Materials, Nanotechnology and Environment LMNE, Faculty of Sciences, Mohammed V University in Rabat, BP 1014 Rabat, Morocco
[f] Laboratory of Human Pathologies and Biology, Faculty of Sciences, Mohammed V University in Rabat, Morocco
[g] Chemistry Interdisciplinary Project (ChIP), School of Pharmacy, University of Camerino, Via Madonna delle Carceri, 62032 Camerino, Italy

ARTICLE INFO

Keywords:
Chemical composition
Geographical origin
Opuntia ficus-indica
Phenolic content
Tocopherols

ABSTRACT

The aim of this study was the valorisation of cactus (or prickly pear, *Opuntia ficus-indica*) seeds growing in six different regions of Morocco. Moisture, proteins, lipids profile, total polyphenols content, oxidative stability, and antioxidant activity were investigated. The Folin-Ciocalteu test highlighted the abundant presence of phenolic compounds (165 to 225 mg EAG/100 g of extract) and a significant antioxidant capacity against DPPH free radicals. The seeds contained protein (7–9.25%) and lipids (2.7–5%). Cactus oil quality indices such as acidity and peroxide value were below 1.2% and 10 mEq.O_2/kg, respectively. GC analysis revealed that linoleic and oleic acid percentages ranged from 57.1 to 63.8%, and 13.5 to 18.7%, respectively. Cactus seed oil was rich in tocopherols (500–680 mg/kg) and phytosterols (8000–11,100 mg/kg) with a predominance of γ-tocopherols and β-sitosterol. Triacylglycerols, fatty acids and sterols composition showed small variation depending on the geographical origin, while the individual tocopherol profile was significantly influenced.

1. Introduction

Cactus (*Opuntia ficus-indica* (L.) Mill.), is a shrub of the Cactaceae family, mainly growing in arid and semi-arid areas (America, the Mediterranean, Africa, the Middle East, Australia and India) (Barbera, Inglese, Pimienta-Barrios, & Arias-Jiménez, 1995). This species is of great agronomic importance, both for edible fruit and cladodes, which can be used as fodder or vegetable (Mulas & Mulas, 2004). Furthermore, cactus flowers are a source of nutrients highly valued by bees, hence the possibility of developing beekeeping (Arba, 2009). The seeds are rich in minerals, predominating phosphorus (152 mg/100 g) and potassium (163 mg/100 g). The seeds also contain important amounts of magnesium (74.8 mg/100 g), sodium (67.6 mg/100 g), and calcium (16.2 mg/100 g) (Chavez-Santoscoy, Gutierrez-Uribe, & Serna-Saldívar, 2009; El Kossori, Villaume, El Boustani, Sauvaire, & Méjean, 1998). In addition, they are rich in phenolic compounds (268.4 mg/100 g) and also contain 6.0% of protein and 5.5% of oil (Tlili, Sakouhi, Elfalleh, Triki, & Khaldi, 2011). Cactus seed oil belongs to the oleic-linoleic oils group, which makes it particularly interesting in cosmetic, nutritional and pharmacological fields (Chbani, El Harkaoui, Willenberg, & Matthäus, 2023; Coşkuner & Tekin, 2003; Ramadan & Mörsel, 2003). The oil is also rich in tocopherols and phytosterols known for a wide spectrum of pharmacological activities (Chbani et al., 2023). Alqurashi, Al Masoudi, Hamdi, and Abu Zaid (2022) reported the importance of

* Correspondence to: Giovanni Caprioli, Chemistry Interdisciplinary Project (ChIP) research center, School of Pharmacy, Via Madonna delle Carceri snc, Camerino 62032, Italy.
** Correspondence to: Abdelhakim Bouyahya, Faculté des Sciences, 4 Av. Ibn Battouta, B.P. 1014 RP Rabat, Morocco.
E-mail addresses: h.harhar@um5r.ac.ma (H. Harhar), a.bouyahya@um5r.ac.ma (A. Bouyahya), giovanni.caprioli@unicam.it (G. Caprioli), filippo.maggi@unicam.it (F. Maggi), bertrand.matthaeus@mri.bund.de (B. Matthäus).
[1] Contributed equally.

https://doi.org/10.1016/j.fochx.2024.101445
Received 5 February 2024; Received in revised form 13 April 2024; Accepted 4 May 2024
Available online 7 May 2024

cactus seed oil as antioxidant, antibacterial, antifungal, and anticancer agent (Alqurashi et al., 2022).

Recently, many women's cooperatives have been installed in Moroccan cities (Guelmim, Sidi Ifni, Ait Baha, Rhamna, Al-Hoceima) for the packaging of fresh fruits and their transformation into products intended for human and animal consumption. Various products are made at these units, including cactus jam, canned cladodes fillets, dried flowers, cosmetics, and oil extracted from the seeds (Arba, 2009). These products will serve an important socio-economic role for both farmers and rural populations, and help to achieve sustainable development in rural areas. Cactus is, therefore, an adequate species for sustainable agriculture in Morocco.

In order to contribute to the valorization of Moroccan cactus seed oil, this study showed results concerning the influence of the origin of cactus seeds collected from six main planting sites in Morocco (Hoceima, Bejâad, Rhamna, Ait Baha, Tiznit and Sidi Ifni), on the physicochemical characteristics and chemical composition of cactus seeds and seed oil. In addition, accelerated Rancimat oxidation at 120 °C was used to rapidly assess the effect of geographic origin on the oxidative stability of the cactus seed oil. The results of this study may contribute to developing a national standard for cactus seed oil.

2. Materials and methods

2.1. Plant material, solvents, and reagents

Cactus seeds were collected in June 2017 from six cooperatives producing cactus seed oil located (Table 1, Fig. 1): Bejaad (*32°46′15" N, 6°23′28" W*), Ait Baha (30°4′7.9" N, 9°9′10" W), Rhamna (*32°28′12" N, 7°57′29" W*), Tiznit (29°42′1" N, 9°43′43" W), Hoceima (*35°14′41" N, 3°55′60" W*), and Sidi Ifni (*29°22′45" N, 10°10′17.6" W*).

The extraction was performed by using a KOMET D85 type worm press. The oils were then filtered and stored in brown glass bottles at 4 °C until analysis.

All reagents and solvents used were of analytical grade, except for the mobile phase used for HPLC, which was of chromatographic grade and purchased from VWR international (Darmstadt, Germany).

2.2. Moisture content, specific extinction, peroxide index (PV) and free fatty acids

All moisture content, free fatty acid, peroxide index (PV), and UV spectroscopy (Specific UV absorbance at 232 nm and 270 nm) were determined following the ISO 662:2016, DGF C—V 2 (06), DGF C-VI 6a Part 1 (05) and DGF C-IV 6 (13) (Dgf, 1998) methods, respectively.

2.3. Protein content

The nitrogen content was determined using the Kjeldahl procedure with Gerhardt model Vapodest 20 instrument. A factor of 6.25 was then used to convert the measured nitrogen to protein content expressed as a percentage (g/100 g) (Deutsche Gesellschaft für Fettwissenschaften).

Table 1
Geographical data of cactus seed collection sites.

	Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
Altitude (meters above sea level)	133	680	491	604	252	50
Average temperature (°C)	18.1	17.3	17.6	18.1	19.2	19.2
Rainfall (mm/ Year)	272	438	312	229	158	133

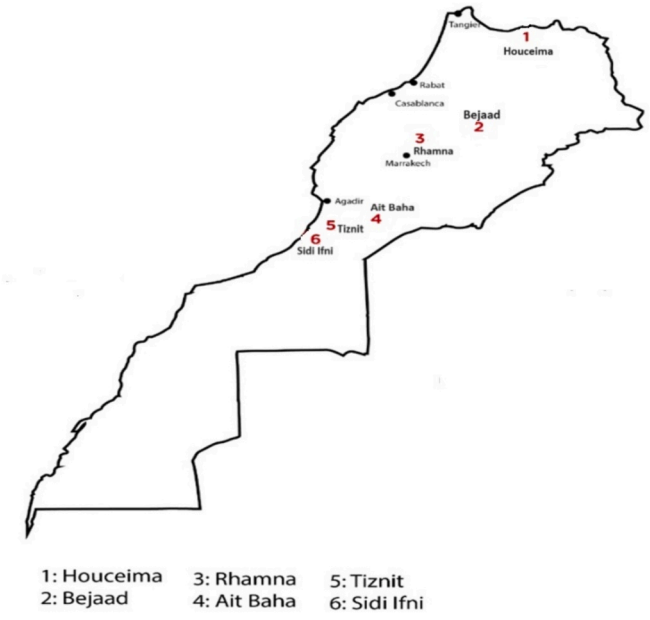


Fig. 1. Location of the evaluated cactus seed sites of production.

2.4. Triacylglycerol composition

The DGF C-VI 14 (08) (Dgf, 1998) method was used to determine the triglyceride composition. The analysis was carried out with an Agilent 6890 gas chromatography system combined with an Agilent 7683B injector (Waldbronn, Germany) and equipped with a flame ionization detector (FID) and a RTX-65 column (30 m × 0.32 mm i.d, 0.1 μm film thickness). The injected volume was 1 μL of a solution of oil in isooctane (0.5 mg/mL). Carrier gas used was hydrogen (flow rate 1 mL/min). The oven temperature was programmed from 300 to 360 (2 °C/min). Injector and detector temperature were set at 380 °C. Triacylglycerols were identified by comparing their retention time, under the same analytical conditions, to that of other oils and data from literature.

2.5. Fatty acid composition

Fatty acids were converted to fatty acid methyl esters (FAMEs), according to DGF C-VI 10a (00) (Dgf, 1998). FAMEs analysis was carried out with an Agilent HP5890 gas chromatography system (Waldbronn, Germany) equipped with a FID and a CP-Sil 88 capillary column (100 m × 250 μm i.d, 0.2 μm film thickness). The hydrogen was used as carrier gas (flow rate: 1 mL/min). Detector and injector temperature was 250 °C. The initial oven temperature of 155 °C was increased to 230 °C on a scale of 1.5 °C/min. The injection volume was 1 μL in a split mode (1:50). The standard mixture of fatty acid methyl esters (Sigma Chemical Co.) was used for the identification of the peaks.

2.6. Phytosterol composition

Phytosterol composition and content were measured according to the DGF F-III 1 (98) method (Dgf, 1998). After a trimethylsilylation of the crude sterol fraction, the composition was carried out using an Agilent 6890 gas chromatography system equipped with a flame ionization detector. The SE 54 CB column capillary column (50 m × 320 μm i.d., 0.25 μm film thickness) was used. The carrier gas used was hydrogen with a flow rate of 1.6 mL/min. The oven temperature was programmed from 245 to 260 °C (5 °C/min). Injector and detector temperature was 320 °C. The injection volume was 1 μL in the split mode (1:20). The results were expressed as relative percentage of the area of each individual sterol peak to the area of all sterol peaks. Cholestan-3-ol was used as an internal standard.

2.7. Tocopherol composition

Tocopherol composition was determined according to the Method DGF F-II 4a (00) (Dgf, 1998). The analysis was conducted using a Merck-Hitachi low-pressure gradient system, equipped with a L-6000 pump, a fluorescence spectrophotometer (Merck-Hitachi F-1000) and a Chem-Station integration system. Twenty µL of a filtrated solution (150 mg/mL of oil in n-heptane) was directly injected onto a Diol phase HPLC column (25 cm × 4.6 mm i.d.). A mixuture of *n*-heptane and tert-butyl methyl ether (99:1, V:V) was used as mobile phase, with a flow rate of 1.3 mL/min. Detector wavelengths were 295 nm for excitation, and 330 nm for emission. Identification was done using α-, β, δ-, and γ-tocopherol reference standards (chromatographic purity 97.6–99.6%, Merck KGaA, Darmstadt, Germany) and quantified through external calibration.

2.8. Oxidative stability

The oxidative stability of the extracted oil (3 g) was evaluated by the Rancimat method using a 743 Rancimat (Methrom AG, Herisau, Switzerland). The heating block was set at 120 °C with an air flow of 20 L/h.

2.9. Total phenol content and antioxidant activity

The total content of phenolic compounds extracted was determined using Folin-Ciocalteu reagent (Bouzid et al., 2023; Yoo, Lee, Park, Lee, & Hwang, 2004)). Briefly, 2.5 mL of diluted Folin-Ciocalteu reagent in water (1:10) and 4 mL of Na_2CO_3 (7.5%, *w/v*) were added to 0.5 mL of sample solution. The mixture was then allowed to stand at 45 °C in a water bath for 30 min and the absorbance measured at 765 nm using a UV–Vis spectrophotometer against a blank sample.

The antioxidant activity of cactus seed extracts was evaluated according to the method described by Scherer et al. (2009) (Scherer & Godoy, 2009). Briefly, 2.5 mL of plant extract was mixed with 0.5 mL of a 0.2 mM solution of DPPH (1.1-diphenyl-2-picrylhydrazyl) in ethanol. The mixuture was allowed to stand at room temperature for 30 min and the absorbance was measured at 517 nm using a UV–Vis spectrophotometer against blank samples.

2.10. Data analysis

Data are presented as means ± standard deviation and analysis of variance was performed with Tukey's test at (95% confidence level) using the software IBM SPSS Statistics 21. Grouped barplot representing the influence of location on oil quality parameters was illustrated using RStudio. Associations between physicochemical parameters of cactus seed oils in this study were performed using the Pearson correlation coefficient (r) with the metan package using RStudio version 1.3.1093.

3. Results and discussion

3.1. Seed analysis

3.1.1. Moisture content

Water content in oil-containing seeds should be monitored with precaution to avoid oxidation risk. The results in Table 2 showed that the moisture content of seeds from different provenances ranges from 4.25 ± 0.1 to 8.97 ± 0.3 g/100 g. The average is around 8% and agrees with the values proposed for the safe storage of oil seeds (Bouzid et al., 2023; Yoo et al., 2004). Our results corroborate those reported in the literature (De Wit, Hugo, & Shongwe, 2017; El Mannoubi, Barrek, Skanji, Casabianca, & Zarrouk, 2009). Furthermore, the results showed an effect of the origin of the seeds on the moisture content, with zones of a high altitude (Ait Baha and Bejaâd) having a lower moisture content than the zones of low altitude (Hoceima, Rhamna, Tiznit and Sidi Ifni). Indeed, cactus seeds are generally dried in the open air under sunshine.

Table 2
Effect of seeds geographical origin on moisture content, total protein, oil yield, total polyphenols content (TPC) and antioxidant power.

	Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
Moisture (%)	8.05 ± 0.2^{a}	4.25 ± 0.1^{b}	8.4 ± 0.3^{a}	7.0 ± 0.1^{c}	8.97 ± 0.3^{d}	8.05 ± 0.3^{a}
Total proteins (%)	8.25 ± 0.08^{a}	7.98 ± 0.27^{ab}	7.78 ± 0.19^{bc}	7.5 ± 0.02^{c}	8.24 ± 0.29^{a}	7.01 ± 0.2^{d}
Oil yield (%)	4.4 ± 0.5^{a}	4.2 ± 0.1^{a}	3.4 ± 0.2^{b}	2.7 ± 0.1^{c}	5.0 ± 0.4^{d}	4.1 ± 0.2^{a}
TPC (mg EAG/ 100 g)	165.6 ± 1.1^{a}	187.0 ± 1.5^{b}	208.5 ± 1.9^{c}	220.7 ± 1.7^{d}	215.5 ± 1.8^{e}	225.9 ± 2.1^{f}
IC_{50} value (mg/mL)	1.03 ± 0.05^{a}	0.77 ± 0.04^{bc}	0.9 ± 0.03^{ab}	0.72 ± 0.3^{bc}	0.53 ± 0.06^{c}	0.58 ± 0.1^{c}

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p < 0.05$).

Farmers judge the dryness of the seeds according to their knowledge before storing them in different polyethylene bags. Therefore, the drying process and storage facilities could explain the recorded moisture values.

3.1.2. Total proteins

The cactus fruit is an important source of biomolecules like proteins (Taoufik et al., 2015). The protein content of cactus seeds according to the geographical origin is presented in Table 2. Our results showed that the seeds protein content varied from 7.01 to 8.25 g/100 g of dry matter. These values were similar to those described by Tlili et al. (2011) (Tlili et al., 2011), but lower then those obtained from Tunisian cactus seeds (17.34 g/100 g) (Albergamo et al., 2022). Geographical origin had a significant influence ($p < 0.05$) on the protein content. The highest protein contents were recorded for the seeds of Hoceima (8.25 g/100 g) and Tiznit (8.24 g/100 g), while Sidi Ifni seeds (7.01 g/100 g) showed the lowest content.

3.1.3. Oil yield

The results in Table 2 show that the oil content varied between 2.7 g/100 g (Ait Baha) and 5.0 g/100 g (Tiznit), suggesting that the origin of the seeds influenced the oil yield. Similar oil yields were reported for 12 cold-pressed South African cactus seeds (2.51 to 5.96 g/100 g) (De Wit, Motsamai, & Hugo, 2021). In contrast, our results were different to those reported by Albergamo et al. (2022) from Tunisian cactus seeds (9.65 g/100 g) (Albergamo et al., 2022).

Different moisture levels (4 to 9 g/100 g) did not affect the oil content of cactus seeds. The same phenomenon was observed for grape seed extraction, where the authors showed that at low moisture levels (5.5 to 7.5 g/100 g) no effect on oil yield was observed (Rombaut et al., 2015). Several factors can influence the yield of cactus seed oil, such as soil and climatic conditions, environmental conditions, and extraction methods (Matthäus & Özcan, 2011; Sakar et al., 2022).

3.1.4. Total phenols content and antioxidant activity

Our results showed that the total polyphenol content in the water/methanol (20/80) extracts varied from 165 to 225 mg EAG/100 g of extract. The highest content was found in Sidi Ifni seed extract (225.9 ± 0.05 mg EAG/100 g of extract), followed by the extracts from Ait Baha (220.7 ± 0.05 mg EAG/100 g), Tiznit (215.5 ± 1.8 mg EAG/100 g) and Rhamna (208.5 ± 0.04 mg EAG/100 g). However, the seed extracts from Bejâad and Hoceima were less rich in polyphenols (187.0 ± 0.03 and 165.6 ± 0.03 mg EAG/100 g, respectively).

Our results were similar to those of Tlili et al. (2011) for Tunisian cactus seeds (222 mg EAG/100 g of extracts) (Tlili et al., 2011). Albano

et al. (2015) worked on two cactus genotypes (purple and orange fruit) from Italy (Albano et al., 2015). They reported a total polyphenol content of 89.2 mg EAG/100 g fresh weight for the purple fruit and 69.8 mg EAG/100 g for the orange fruit. Factors such as cultivar type, climatic conditions and soil composition could explain these variations (Bijla et al., 2021; Gagour et al., 2022; Ibourki et al., 2021; Ibourki et al., 2022). A deep study of the phenolic profile of Moroccan cactus seed oil was done by Chbani et al. (2020). The authors identified 7 polyphenol compounds present in Moroccan cactus seed oil with a predominance of vanillin (3.9 mg/kg – 32.4 mg/kg), syringaldehyde (2.3 mg/kg – 12.3 mg/kg), and ferulaldehyde (2.6 mg/kg – 5.7 mg/kg) in all the studied regions. They concluded that the quantitatve variations between oils from different origins were the result of different climatic, storage or different processing conditions (Chbani et al., 2020).

Statistical analysis of our results showed that the total polyphenol contents in seed extracts depend on the origin of the seeds. A negative correlation between TPC and Rainfall (mm) ($r^2 = -0.6002$) was recorded. We noticed that the concentration of total polyphenols increased significantly from north to south. This can be explained by increasing aridity, inducing stress favouring the appearance of polyphenols.

Polyphenols have significant antioxidant power which was tested using the DPPH test. The results in Table 2 show that there is a significant difference between regions. Tiznit and Sidi Ifni seeds recorded the lowest IC_{50} value (around 0.5 mg/mL). However, the seeds from Hoceima region showed the lowest antioxidant activity ($IC_{50} = 1.03$ mg/mL).

Our results also showed that the IC_{50} is negatively correlated ($r^2 = -0.722$) with the content of phenolic compounds. Indeed, the seeds from Sidi Ifni and Tiznit had significant antioxidant power and a high polyphenol content compared to other origins. Other studies also showed that anti-radical activity is correlated with the level of polyphenols and flavonoids in extracts from medicinal plants (Ait Bouzid et al., 2022; Bijla et al., 2021; Gagour et al., 2022; Ibourki et al., 2022)().

3.2. Oils analysis

3.2.1. Free fatty acids

The acidity, referred to as free fatty acids (FFA), represents the proportion of free fatty acids produced by the hydrolysis of triacylglycerols in the oil. It is a parameter that indicates the quality of the vegetable oil (Gharby et al., 2012; Gharby & Charrouf, 2022). The results in Table S1 and Fig. 2 show that the FFA does not exceed 1.3% (as oleic acid %). Özcan and Al Juhaimi (2011) reported a similar value (1.41%) for Turkish prickly pear seed oil (Özcan & Al Juhaimi, 2011). Our results showed that different cactus seed oils studied (Hoceima, Bejaâd, Rhamna, Ait Baha, Tiznit and Sidi Ifni) respected the Codex Alimentarius, (2019) limit of 2 g of free oleic acid per 100 g of oil) for virgin oils. The very low level of acidity in all samples (<1.3%) indicates no enzymatic or chemical hydrolysis of triacylglycerols and demonstrates the excellent quality of the oils from different locations.

The geographical origin of the seeds had a significant influence ($p < 0.05$) on the FFA of the oil. The average acidity value was 0.5%, with the lowest value of 0.17% for Hoceima oil and the highest value of 1.23% for Sidi Ifni oil.

3.2.2. Peroxide value (PV)

The obtained results showed that all samples had a peroxide value between 3 and 10 mEq.O_2/kg (Table S1 and Fig. 2), which remains below the limit of 20 mEq.O_2/kg according to the Codex Alimentarius Commission (2019). De Wit et al. (2017) obtained a higher peroxide value for South African cactus seed oils varying between 9.50 and 33.67 mEq.O_2/kg (De Wit et al., 2017). Compared to other oils, the relatively high PVs of cactus seed oil can be explained by its high content of unsaturated fatty acids such as linoleic acid, which can lead to limited stability.

The statistical analysis of our results showed that the geographical origin had an influence ($p < 0.05$) on the peroxide value of our oils and an average positive correlation was found between the peroxide value and the moisture content ($r^2 = 0.604$). High humidity promotes the formation of peroxides (Womeni et al., 2003). The regions of Tiznit, Ait Baha and Rhamna recorded the highest values from 8.65 to 9.97 mEq.O_2/kg. The lowest value was obtained in oil from the region of Bejâad (3.51 mEq.O_2/kg).

3.2.3. Specific extinction (K_{232} & K_{270})

The analysis of specific absorbance showed that all the oils studied had K_{232} absorbance values ranging between 2.8 and 4.6 and specific K_{270} values between 0.4 and 0.7 (Table S1 and Fig. 2). Zine, Gharby, and El Hadek (2013) reported lower values for K_{232} (1.72) and K_{270} (0.31) (Zine et al., 2013).

The origin of the seeds influenced the specific coefficients K_{232} and K_{270}. The highest values were recorded for oil from Ait Baha, while the

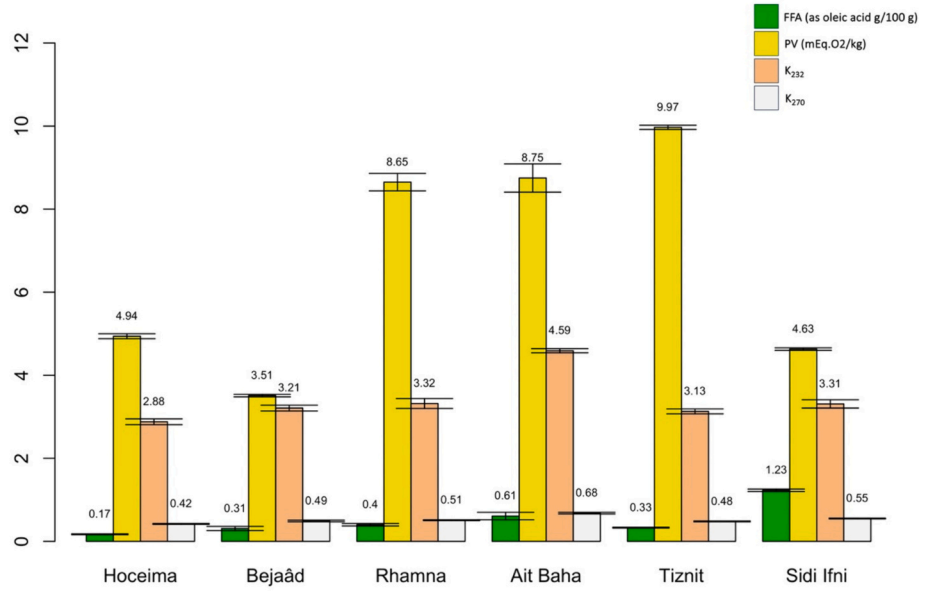


Fig. 2. Effect of seeds geographical origin on acidity (FFA), peroxide value (PV) and specific extinctions (232 and 270 nm).

lower values were observed for oil from Hoceima. The strong and positive correlation between K_{232} and K_{270} ($r^2 = 0.969$, Table S2, Table S3 and Fig. 3) explains the rapid transformation of primary oxidation products into secondary products.

3.2.4. Triacylglycerols composition

The glyceric fraction of cactus seed oil is mainly composed of triacylglycerols (93.37%). Diacylglycerols and monoacylglycerols represent only 2.6 and 0.2%, respectively. Free fatty acids were determined at 1.6%. Eleven triacylglycerols have been identified (Table 3). The three major compounds were palmito-dilinoleic (PLL) (19.9–24.3%), oleo-dilinoleic (LLO) (17.8–19.12%) and trilinoleic (LLL) (17.3–21.28%). Minor triacylglycerols were trioleic (OOO) (0.6–1.6%), palmito-diolein (POO) (1.3–2.7%), and oleo-dipalmitic (POP) (0.8–1.3%). Comparable results were found for the Tunisian cactus seed oil with a predominance of LLL (25%) OLL (21%), and PLL (15%) (El Mannoubi et al., 2009). The results in Table 3 show that the geographic origin of cactus seed oil had little influence on the triacylglycerols composition.

3.2.5. Fatty acids composition

The fatty acid composition of different cactus seed oils was determined after the trans-esterification of fatty acids to methyl esters. The results obtained are reported in Table 4.

The cactus seed oil contains >81% unsaturated fatty acids, regardless of their geographical origin. The main unsaturated fatty acids were linoleic acid (57.1 to 63.8%) and oleic acid (13.5 to 18.7%). Therefore, cactus seed oil belongs to the group of oleic-linoleic oils. The minor unsaturated fatty acids were palmitoleic acid (0.6%), linolenic acid and gadoleic acid at a content of 0.2%. The cactus seed oil also contained 15% saturated fatty acids, mainly palmitic acid (10.1 to 11.4%) and stearic acid (3.1 to 4.1%). The fatty acid contents obtained in this study were similar to those described by Matthäus and Özcan (2011) (Matthäus & Özcan, 2011), who reported a predominance of linoleic acid (49.3 (Kepez) to 62.1% (Hatay-2)), oleic acid (13.0 (Hatay-2) to 23.5% (Kepez)) and palmitic acid (10.6 (Mut) to 12.8% (Kepez)). In another study, Nkoi, Wit, Fouche, Coetzer, and Hugo (2021) found that higher nitrogen fertilization rates significantly increased the oleic and stearic acid content. In contrast, the content of palmitic acid and *cis*-vaccenic acid decreased. However, linoleic fatty acid was not significantly affected (Nkoi et al., 2021). The richness of cactus seed oil in UFA leads to forming the main oxidation products. This finding tends to confirm the previous results with the PV ($r^2 = 0.816$), K_{232} ($r^2 = 0.676$) and K_{270} ($r^2 = 0.714$) indices.

A negative correlation ($r^2 = -0.834$, Table S2 and Fig. 3) was noted between SFA and UFA. An increase in UFA caused a decrease in SFA, which disturbs the stability of the oil due to the increase in double bonds. A negative correlation between SFA and the peroxide value was therefore expected.

The fatty acids in our oils were not significantly different, depending on the geographical origin. This indicates that localization only slightly influenced the composition of fatty acids. These results differ from those reported by Matthäus and Özcan (2011) (Matthäus & Özcan, 2011). They showed that the fatty acid composition of cactus seeds grown in different locations in Turkey differed greatly.

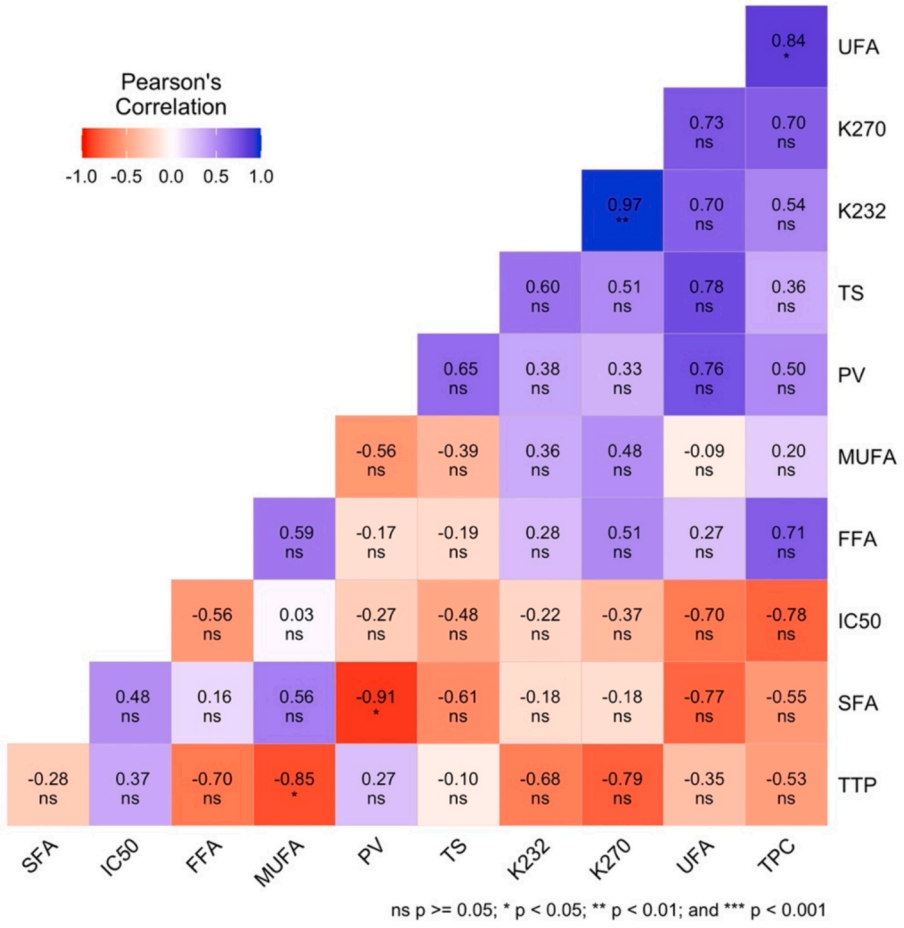


Fig. 3. Pearson's correlation between the variables: FFA, PV, K_{232}, K_{270}, Total sterol (TS), Total tocopherol (TTP), SFA, UFA, Polyphenol, and DPPH (1/IC50 value) of the different samples of cactus oils.

Table 3
Effect of seeds origin on the triglyceride composition of cactus seed oil (%).

	POP	PLP	POO	PLS	PLO	PLL	OOO	SLO	OLO	LLO	LLL
Hoceima	1.1 ± 0.1^{a}	5.5 ± 0.2^{a}	1.9 ± 0.1^{ab}	3.2 ± 0.5^{a}	10.1 ± 1.1^{ab}	24.3 ± 2.7^{a}	0.6 ± 0.1^{a}	2.4 ± 0.2^{a}	3.8 ± 0.5^{a}	17.9 ± 1.1^{a}	19.6 ± 1.9^{ab}
Bejaâd	1.2 ± 0.2^{a}	4.4 ± 0.3^{bd}	2.3 ± 0.8^{b}	3.3 ± 0.4^{a}	10.4 ± 1.3^{ab}	20.9 ± 0.8^{ab}	1.6 ± 0.2^{b}	3.6 ± 0.9^{b}	5.0 ± 1.1^{ab}	18.3 ± 1.6^{a}	17.3 ± 2.3^{a}
Rhamna	1.1 ± 0.5^{a}	5.0 ± 0.1^{abc}	2.1 ± 0.2^{ab}	3.0 ± 0.8^{a}	10.4 ± 1.5^{ab}	22.5 ± 1.8^{ab}	0.8 ± 0.1^{a}	2.8 ± 0.7^{ab}	4.6 ± 0.9^{ab}	18.0 ± 2.6^{a}	18.2 ± 0.9^{a}
Ait Baha	1.1 ± 0.3^{a}	4.2 ± 0.5^{d}	2.1 ± 0.3^{ab}	2.9 ± 0.6^{a}	9.9 ± 0.9^{ab}	19.9 ± 2.1^{b}	1.1 ± 0.8^{ab}	3.8 ± 0.9^{b}	5.2 ± 1.2^{ab}	19.1 ± 1.8^{a}	18.4 ± 1.4^{a}
Tiznit	0.8 ± 0.1^{a}	5.2 ± 0.8^{ac}	1.3 ± 0.1^{a}	2.5 ± 0.1^{a}	8.7 ± 1.1^{a}	22.8 ± 2.5^{ab}	0.8 ± 0.1^{a}	2.8 ± 0.3^{ab}	3.7 ± 0.5a	17.8 ± 1.5^{a}	21.8 ± 2.4^{b}
Sidi Ifni	1.3 ± 0.6^{a}	4.6 ± 0.2^{bcd}	2.7 ± 0.7^{b}	3.2 ± 0.5^{a}	11.4 ± 0.8^{b}	20.8 ± 1.9^{ab}	1.1 ± 0.3^{ab}	3.4 ± 0.4^{ab}	5.6 ± 0.7^{b}	18.7 ± 3.1^{a}	16.6 ± 1.4^{a}

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p < 0.05$). O, oleic acid; L, linoleic acid; P, palmitic acid; S, stearic acid.

Table 4
Effect of seeds origin on fatty acids composition (%) of cactus seed oil.

Fatty acid		Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
Palmitic acid	C16:0	11.4 ± 0.09^{a}	10.7 ± 0.05^{b}	11.1 ± 0.03^{c}	10.18 ± 0.08^{d}	11.4 ± 0.04^{a}	10.9 ± 0.04^{e}
Palmitoleic acid	C16:1	0.6 ± 0.01^{a}	0.6 ± 0.01^{a}	0.6 ± 0.01^{a}	0.6 ± 0.01^{a}	0.5 ± 0.01^{b}	0.6 ± 0.01^{a}
Stearic acid	C18:0	3.6 ± 0.01^{a}	4.1 ± 0.01^{b}	3.4 ± 0.008^{c}	3.7 ± 0.01^{d}	3.1 ± 0.01^{e}	3.8 ± 0.01^{f}
Oleic acid	C18:1Δ9	14.8 ± 0.1^{a}	17.6 ± 0.04^{b}	16.5 ± 0.02^{c}	17.3 ± 0.01^{b}	13.5 ± 0.03^{d}	18.7 ± 0.02^{e}
Elaidic acid	C18:1Δ9	0.3 ± 0.01^{a}	0.3 ± 0.01^{a}	0.3 ± 0.01^{a}	0.2 ± 0.01^{b}	0.1 ± 0.01^{c}	0.2 ± 0.01^{b}
Vaccenic acid	C18:1Δ11	4.9 ± 0.03^{a}	5.2 ± 0.03^{b}	5.0 ± 0.01^{c}	4.9 ± 0.01^{a}	4.3 ± 0.01^{d}	4.8 ± 0.02^{e}
Linoleic acid	C18:2	59.8 ± 1.2^{a}	57.3 ± 0.5^{a}	59.1 ± 2.1^{a}	59.2 ± 1.2^{a}	63.8 ± 3.3^{b}	57.1 ± 1.1^{a}
Linolenic acid	C18:3	0.2 ± 0.01^{a}	0.2 ± 0.02^{a}	0.2 ± 0.01^{a}	0.3 ± 0.03^{a}	0.2 ± 0.01^{a}	0.2 ± 0.01^{a}
Arachidic acid	C20:0	0.4 ± 0.02^{a}	0.4 ± 0.01^{a}	0.4 ± 0.03^{a}	0.4 ± 0.01^{a}	0.4 ± 0.02^{a}	0.4 ± 0.01^{a}
Gadoleic acid	C20:1	0.2 ± 0.01^{a}	0.2 ± 0.02^{a}	0.3 ± 0.03^{a}	0.3 ± 0.01^{a}	0.2 ± 0.06^{a}	0.3 ± 0.05^{a}
Behenic acid	C22:0	0.2 ± 0.04^{a}	0.2 ± 0.02^{a}	0.2 ± 0.01^{a}	0.2 ± 0.03^{a}	0.2 ± 0.01^{a}	0.2 ± 0.05^{a}
SFA	–	15.8 ± 2.3	15.7 ± 1.9	15.4 ± 0.5	15.5 ± 1.6	15.2 ± 0.4	15.7 ± 0.9
UFA	–	81.4 ± 3.5	82.1 ± 1.6	82.5 ± 0.9	83.3 ± 1.3	83.1 ± 1.1	82.4 ± 0.8
MUFA	–	21.0 ± 1.8	24.2 ± 0.4	22.8 ± 1.3	23.5 ± 0.5	18.8 ± 2.1	24.7 ± 0.3

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p < 0.05$). SFA-saturated fatty acids, UFA-unsaturated fatty acids, MUFA monounsaturated fatty acids.

3.2.6. *Phytosterols composition*

The total phytosterol content of the unsaponifiable fraction of cactus seed oil from different sources ranged between 8000 and 11,100 mg/kg (Table 5). The geographic origin of the seeds significantly influenced the amount of total phytosterols. The highest levels were observed for cactus seed oils from Ait Baha (11,091 mg/kg) and Tiznit (10,859 mg/kg), which were significantly different from a second cluster that includes Hoceima (8547 mg/kg), Rhaman (8492 mg/kg).

The sterol fraction of cactus seed oil was mainly composed of β-sitosterol (5973–7142 mg/kg of total phytosterols). Other phytosterols were also detected; campesterol (9–13%), Δ-5-avenasterol (approximately 5%) and sitostanol (3%). The minor phytosterols were Δ-7-avenasterol, Δ-7-Stigmasterol, Δ-5,23-stigmastadienol, Δ-5,24-stigmastadienol and stigmasterol; their proportions did not exceed 4%.

3.2.7. *Tocopherols composition*

The total tocopherol content of cactus seed oil ranged from 500 to 688 mg/kg (Table 6). The geographical origin of the seeds had a significant influence on the composition of tocopherols ($p < 0.05$). Cactus seed oil from Hoceima recorded the highest tocopherol content (687.3 mg/kg), followed by Tiznit oil (679.7 mg/kg), while Ait Baha oil showed the lowest content (502.1 mg/kg).

The major tocopherol was γ-tocopherol, representing on average 92% (445–654 mg/kg) of total tocopherols, followed by α-tocopherol (10–30 mg/kg) and δ-tocopherol (5.4–10.7 mg/kg), while β-tocopherol was not detected. These results agreed with the conclusions of other authors (Matthäus & Özcan, 2011; Ramadan & Mörsel, 2003). Furthermore, the origin of the seeds influenced individual tocopherols. Hoceima and Tiznit contained the highest level of γ-tocopherol (654.5 mg/kg and 649.8 mg/kg, respectively), while Ait Baha oil showed the lowest value (445.3 mg/kg). Studies have revealed a negative correlation between total tocopherol content and both altitude and distance from the coast in olive oil (Mohamed Mousa, Gerasopoulos, Metzidakis, & Kiritsakis, 1996) and argan oil (Elgadi et al., 2021). This phenomenon likely contributes to the elevated tocopherol levels observed in Hoceima. Moreover, water deficit conditions have been documented to enhance

Table 5
Effect of seeds origin on individual and total sterol composition of cactus seed oil.

Phytosterols (mg/kg)	Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
Total sterols	8547 ± 199^{a}	9207 ± 154^{b}	8492 ± 308^{a}	11,091 ± 327^{c}	10,859 ± 306^{c}	8292 ± 11^{a}
Campesterol	837.6 ± 0.7^{a}	892.1 ± 1.6^{a}	811.8 ± 0.9^{a}	1405.2 ± 1.4^{b}	1395.4 ± 0.4^{b}	839.1 ± 0.1^{a}
Stigmasterol	119.6 ± 0.1^{a}	165.7 ± 0.1^{bc}	131.6 ± 0.1^{ab}	222.9 ± 0.2^{c}	223.7 ± 0.1^{c}	145.9 ± 0.4^{abc}
Δ-7-Campesterol	135.9 ± 0.2^{a}	167.7 ± 0.^{2a}	138.4 ± 0.1^{a}	226.2 ± 0.3^{a}	230.2 ± 0.6^{a}	136.8 ± 0.2^{a}
Δ − 5,23-stigmastadienol	99.1 ± 0.4^{a}	116.9 ± 0.1^{a}	106.1 ± 0.3^{a}	133.1 ± 0.1^{a}	152.0 ± 0.2^{a}	97.0 ± 0.5^{a}
β-Sitosterol	6151.3 ± 1.9^{a}	6490.9 ± 2.1^{ab}	6111.7 ± 2.5^{a}	7142.6 ± 3.1^{c}	7114.8 ± 4^{bc}	5973.5 ± 2.9^{a}
Sitostanol	275.2 ± 0.3^{a}	342.5 ± 0.9^{a}	288.7 ± 0.7^{a}	364.9 ± 0.8^{a}	377.9 ± 0.5^{a}	282.7 ± 1.2^{a}
Δ − 5-Avenasterol	391.4 ± 1.2^{a}	462.2 ± 1.2^{a}	360.0 ± 1.9^{a}	473.6 ± 1.1^{a}	384.4 ± 1.3^{a}	339.1 ± 0.2^{a}
Δ − 5,24-stigmastadienol	98.3 ± 0.1^{ab}	118.8 ± 0.2^{b}	85.8 ± 0.1^{ac}	99.8 ± 0.1^{c}	55.4 ± 0.1^{d}	82.1 ± 0.1^{ac}
Δ − 7-Stigmasterol	100.0 ± 0.1^{a}	95.7 ± 0.1^{a}	120.6 ± 0.1^{a}	401.5 ± 0.5^{b}	393.1 ± 0.9^{b}	97.0 ± 0.3^{a}
Δ − 7-Avenasterol	163.2 ± 0.2^{ab}	187.8 ± 0.8^{ab}	181.7 ± 0.6^{ab}	326.1 ± 0.9^{b}	255.2 ± 0.4a^{b}	149.2 ± 0.2^{a}

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p < 0.05$).

Table 6
Effect of seeds origin on individual, total tocopherol composition and oxidative induction time (OIT) of cactus seed oil.

Tocopherols (mg/kg)	Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
Total tocopherols	687.8 ± 3[a]	553.8 ± 2[b]	634.5 ± 2.5[c]	502.1 ± 2[d]	679.7 ± 4[e]	512.8 ± 3[f]
α-Tocopherol	9.9 ± 0.3[a]	23.0 ± 0.5[b]	18.3 ± 0.4[c]	30.4 ± 0.5[d]	10.4 ± 0.3[a]	17.6 ± 0.5[c]
γ-Tocopherol	654.5 ± 3[a]	496.0 ± 2[b]	596.7 ± 2.5[c]	445.3 ± 3.1[d]	649.8 ± 4[a]	473.7 ± 2.5[e]
δ-Tocopherol	10.7 ± 0.9[a]	9.0 ± 0.2[b]	5.4 ± 0.3[c]	5.9 ± 0.3[c]	7.3 ± 0.6[d]	5.9 ± 0.9[c]
OIT (hours at 120 °C)	3.9 ± 0.4[ab]	3.5 ± 0.3[b]	3.0 ± 0.1[a]	2.2 ± 0.1[c]	3.1 ± 0.1[ab]	3.1 ± 0.2[ab]

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p < 0.05$).

tocopherol levels (Carrera & Seguin, 2016). Nonetheless, the determination of significance is complicated by various confounding factors such as sunlight duration, relative humidity, and soil composition. Expanding the sample size may facilitate a more comprehensive understanding of the influence of these factors.

A significant positive correlation of α-tocopherol content with the increase in the level of K_{270} was observed ($r^2 = 0.853$, Table S2, Table S3 and Fig. 3), which could be explained by its pro-oxidant activity in the first stages of the auto-oxidation.

3.2.8. Oxidation induction time (OIT)

Oxidative stability is usually determined under standardized conditions, but accelerated methods can be used. Therefore, the Rancimat test was used for the present study to determine the induction period at 120 °C (393 K) (Table 6).

The oxidative stability of oils was slightly influenced by the geographic origin of the cactus seeds. The oil from Hoceima (3.5 h) and Bejâad (3.4 h) gave the highest OIT values compared to the oil from Ait Baha (2.2 h) at some conditions. The other origins (Rhamna, Tiznit and Sidi Ifni) recorded similar values of around 3.1 h. Nounah et al. (2021) reported that roasting the cactus seeds improved the oxidative stability of the oil. Induction time increased from 3.1 h for non-roasted seeds to 7.6 h after 40 min of roasting. In addition, the stability of cactus seed oil might be enhanced by mixing it with other oils (Nounah et al., 2021). Indeed, blending cactus seed oil with 25% Moringa oil is also a promising way to improve the stability of cactus seed oil (Salama et al., 2020). Gharby et al. (2021) and Taneva et al. (2021) have shown that cactus oil is too sensitive to oxidation compared to argan oil (Gharby et al., 2021; Taneva et al., 2021).

Regarding the sensitivity to oxidation of cactus seed oil due to its high linoleic acid content, special precautions, such as protection against heat and light, should be considered for the prolonged storage of seed cactus oil (Gharby et al., 2012; Gharby et al., 2018; Harhar et al., 2011; Harhar, Gharby, Guillaume, & Charrouf, 2010).

Significant negative correlations were reported between OIT and, PV ($r^2 = -0.6036$), K_{232} ($r^2 = -0.9367$) and K_{270} ($r^2 = -0.9163$). An increase in these values resulted in a decrease in the OIT value. This correlation showed that the OIT values of these oils were highly dependent on the primary and secondary oxidation products (Gharby et al., 2011; Gharby et al., 2023; Hajib et al., 2021).

The low oxidative stability and the variability of OIT values between origins could be explained by the level of tocopherols, phytosterols and unsaturated fatty acids in cactus seed oil. Unsaturated fatty acids are very susceptible to oxidation reactions due to several double bonds; therefore, a negative correlation ($r^2 = -0.7818$, Table S2, Table S3 and Fig. 3) between UFA and the OIT was observed. Table S2 reports a positive correlation of OIT with δ-tocopherol ($r^2 = 0.6492$) and γ-tocopherol ($r^2 = 0.4902$) suggesting an improved oxidative stability in the presence of tocopherols. The action of tocopherols on the oxidative stability of vegetable oils is a complex phenomenon because they are effective antioxidants at low concentrations. Still, they gradually lose their effectiveness as their concentrations in vegetable oils increase (Kamal-Eldin, 2006).

4. Conclusions

The present study shows that cactus seeds are rich in biomolecules (proteins, oil, and polyphenols) that may be appealing for the dermo-cosmetics industry. Indeed, the extracted polyphenols have strong antioxidant potential, suggesting an anti-ageing application. The study of the oil's physicochemical characteristics showed that the quality indices (acidity, peroxide value, and UV-absorbance) recorded slight variations depending on the geographical origin. These variations may be controlled if correct manufacturing practices, including packaging and storage, are followed. Cactus seed oil is an oleic-linoleic type, rich in phytosterols and tocopherols. Its high linoleic acid content suggests important nutritional and dermo-cosmetic properties; however, its oxidation resistance is low. The composition of triacylglycerols, fatty acids and phytosterols showed little variation depending on the geographical origin, while the origin might influence the tocopherol content. Cactus seed oil is the flagship product of the *O. ficus-indica* plant. Consequently, more efforts should be done to ensure its authenticity using quality markers such as phenolic or volatile compounds. Indeed, packaging and oil storage are critical steps in the production chain; therefore, optimization and storage experiments are recommended.

CRediT authorship contribution statement

Issmail Nounah: Writing – original draft, Methodology. **Said El Harkaoui:** Writing – original draft, Methodology. **Ahmed Hajib:** Methodology, Data curation. **Said Gharby:** Validation, Conceptualization. **Hicham Harhar:** Supervision, Data curation. **Abdelhakim Bouyahya:** Writing – review & editing, Supervision. **Giovanni Caprioli:** Writing – review & editing, Supervision, Resources. **Filippo Maggi:** Supervision. **Bertrand Matthäus:** Supervision, Software, Resources. **Zoubida Charrouf:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

The authors grateful thank for the financial support of the research project "*Quality and safety of Moroccan virgin cactus seed oil (Opuntia ficus-indica) form the plant to the bottle*" (FKZ: 01DH17019) financed within German Federal Ministry for Education and Research. The authors would also like to thank the Ministry of High Education Scientific Research, Morocco, for the financial support of the project. We also thank Association Ibn Al Baytar and cooperatives of cactus seed oils for

their support and assistance in this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2024.101445.

References

Ait Bouzid, H., Sakar, E. H., Bijla, L., Ibourki, M., Zeroual, A., Gagour, J., … Gharby, S. (2022). Physical fruit traits, proximate composition, antioxidant activity, and profiling of fatty acids and minerals of wild Jujube (*Ziziphus lotus* L.(Desf.)) fruits from eleven Moroccan origins. *Journal of Food Quality*, 1–15. https://doi.org/10.1155/2022/9362366

Albano, C., Negro, C., Tommasi, N., Gerardi, C., Mita, G., Miceli, A., … Blando, F. (2015). Betalains, phenols and antioxidant capacity in cactus pear [Opuntia ficus-indica (L.) Mill.] fruits from Apulia (South Italy) genotypes. *Antioxidants, 4*, 269–280. https://doi.org/10.3390/antiox4020269

Albergamo, A., Potortí, A. G., Di Bella, G., Amor, N. B., Lo Vecchio, G., Nava, V., … Lo Turco, V. (2022). Chemical characterization of different products from the Tunisian Opuntia ficus-indica (L.) Mill. *Foods, 11*, 155. https://doi.org/10.3390/foods11020155

Alqurashi, A. S., Al Masoudi, L. M., Hamdi, H., & Abu Zaid, A. (2022). Chemical composition and antioxidant, antiviral, antifungal, antibacterial and anticancer potentials of Opuntia ficus-indica seed oil. *Molecules, 27*, 5453. https://doi.org/10.3390/molecules27175453

Arba, M. (2009). Le cactus Opuntia, une espèce fruitière et fourragère pour une agriculture durable au Maroc. *Cana Print Rabat*, 14–16.

Barbera, G., Inglese, P., Pimienta-Barrios, E., & Arias-Jiménez, E. D. J. (1995). *Agro-ecology*. Cultivation and Uses of Cactus Pear.

Bijla, L., Aissa, R., Bouzid, H. A., Sakar, E. H., Ibourki, M., & Gharby, S. (2021). Spent coffee ground oil as a potential alternative for vegetable oil production: Evidence from oil content, lipid profiling, and physicochemical characterization. *Biointerface Research in Applied Chemistry, 12*, 6308–6320. https://doi.org/10.33263/BRIAC125.63086320

Bouzid, H. A., Bijla, L., Ibourki, M., Oubannin, S., Elgadi, S., Koubachi, J., … Gharby, S. (2023). *Ziziphus lotus* (L.) Lam. almonds nutritional potential: Evidence from proximate composition, mineral, antioxidant activity, and lipid profiling reveals a great potential for valorization. *Biomass Conversion and Biorefinery*, 1–15.

Carrera, C. S., & Seguin, P. (2016). Factors affecting tocopherol concentrations in soybean seeds. *Journal of Agricultural and Food Chemistry, 64*, 9465–9474. https://doi.org/10.1021/acs.jafc.6b03902

Chavez-Santoscoy, R. A., Gutierrez-Uribe, J. A., & Serna-Saldívar, S. O. (2009). Phenolic composition, antioxidant capacity and in vitro cancer cell cytotoxicity of nine prickly pear (Opuntia spp.) juices. *Plant Foods for Human Nutrition, 64*, 146–152. https://doi.org/10.1007/s11130-009-0117-0

Chbani, M., El Harkaoui, S., Willenberg, I., & Matthäus, B. (2023). Analytical extraction methods, physicochemical properties and chemical composition of cactus (Opuntia ficus-indica) seed oil and its biological activity. *Food Reviews International, 39*, 4496–4512. https://doi.org/10.1080/87559129.2022.2027437

Chbani, M., Matthäus, B., Charrouf, Z., El Monfalouti, H., Kartah, B., Gharby, S., & Willenberg, I. (2020). Characterization of phenolic compounds extracted from cold pressed cactus (*Opuntia ficus-indica* L.) seed oil and the effect of roasting on their composition. *Foods, 9*, 1098. https://doi.org/10.3390/foods9081098

Coşkuner, Y. N., & Tekin, A. (2003). Monitoring of seed composition of prickly pear (Opuntia ficus-indica L) fruits during maturation period. *Journal of the Science of Food and Agriculture, 83*, 846–849. https://doi.org/10.1002/jsfa.1423

De Wit, M., Hugo, A., & Shongwe, N. (2017). Quality assessment of seed oil from selected cactus pear cultivars (*Opuntia ficus-indica* and *Opuntia robusta*). *Journal of Food Processing and Preservation, 41*, Article e12898. https://doi.org/10.1111/jfpp.12898

De Wit, M., Motsamai, V., & Hugo, A. (2021). Cold-pressed cactus pear seed oil: Quality and stability. *Grasas y Aceites, 72*, e415. https://doi.org/10.3989/gya.0329201

Dgf. (1998). *Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten, Tensiden und verwandten Stoffen*. Wissenschaftliche Verlagsgesellschaft Stuttgart.

El Kossori, R. L., Villaume, C., El Boustani, E., Sauvaire, Y., & Méjean, L. (1998). Composition of pulp, skin and seeds of prickly pears fruit (Opuntia ficus indica sp.). *Plant Foods for Human Nutrition, 52*, 263–270.

El Mannoubi, I., Barrek, S., Skanji, T., Casabianca, H., & Zarrouk, H. (2009). Characterization of Opuntia ficus indica seed oil from Tunisia. *Chemistry of Natural Compounds, 45*, 616–620.

Elgadi, S., Ouhammou, A., Zine, H., Maata, N., Aitlhaj, A., El Allali, H., & El Antari, A. (2021). Discrimination of geographical origin of unroasted kernels Argan oil (Argania spinosa (L.) Skeels) using tocopherols and Chemometrics. *Journal of Food Quality, 2021*, 1–9. https://doi.org/10.1155/2021/8884860

Gagour, J., Ahmed, M. N., Bouzid, H. A., Oubannin, S., Bijla, L., Ibourki, M., … Gharby, S. (2022). Proximate composition, physicochemical, and lipids profiling and elemental profiling of rapeseed (*Brassica napus* L) and sunflower (*Helianthus annuus* L) grown in Morocco. *Evidence-based Complementary and Alternative Medicine, 2022*, Article 3505943. https://doi.org/10.1155/2022/3505943

Gharby, S., & Charrouf, Z. (2022). Argan oil: Chemical composition, extraction process, and quality control. *Frontiers in Nutrition, 8*, Article 804587. https://doi.org/10.3389/fnut.2021.804587

Gharby, S., Guillaume, D., Nounah, I., Harhar, H., Hajib, A., Matthäus, B., & Charrouf, Z. (2021). Shelf-life of Moroccan prickly pear (Opuntia ficus-indica) and argan (Argania spinosa) oils: A comparative study. *Grasas y Aceites, 72*, e397. https://doi.org/10.3989/gya.1147192

Gharby, S., Harhar, H., El Monfalouti, H., Kartah, B., Maata, N., Guillaume, D., & Charrouf, Z. (2012). Chemical and oxidative properties of olive and argan oils sold on the Moroccan market. A comparative study. *Mediterranean Journal of Nutrition and Metabolism, 5*, 31–38. https://doi.org/10.1007/s12349-011-0076-5

Gharby, S., Harhar, H., Farssi, M., Taleb, A. A., Guillaume, D., & Laknifli, A. J. O. (2018). Influence of roasting olive fruit on the chemical composition and polycyclic aromatic hydrocarbon content of olive oil. *OCL, 25*, A303. https://doi.org/10.1051/ocl/2018013

Gharby, S., Harhar, H., Guillaume, D., Haddad, A., Matthäus, B., & Charrouf, Z. (2011). Oxidative stability of edible argan oil: A two-year study. *LWT- Food Science and Technology, 44*, 1–8. https://doi.org/10.1016/j.lwt.2010.07.003

Gharby, S., Kartah, B., El Monfalouti, H., El-sayed, M. E., Abdin, M., Salama, M. A., Charrouf, Z., & Matthäus, B. (2023). Lipid profile, volatile compounds and oxidative stability during the storage of Moroccan Opuntia ficus-indica seed oil. *Grasas y Aceites, 74*, e486. https://doi.org/10.3989/gya.1129212

Hajib, A., Nounah, I., Harhar, H., Gharby, S., Kartah, B., Matthäus, B., Bougrin, K., & Charrouf, Z. (2021). Oil content, lipid profiling and oxidative stability of "Sefri" Moroccan pomegranate (*Punica granatum* L.) seed oil. *OCL, 28*, 5. https://doi.org/10.1051/ocl/2020069

Harhar, H., Gharby, S., Guillaume, D., & Charrouf, Z. (2010). Effect of argan kernel storage conditions on argan oil quality. *European Journal of Lipid Science and Technology, 112*, 915–920. https://doi.org/10.1002/ejlt.200900269

Harhar, H., Gharby, S., Kartah, B., El Monfalouti, H., Guillaume, D., & Charrouf, Z. (2011). Influence of argan kernel roasting-time on virgin argan oil composition and oxidative stability. *Plant Foods for Human Nutrition, 66*, 163–168. https://doi.org/10.1007/s11130-011-0220-x

Ibourki, M., Azouguigh, F., Jadouali, S. M., Sakar, E. H., Bijla, L., Majourhat, K., Gharby, S., & Laknifli, A. (2021). Physical fruit traits, nutritional composition, and seed oil fatty acids profiling in the main date palm (*Phoenix dactylifera* L.) varieties grown in Morocco. *Journal of Food Quality, 2021*, 1–12. https://doi.org/10.1155/2021/5138043-

Ibourki, M., Bouzid, H. A., Bijla, L., Aissa, R., Ainane, T., Gharby, S., & El Hammadi, A. (2022). Physical fruit traits, proximate composition, fatty acid and elemental profiling of almond [*Prunus dulcis* mill. DA Webb] kernels from ten genotypes grown in southern Morocco. *OCL, 29*, 9. https://doi.org/10.1051/ocl/2022002

Kamal-Eldin, A. (2006). Effect of fatty acids and tocopherols on the oxidative stability of vegetable oils. *European Journal of Lipid Science and Technology, 108*, 1051–1061. https://doi.org/10.1002/ejlt.200600090

Matthäus, B., & Özcan, M. M. (2011). Habitat effects on yield, fatty acid composition and tocopherol contents of prickly pear (*Opuntia ficus-indica* L.) seed oils. *Scientia Horticulturae, 131*, 95–98. https://doi.org/10.1016/j.scienta.2011.09.027

Mohamed Mousa, Y., Gerasopoulos, D., Metzidakis, I., & Kiritsakis, A. (1996). Effect of altitude on fruit and oil quality characteristics of 'Mastoides' olives. *Journal of the Science of Food and Agriculture, 71*, 345–350.

Mulas, M., & Mulas, G. (2004). Potentialités d'utilisation stratégique des plantes des genres Atriplex et Opuntia dans la lutte contre la désertification. In *Short and Medium-Term Priority Environnemental Action Programme (SMAP)*. Université des études de SASSAR 112.

Nkoi, V., Wit, M. D., Fouche, H., Coetzer, G., & Hugo, A. (2021). The effect of nitrogen fertilization on the yield, quality and fatty acid composition of Opuntia ficus-indica seed oil. *Sustainability, 13*, Article 10123. https://doi.org/10.3390/su131810123

Nounah, I., Gharby, S., Hajib, A., Harhar, H., Matthäus, B., & Charrouf, Z. (2021). Effect of seeds roasting time on physicochemical properties, oxidative stability, and antioxidant activity of cactus (Opuntia ficus-indica L.) seed oil. *Journal of Food Processing and Preservation, 45*, Article e15747. https://doi.org/10.1111/jfpp.15747

Özcan, M. M., & Al Juhaimi, F. Y. (2011). Nutritive value and chemical composition of prickly pear seeds (Opuntia ficus indica L.) growing in Turkey. *International Journal of Food Sciences and Nutrition, 62*, 533–536. https://doi.org/10.3109/09637486.2011.552569

Ramadan, M. F., & Mörsel, J.-T. (2003). Oil cactus pear (*Opuntia ficus-indica* L.). *Food Chemistry, 82*, 339–345. https://doi.org/10.1016/S0308-8146(02)00550-2

Rombaut, N., Savoire, R., Thomasset, B., Castello, J., Van Hecke, E., & Lanoisellé, J.-L. (2015). Optimization of oil yield and oil total phenolic content during grape seed cold screw pressing. *Industrial Crops and Products, 63*, 26–33. https://doi.org/10.1016/j.indcrop.2014.10.001

Sakar, E. H., Khtira, A., Aalam, Z., Zeroual, A., Gagour, J., & Gharby, S. J. A. (2022). Variations in physicochemical characteristics of olive oil (cv 'Moroccan Picholine') according to extraction technology as revealed by multivariate analysis. *AgriEngineering, 4*, 922–938. https://doi.org/10.3390/agriengineering4040059

Salama, M. A., El Harkaoui, S., Nounah, I., Sakr, H., Abdin, M., Owon, M., Osman, M., Ibrahim, A., Charrouf, Z., & Matthäus, B. (2020). Oxidative stability of Opuntia ficus-indica seeds oil blending with Moringa oleifera seeds oil☆. *OCL, 27*, 53. https://doi.org/10.1051/ocl/2020045

Scherer, R., & Godoy, H. T. (2009). Antioxidant activity index (AAI) by the 2, 2-diphenyl-1-picrylhydrazyl method. *Food Chemistry, 112*, 654–658. https://doi.org/10.1016/j.foodchem.2008.06.026

Taneva, S., Mechqoq, H., Totseva, I., Nikolova, Y., Kamenova-Nacheva, M., Yaagoubi, M. E., … Momchilova, S. (2021). Lipid composition and oxidative stability of Argan and Cactus Opuntia ficus indica seed oils from Morocco-assessment of two extraction methods. *Journal Of Chemical Technology And Metallurgy, 56*, 548–560.

Taoufik, F., Zine, S., El Hadek, M., Idrissi Hassani, L. M., Gharby, S., Harhar, H., & Matthäus, B. (2015). Oil content and main constituents of cactus seed oils Opuntia

Ficus Indica of different origin in Morocco. *Mediterranean Journal of Nutrition and Metabolism, 8*, 85–92.

Tlili, N., Sakouhi, F., Elfalleh, W., Triki, S., & Khaldi, A. (2011). Fatty acids, total phenols and protein of caper, Acacia, Cactus and carob seeds. *Asian Journal of Biotechnology and Research, 2*, 384.

Womeni, H. M., Tchagna, D. T., Ndjouenkeu, R., Kapseu, C., Mbiapo, F. T., Linder, M., Parmentier, M., & Fanni, J. (2003). *Influence des traitements traditionnels des graines et amandes de karité sur la qualité du beurre* (pp. 1–8). Yaoundé, Cameroun: Food Africa: Improving food systems in sub-Saharan Africa: Responding to a changing environment.

Yoo, K. M., Lee, K. W., Park, J. B., Lee, H. J., & Hwang, I. K. (2004). Variation in major antioxidants and total antioxidant activity of Yuzu (Citrus junos Sieb ex Tanaka) during maturation and between cultivars. *Journal of Agricultural and Food Chemistry, 52*, 5907–5913. https://doi.org/10.1021/jf0498158

Zine, S., Gharby, S., & El Hadek, M. (2013). Physicochemical characterization of Opuntia ficus-indica seed oil from Morocco. *Biosciences, Biotechnology Research Asia, 10*.

Manuscript III

Insights into date seed oil composition: Geographical variability and potential applications

Said El Harkaoui, Katharina N'Diaye, Said Gharby, Maryam Al-Hilal, Zoubida Charrouf, Sascha Rohn, Bertrand Matthäus

European Journal of Lipid Science and Technology, 2024, 126, 2400061

Accepted manuscript available via https://doi.org/10.1002/ejlt.202400061

Received: 14 March 2024 | Revised: 24 May 2024 | Accepted: 31 May 2024

DOI: 10.1002/ejlt.202400061

European Journal of Lipid Science and Technology

RESEARCH ARTICLE

Insights into date seed oil composition: Geographical variability and potential applications

Said El Harkaoui[1,2,3] | **Katharina N'Diaye**[1] | **Said Gharby**[4] | **Maryam Al-Hilal**[5] | **Zoubida Charrouf**[6] | **Sascha Rohn**[2] | **Bertrand Matthäus**[1]

[1]Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Detmold, Germany

[2]Department of Food Chemistry and Analysis, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany

[3]Department of Food Technology and Food Material Science, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany

[4]Biotechnology, Analytical Sciences and Quality Control Team, Polydisciplinary Faculty of Taroudant, University Ibn Zohr, Agadir, Morocco

[5]National Center for Health Information, Ministry of Health, Safat, Kuwait

[6]Department of Chemistry, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco

Correspondence
Said El Harkaoui, Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Schützenberg 12, 32756 Detmold, Germany.
Email: Said.Elharkaoui@mri.bund.de

Funding information
Federal Ministry of Food and Agriculture (BMEL), Grant/Award Number: FKZ 2819DOKA03

Abstract

Date fruit (*Phoenix dactylifera* L.) processing generates substantial quantities of date seeds, constituting a potential source of valuable oil yet currently considered waste. This study aimed at characterizing date seed oil (DSO) as a prospective approach for date seed valorization while investigating the influence of geographical origin on its composition. Samples from three Moroccan palm groves (*Allougoum*, *Alnif*, and *Errachidia*) and four non-Moroccan samples were analyzed. Liquid and gas chromatography were applied to determine the chemical composition. The prevalent fatty acids were oleic acid (Ol), ranging from 40.8% to 50.2%, followed by lauric acid (La, 14.0%–24.2%), myristic acid (My, 9.0%–12.6%), palmitic acid (Pa, 9.0%–11.6%), linoleic acid (7.1%–10.7%), and stearic acid (St, 2.4%–4.8%). Major triacylglycerols identified included LaOlLa, LaOlMy, LaOlPa, LaOlOl, MyOlPa, MyOlOl, and PaOlOl. Notably, DSO contains significant tocochromanol (424–760 mg kg^{-1}) and phytosterol content (3422–4827 mg kg^{-1}), with specific phytosterols identified for the first time. Multivariate analysis employing principal component analysis and hierarchical cluster analysis, supplemented by heatmaps, underscored the influence of geographical origin on DSO composition. Moroccan samples exhibited distinctive profiles rich in St, Ol, Pa, PaOlSt, MyOlPa, PaOlPa, γ-tocotrienol, and δ-tocotrienol, leading to a differentiation from non-Moroccan oils. Challenges in classifying DSO within selected Moroccan palm groves were discussed, emphasizing the necessity to consider major and minor compounds, along with date varieties, in geographical effect studies.

Practical Applications: By delving into the chemical properties and potential applications of DSO, this research strives to contribute to sustainable and economically viable approaches for utilizing agricultural waste products.

KEYWORDS
date seed oil, fatty acid, geographical origin, multivariate analysis, *Phoenix dactylifera*, phytosterol, sustainable utilization, tocochromanol, triacylglycerol, valorization

1 | INTRODUCTION

The date palm tree (*Phoenix dactylifera* L.) is widely cultivated in arid and semiarid regions such as the Middle East and North Africa.[1] Especially, it is the most important arboricultural plant in the arid area of Morocco.[2] Its fruits are quite delicious and known for having a high nutritional value and being a source of complex sugars, proteins, and antioxidant compounds, some of them even bearing pharmacological properties.[2] There are a number of date products, such as syrup, jam, and vinegar.[3] However, residual seeds of the date fruits are considered waste and mainly used only as animal feed.[4] The latest FAOSTAT's data on crops and livestock products indicated that date palm production in Morocco increased from 101 351 t in 2010 to 150 301 t in 2021, generating approximately 15 000 t of seeds.[5]

Date seeds of different varieties and clones have different morphological appearances (Figure 1). More or less, date production in Morocco is nowadays dominated by clones from naturally occurring seedlings, locally known as "khalts". Some varieties with big seeds are potentially useful for oil production.[4] Consequently, date seed oil (DSO) extraction should be considered an economic resource with a higher added value that could help in the exploitation of by-products.

Date seeds contain 6%–12% oil, which has a chemical composition interesting with regard to nutritional value and techno-functionality. Many of DSO's biological activities have been already described.[4, 6] Numerous studies showed that the oils are characterized by a high content of saturated and monounsaturated fatty acids (FA), mainly lauric (La), palmitic (Pa), and oleic acids (Ols).[4] This makes it a comparatively stable oil against lipid peroxidation, suggesting its potential use as cooking or frying oil.[4] It is an excellent source of important fat-soluble antioxidants such as carotenoids, tocopherols, and further phenolic compounds, which play an important role in reducing the risk of many diseases.[6] Other potential applications have been listed in some recently published reviews, some of them also covering the non-food sector.[4, 6–9]

Obviously, DSO properties and potential applications are linked to its chemical composition, which varies depending on the date cultivar and geographical origin. In Morocco, 223 date varieties are listed, and many date clones are growing in different groves of the whole country.[2] However, studies on the composition of Moroccan date seeds are very limited, especially with regard to a differentiation of the geographical origin based on the chemical composition of DSO.

The aim of the present study was to characterize the chemical composition of different Moroccan DSO and, subsequently, compare them among themselves and against non-Moroccan samples in order to illustrate how the chemical composition can be related to their geographical origin as well. The samples were investigated in terms of FA, triacylglycerols (TAG), tocochromanols, and phytosterols. Taking chemometric methods into account, DSO samples were grouped or classified according to their geographical origin at hand of the aforementioned analytical parameters. By exploring the chemical properties, this research aims to contribute to sustainable and economically viable approaches for maximizing the value of agricultural by-products.

2 | MATERIALS AND METHODS

2.1 | Material

This study was carried out at hand of 26 date seed samples from three Moroccan provenances: *Allougoum* (Ag), *Alnif* (Al), and *Errachidia*

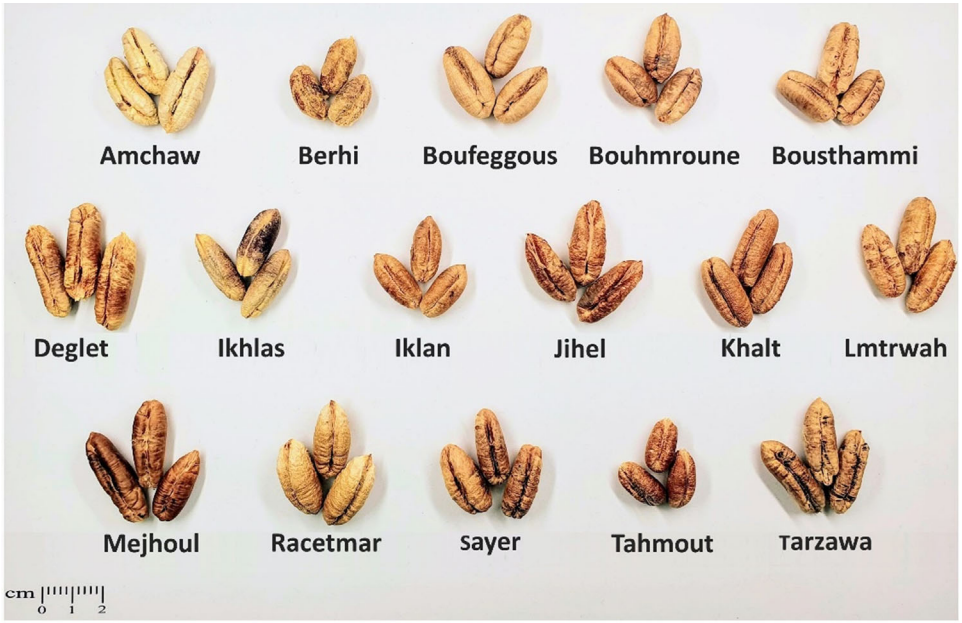


FIGURE 1 Variation in date seed morphology among common date varieties investigated in the present study.

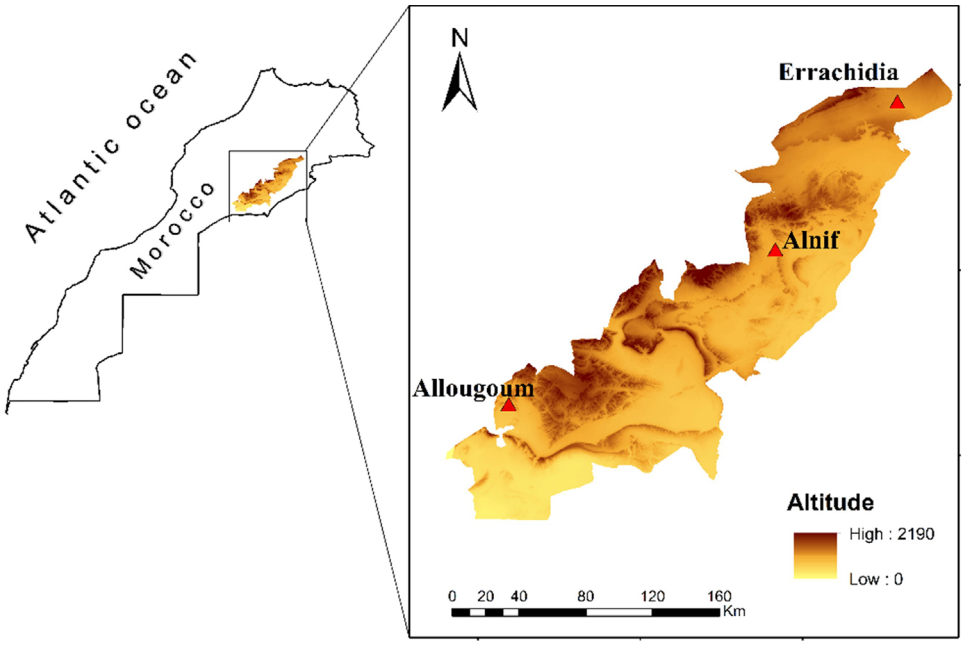


FIGURE 2 Sampling provenances *Allougoum*, *Alnif*, and *Errachidia* (red triangles) with regard to location and altitude. *Source*: Software QGIS 3.14.

TABLE 1 Geographical parameters, temperature, and rainfall of the sampling provenances.

	Latitude	Longitude	Altitude (meters above sea level)	Distance from the Atlantic coast (km)	Temperature min and max (°C)	Rainfall (mm/year)
Allougoum	30°16′31″	06°49′34″	783	271	2.5–44.6	100.4
Errachidia	31°54′49″	04°23′18″	1015	487	1.6–42.6	101.2
Alnif	31°07′07″	05°09′52″	872	444	2.4–43.1	102.7

Note: Temperature and rainfall for the year 2020 were obtained from the database Climate Engine.[57]

(Er). The exact location of the sampling provenances and their altitude, which represent different climatic conditions as well, are mentioned in Figure 2 and Table 1. The varieties of date seeds investigated in this study were "Amchaw", "Berhi", "Boufeggous", "Bouhmroune", "Bousthammi", "Ikhlas", "Iklan", "Jihel", "Khalt", (clone) "Lmtrwah", "Mejhoul", "Racetmar", "Sayer", "Tahmout", and "Tarzawa". In addition, four non-Moroccan varieties, "Deglet Nour", (Algeria) "Berhi", (Kuwait) "Ikhlas", (Saudi Arabia) and "Mejhoul" (Palestine), were purchased from local markets and included for comparative purposes (Figure 1). The complete list of the date seed samples investigated, their names, and abbreviations are also reported in Table S1.

Date fruits were pitted, the seeds cleaned, sun dried, and pre-milled using a Variostuhl roller mill (Bühler AG). Fine-milled using a cross hammer mill (SKB-200, Kvarnaskine, Sweden) equipped with an 800 µm sieve finished raw material preparation. DSO was extracted with petroleum ether for 6 h in a Twisselmann apparatus according to standard Methods described by the *German Society for Fat Science* (DGF B-I 5 (12) and B-II 4a (09)).[10]

2.2 | Determination of the fatty acid composition

The FA composition was analyzed using gas chromatography, following the standard methods DGF C-VI 10a (00) and C-VI 11d (19).[10] To start the analysis, a drop of oil was dissolved in 1 mL *n*-heptane (for liquid chromatography; LiChrosolv®, Supelco®, Merck KGaA) and mixed with 50 µL sodium methylate (30% solution in methanol; Merck KGaA). The mixture was agitated for 1 min at room temperature (22°C), and then 100 µL of Millipore water was added. After centrifuging at 1550 × *g* for 5 min, the lower aqueous phase was removed, and 50 µL 1 M hydrochloric acid (min. 25.0%, p.a; Chemsolute®) was added along with some drops of the indicator methyl orange (ACS reagent, dye content 85%; Merck KGaA). The solution was briefly mixed, and the lower aqueous phase was removed. Twenty milligrams sodium hydrogen sulfate (EMSURE® ACS, ISO, Reag. Ph Eur, Merck KGaA) were added, and after centrifuging at 1550 × *g* for 5 min, the top *n*-heptane phase was transferred to a vial and injected into an HP5890 gas chromatograph (Agilent Technologies Deutschland GmbH). The chromatograph was equipped with a CP-Sil 88 capillary

column (100 m × 0.25 mm × 0.25 µm; Agilent Technologies Deutschland GmbH). The temperature was increased gradually from 150 to 250°C at a rate of 1.5°C min^{-1} and maintained at 250°C for 5 min. The injector and detector were set to 260°C, and the carrier gas (H_2) flow rate was 1.7 mL min^{-1} with a 1:50 split ratio. The detector was operated using 40 mL min^{-1} hydrogen, 400 mL min^{-1} air, and 40 mL min^{-1} nitrogen. The injection volume was 1 µL. The fatty acid methyl esters (FAME) were identified by comparing their retention times with a standard mix (Supelco37 Component FAME Mix; Merck KGaA), and their composition was quantified as a percentage of the total FA.

2.3 | Determination of the triacylglycerol composition

The composition of TAG was analyzed using gas chromatography according to the DGF standard method DGF C-VI 14 (08).[10] The analysis was performed with an Agilent 6890 gas chromatograph equipped with a flame ionization detector (Agilent Technologies Deutschland GmbH). Fifty milligrams of oil were dissolved in 10 mL isooctane (EMSURE, ACS, Reag. Ph Eur; Merck KGaA), and 1 µL was injected into a Restek™ RTX®65TG column (30 m × 0.25 mm × 0.1 µm; Restek Corp.). The oven temperature was held at 300°C for 1 min and then increased from 300 to 360°C at a rate of 2°C min^{-1} and maintained at 360°C for 10 min. The injector and detector were set to 370°C, and the carrier gas (H_2) flow rate was 1 mL min^{-1} with a 1:40 split ratio. The detector was operated using 40 mL min^{-1} hydrogen, 450 mL min^{-1} air, and 45 mL min^{-1} nitrogen. According to the DGF method, TAGs were identified by comparing their retention times with those of sesame oil, which had been previously established for TAG composition analysis. However, in the case of DSO, this approach was insufficient as it contained additional TAG that were not present in the sesame oil. To identify these additional TAG, a Trace GC/Trace DSQ II GC-MS-system (Thermo Scientific Life Technologies GmbH) was used following the method proposed by Ruiz-Samblás et al. [11] with some modifications. The GC injector temperature was set to 320°C. In addition, helium was used as the carrier gas at a flow rate of 1.5 mL min^{-1}, with a 1:15 split ratio and the final temperature was 350°C. The MS parameters were as follows: electron impact source (EI), ionization energy of 70 eV, source temperature of 230°C, and mass range of 50–910 *m/z*. Selected ion monitoring mode was utilized to analyze the ions $[M]^+$, $[RCO]^+$, $[RCO+74]^+$, $[RCO+128]^+$, and $[M-RCO2]^+$, which were the prominent peaks in the EI mass spectra and having the potential for identifying TAG. Commercial TAG standards, tricaprin, trilaurin, trimyristin, and tripalmitin (99%, Merck KGaA), were used for the identification process, as well. After the identification of unknown peaks using GC-MS, GC-FID was used as a routine method to analyze the entire sample set. The quantification values were reported as a percentage of total TAG. The common annotation was used to name the different TAG, involving often used abbreviations of the following FA: caprylic acid (Cy), capric acid (Ca), La, myristic acid (My), Pa, stearic acid (St), Ol, and linoleic acid (Li).

2.4 | Determination of the tocochromanol composition

To determine the tocochromanol content and composition, the standard method DGF F-II 4a (00) [10] was used. Briefly, 150 mg of oil were dissolved in 1 mL of *n*-heptane (for liquid chromatography LiChrosolv, Supelco, Merck KGaA) and filtered using a 1.0 µm filter (Whatmann®) followed by a 0.45 µm filter (Restek Corp.). The filtered solution was then transferred into a vial and afterward injected into the HPLC-FLD. The HPLC system consisted of a pump (L-7100 LaChrom Elite, Merck KGaA), an autosampler (Spark Holland BV), a fluorescence detector (L-2485 LaChrom Elite, Merck KGaA), and the interface 35900E (software CAG BootP Server, Agilent Technologies Inc.). To perform an isocratic separation, a diol phase column (25 cm × 4 mm × 5 µm, LiChroCART® 250-4, Merck KGaA) was utilized. The mobile phase comprised *n*-heptane/*tert*-butyl methyl ether (for liquid chromatography LiChrosolv, Supelco, Merck KGaA) at a flow rate of 1.3 mL min^{-1}. The injection volume for all samples was 20 µL, and the analysis time took 66 min. The fluorescence detector was set to an excitation wavelength of 295 nm and an emission wavelength of 330 nm. The tocochromanols were identified using *α*-, *β*-, *δ*-, and *γ*-tocopherol reference standards (chromatographic purity 97.6%–99.6%, Merck KGaA) and quantified through external calibration with standard solutions (0.25–40 µg mL^{-1}). The calibration curves used for quantification are reported in Table S2.

2.5 | Determination of the phytosterol composition

The phytosterol composition was determined according to the DGF standard method DGF F-III 1 (98).[10] Betulin (98%; Merck KGaA) was prepared as a solution in diisopropyl ether (EMSURE ACS, Reag. Ph Eur; Merck KGaA) at a concentration of 1 mg mL^{-1} and was used as an internal standard for the analysis. After the addition of the internal standard, 250 mg oil were boiled under refluxing with 5 mL ethanolic KOH solution for 15 min. Solid phase extraction on an aluminum oxide column (aluminum oxide 90, Merck KGaA) was used to isolate the unsaponifiable components. In this procedure, the FA anions were retained, whereas the unsaponifiable matter was passed through. The sterol fraction was separated from the unsaponifiable matter through thin-layer chromatography (TLC silica gel 60 plates, 20 cm × 20 cm; Merck KGaA), using a mobile phase of *n*-heptane (for liquid chromatography LiChrosolv, Supelco, Merck KGaA) and distilled diethyl ether (min. 99.5%, stabilized with BHT, Chemsolute) in a 50:50 v/v ratio. After the separation, the sterol fraction was converted into silylated derivatives (TMS) using a silylating agent (*N*-methyl-*N*-trimethylsilyl-heptafluorobutyramid, Macherey-Nagel GmbH & Co. KG). The composition of the sterol fraction was also determined by gas chromatography (HP6890N, Agilent Technologies Deutschland GmbH), with separating the compounds on an SE 54 CB column (50 × 0.25 × 0.1; Macherey-Nagel GmbH & Co. KG). The oven temperature was increased from 245 to 265°C at a rate of 5°C min^{-1}

and maintained at 265°C for 40 min. The injector and detector were set to 320°C, and the carrier gas (H_2) flow rate was 1 mL min^{-1} with a 1:20 split ratio. The detector was operated using 45 mL min^{-1} hydrogen, 450 mL min^{-1} air, and 45 mL min^{-1} nitrogen. The injection volume was 1 µL. Substances were identified by comparing their relative retention times corresponding to betulin to the relative retention times suggested by the DGF method. The assignment of the peaks in the chromatogram was done using the retention times of the individual phytosterols obtained either by comparing to standard compounds (cholesterol, 99%; campesterol, 65%; stigmasterol, 95%; β-sitosterol, 95%; Merck KGaA) or by analyzing them using GC–MS. The quantification was done using the internal standard (betulin).

2.6 | Statistical analysis

All chemical analyses were conducted in triplicate, and the mean values for each set of triplicates were computed for subsequent statistical analysis. To assess the significance of variations among different geographical origins, either a *t*-test or ANOVA followed by the Tukey–Kramer HSD test ($p < 0.05$) were used, employing JMP software (JMP 14.3.0, SAS Institute Inc.). It was opted for parametric tests to assess significant differences as the majority of the variables exhibited a normal distribution (confirmed by the Shapiro–Wilk test and *Q*–*Q* plots) and demonstrated homogeneity of variance (confirmed by Levene's test). Furthermore, principal component analysis (PCA) was performed. Initially, the data underwent log-transformation and Pareto scaling, after which PCA was performed using MetaboAnalyst 5.0.[12]

3 | RESULTS AND DISCUSSION

3.1 | Fatty acid composition

Considering all DSO, oil yield ranged from 8.0 to 10.6 g/100 g, depending on the date seed variety and geographical origin (Table S1).

FA composition of the DSO samples was determined, and the main FAs are reported in Figure 3 (the full list of identified FA is given in Table S3). The prevalent FA was Ol, ranging from 40.8% to 50.2%, followed by La (14.0%–24.2%), My (9.0%–12.6%), Pa (9.0%–11.6%), Li (7.1%–10.7%), and St (2.4%–4.8%) (Figure 3). A small amount (<1%) was determined for the following FA: Cy (8:0), Ca (10:0), margaric acid (17:0), vaccenic acid (18:1Δ11), arachidonic acid (20:0), gondoic acid (20:1Δ11), behenic acid (22:0), and lignoceric acid (24:0) (Table S3). A quite similar FA composition was reported by Ibourki et al. [13], who compared DSO from eight different Moroccan varieties.[13]

Furthermore, unsaturated FA and saturated FA (UFA/SFA) ratios of the analyzed DSO were between 0.9 and 1.5, which is quite similar to the ratio described for palm oil with a usual ratio of 1.0.[14] However, unlike palm oil, the content of SFA in DSO is made up of a range of different FAs, such as La, My, Pa, and St. In contrast, the SFA portion of date palm oil is mostly comprised of Pa.[14, 15] In comparison to other oils, DSO had a lower UFA/SFA ratio than soybean, sunflower, and corn oil, but a higher ratio than coconut oil.[14–16] UFA/SFA ratio and FA composition suggest a comparatively high stability, being highly resistant to oxidative rancidity and thermal treatments but staying liquid at room temperature. These findings support the understanding of the potential application of DSO in culinary and food preservation contexts.

With regard to the effect of geographical origin on the FA composition, present results suggest that Moroccan DSO may contain significantly higher levels of Pa and St but low levels of La when compared to non-Moroccan samples (Figure 3). The variation in the content of these acids could be attributed to differences in varieties and climatic conditions between the Moroccan and non-Moroccan samples. However, it is difficult to make a definitive conclusion based on these findings alone, as comparisons with previous studies on larger sample sets of non-Moroccan DSO also showed contradictory results.[17–19]

When taking a closer look at the variation in FA composition within the Moroccan sample set, no significant differences in the main FA among the samples from different geographical origins were identified (Figure 3).

Rainfall was more or less similar for all three regions studied, so its impact could be assumed to be marginal for the present study. Although the three locations are characterized by different altitudes and distances from the coast, FA composition was the same, and no effects related to altitude and coastal proximity were observed. The relationship between FA composition on the one hand and altitude and climatic conditions on the other seemed to be very complex.[20] Much clearer seems to be the relationship between FA composition and level of maturity and availability of water. However, these factors were constant in the present study, as rainfall was the same in the three Moroccan palm groves and date fruits were collected at the *Tamr* stage. Date fruits are typically harvested during the *Tamr* stage, which marks their final maturation and is characterized by good storability in comparison to earlier stages, such as *Hanabauk*, *Kimri*, *Khalal*, and *Rutab*.[21] Consequently, the seeds generated as biomass ("waste") primarily come from date fruits at the *Tamr* stage. However, if DSO were to become a prominent product and we opted to use lower-quality date fruits exclusively for its extraction, significant consideration should be given to the stage of maturity. As an example of maturity's impact on FA composition, in earlier stages of coriander fruit maturation, palmitic and Sts were the principal FA, yet their concentrations gradually decreased with maturation. Conversely, UFA, mainly petroselinic FA, exhibited an opposing trend, with their concentrations increasing during maturation.[22] This shift was attributed to the activities of specific desaturases that regulate the FA profile during maturation.[22] Similar trends were also reported in other studies.[23–25] On the other hand, several studies highlighted that water deficit significantly reduces the polyunsaturated/saturated ratio, consequently affecting the double bond index of the corresponding oils.[26–29]

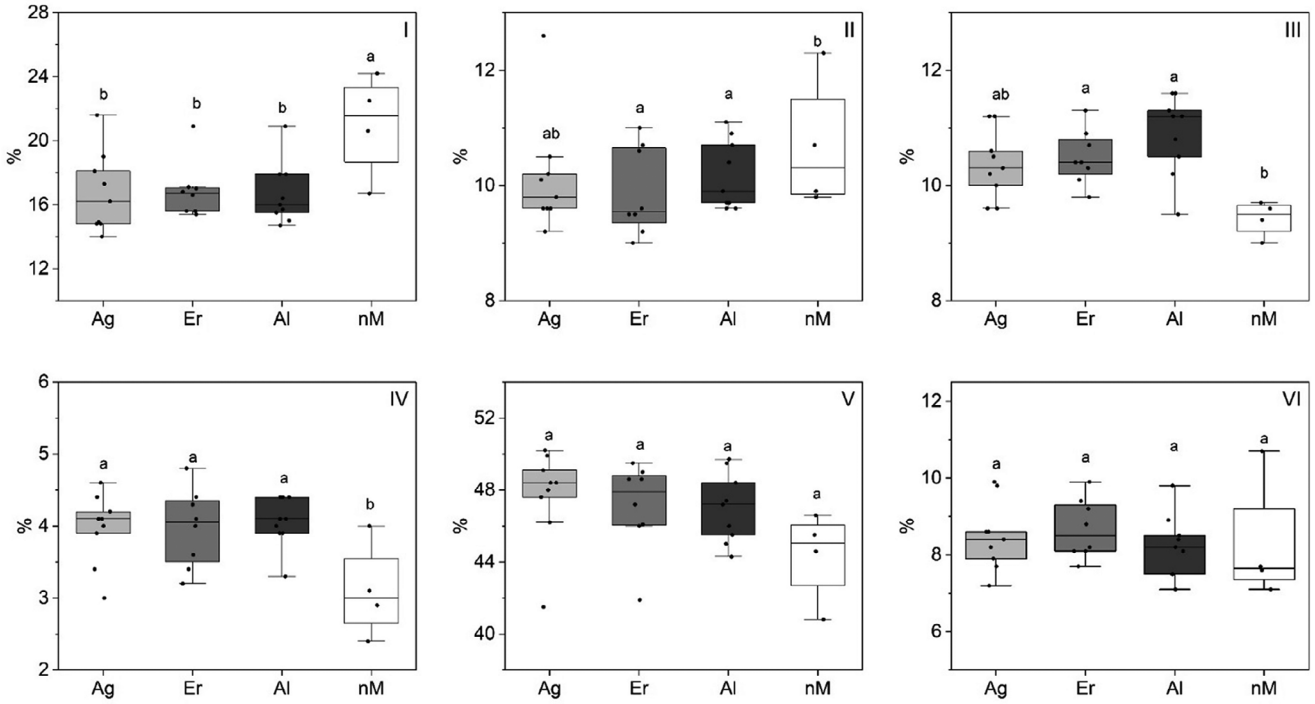


FIGURE 3 Fatty acid composition of date seed oil (DSO) depending on geographic origin. *Allougoum* (Ag), *Errachidia* (Er), *Alnif* (Al), and non-Moroccan (nM); I: lauric acid, II: myristic acid, III: palmitic acid, IV: stearic acid, V: oleic acid, and VI: linoleic acid. Statistical difference was identified by means of a one-way ANOVA test, which was then followed by a Tukey post hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval.

3.2 | Triacylglycerol composition

TAG composition of DSO from the 30 date seed samples is given in Table S4, and the main TAG is reported in Figure 4.

A total of 34 TAGs with a number of carbon atoms ranging from 32 to 54 were identified (Table S4). These accounted for more than 90% of the total area of peaks in the chromatogram. The main TAG identified were LaOlLa, LaOlMy, LaOlPa, LaOlOl, MyOlPa, MyOlOl, and PaOlOl (Figure 4). This confirmed the presence of FA reported above with Ol, La, My, Pa, and Li being the main FA. In terms of predominance according to number of carbon atoms, TAG with 48 C atoms (LaOlOl, PaOlMy, LaLiOl, and PaPaPa), 42 (LaOlLa, LaLiLa, and LaMyPa), 44 (LaOlMy, LaLiMy, and LaPaPa), and 46 C atoms (LaOlPa and LaLiPa) were present as major TAG in all DSO samples. Besbes et al. [30] also quantified TAG in two date seed varieties, "Deglet" and "Allig", from Tunisia and reported similar results with a predominance of a C atom numbers of 48, 42, 44, and 46. [30] In another study, Lieb et al. [31] reported the predominance of 48°C atoms. However, they reported different individual TAG. In that study, main TAG were LaLiOl (9.9%–11.4%), LaLiLa (7.5%–8.4%), LaLiMy (6.8%–8.2%), MyLiMy (6.8%–7.6%), LaLiPa (6.8%–7.6%), and LaOlOl (6.1%–7.0%).[31]

Comparing the content of the main TAG within the three present Moroccan provenances, the following contents were found: for Ag, LaOlLa 9.1%–14.5%, LaOlMy 10.9%–15.2%, LaOlPa 12.3%–13.8%, LaOlOl 8.9%–14.0%; for Al, LaOlLa 8.2%–11.2%, LaOlMy 10.7%–12.9%, LaOlPa 12.6%–14.7%, LaOlOl 11.6%–14.0%; for Er, LaOlLa 9.8%–13.1%, LaOlMy 10.7%–13.3%, LaOlPa 12.3%–14.0%, and LaOlOl 10.7%–14.0%. The results showed that the content ranges are very similar between the provenances with no significant differences identified. The same remark was also valid when comparing the Moroccan to the non-Moroccan DSO (Figure 4). However, there appeared to be some differences according to origin in the less concentrated TAG, such as LaLaLa, LaLaMy, LaMyMy, and LaLiLa (Table S4).

The TAG profile of DSO differed significantly from other common oils, such as coconut, olive, palm, camellia, and sunflower.[32] Unique TAG profiles should be considered a fingerprint for DSO and a strong marker for detecting adulteration.[33]

3.3 | Tocochromanol composition

Five tocochromanol isomers were detected and quantified in the present DSO, and results are shown in Figure 5 and Table S5. DSO contained a significant total content of tocochromanols, ranging from 424 to 760 mg kg^{-1}. The major tocochromanol was α-tocotrienol, regardless of origin and variety. α-Tocotrienol concentration varied from 324 to 563 mg kg^{-1}, followed by β-tocotrienol (46–169 mg kg^{-1}), γ-tocotrienol (14–98 mg kg^{-1}), and δ-tocotrienol (5–30 mg kg^{-1}). α-Tocopherol had the lowest content (0–48 mg kg^{-1}), whereas β-, γ-, and δ-tocopherols were not found in the DSO samples analyzed. A comparable tocochromanol content has also been reported for other

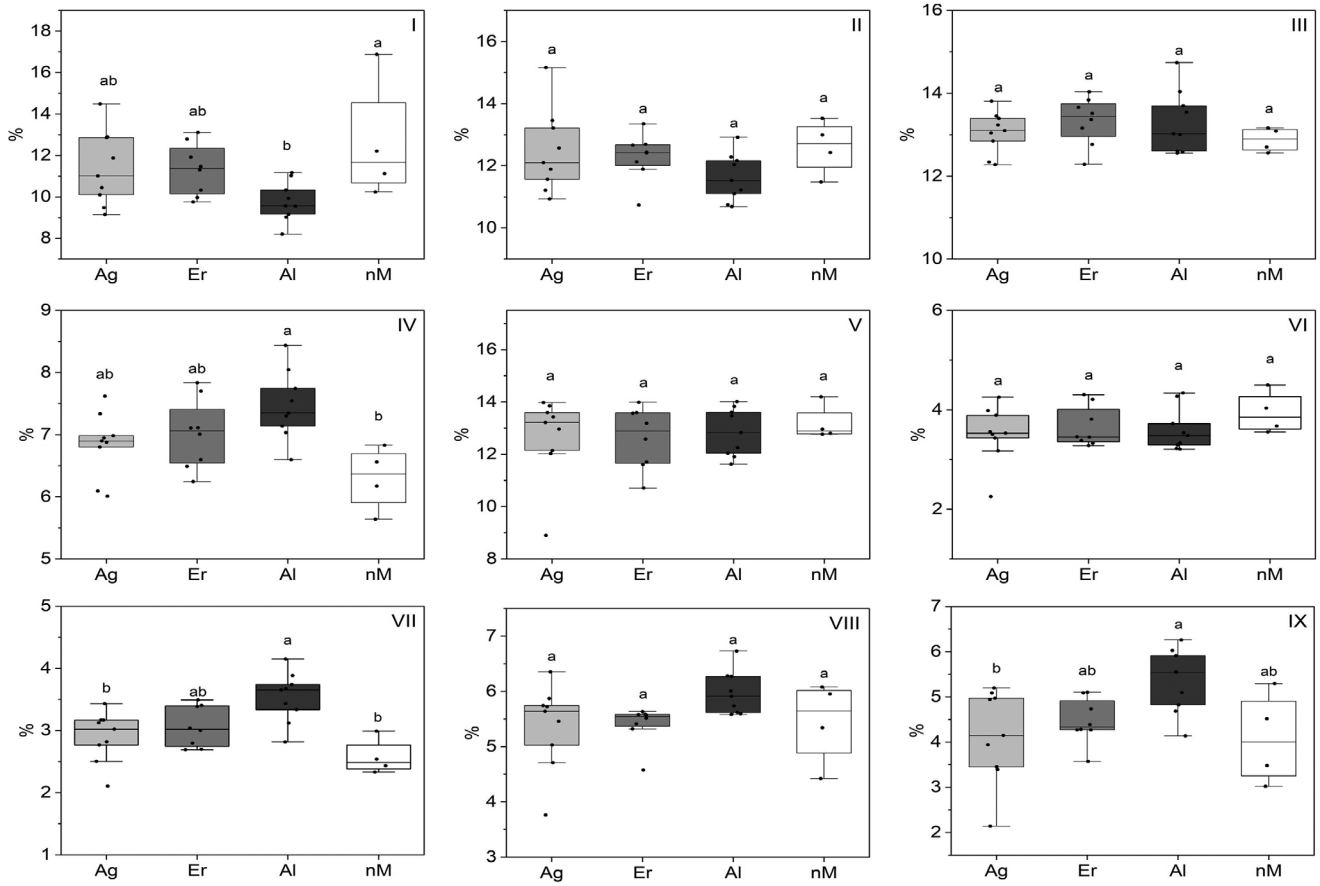


FIGURE 4 Triacylglycerol composition of date seed oil (DSO) depending on geographic origin. *Allougoum* (Ag), *Errachidia* (Er), *Alnif* (Al), and non-Moroccan (nM); I: LaOlLa, II: LaOlMy, III: LaOlPa, IV: MyOlPa, V: LaOlOl, VI: LaLiOl, VII: PaOlPa, VIII: MyOlOl, and IX: PaOlOl. Statistical difference was identified by means of a one-way ANOVA test, which was then followed by a Tukey post hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval. La, lauric acid; My, myristic acid; Pa, palmitic acid; St, stearic acid; Ol, oleic acid; Li, linoleic acid.

"exotic" oils such as argan oil and cactus seed oil.[34–36] The range of the tocochromanol content detected in the present sample set was larger than those described in the literature.[17, 30, 37, 38]

Only one study reported higher values for some Algerian varieties: 946 mg kg^{-1} ("Haloua"), 942 mg kg^{-1} ("Mech-Degla"), 879 mg kg^{-1} ("Deglet-Nour"), and 810 mg kg^{-1} ("Itima").[18] Differences may be attributed to some genetic variation across varieties. In agreement with the present results, α-tocotrienol was the main tocochromanol in six Saudi Arabian varieties with values ranging from 190 to 260 mg kg^{-1}.[38] The predominance of α-tocotrienol was also reported by Al Juhaimi et al. [37] for two Sudanese cultivars (370–330 mg kg^{-1} oil) and two Libyan cultivars (330–340 mg kg^{-1} oil). Another derivative of vitamin E, α-tocopherol acetate, was also already detected in DSO.[17] α-Tocopherol acetate is known for its high stability toward oxygen and light compared to mean tocopherols.[39]

Concerning the effect of geographical origin, no significant differences were observed when comparing the Moroccan samples with the non-Moroccan sample set (Figure 5). On the other hand, it was observed that the samples from Ag were characterized by a lower level of α-tocotrienol (mean value: 369 mg kg^{-1} oil) when compared with samples from Al and Er. As α-tocotrienol is the main tocochromanol in DSO, it could be expected that the samples from Ag had the lowest total tocochromanol amounts.

It has been reported that total tocopherol amount correlates negatively with altitude and distance from the coast for olive oil [40] and argan oil,[41] but in the present study, this hypothesis was not confirmed for DSO. Probably more distinct differences in terms of altitude and distances from the coast are necessary. However, interfering factors such as duration of sunlight, relative humidity, and soil type make it complex to identify significances. Increasing the sample number may also help to clarify the impact of other factors.

3.4 | Phytosterol composition

Twelve phytosterols were identified in the DSO samples. Results are given in Table S6. The total phytosterol content ranged from 3422 to 4827 mg kg^{-1} oil. The main phytosterols were β-sitosterol (2168–3079 mg kg^{-1} oil), Δ5-avenasterol (325–1101 mg kg^{-1} oil), campesterol (307–672 mg kg^{-1} oil), and stigmasterol (114–240 mg kg^{-1}

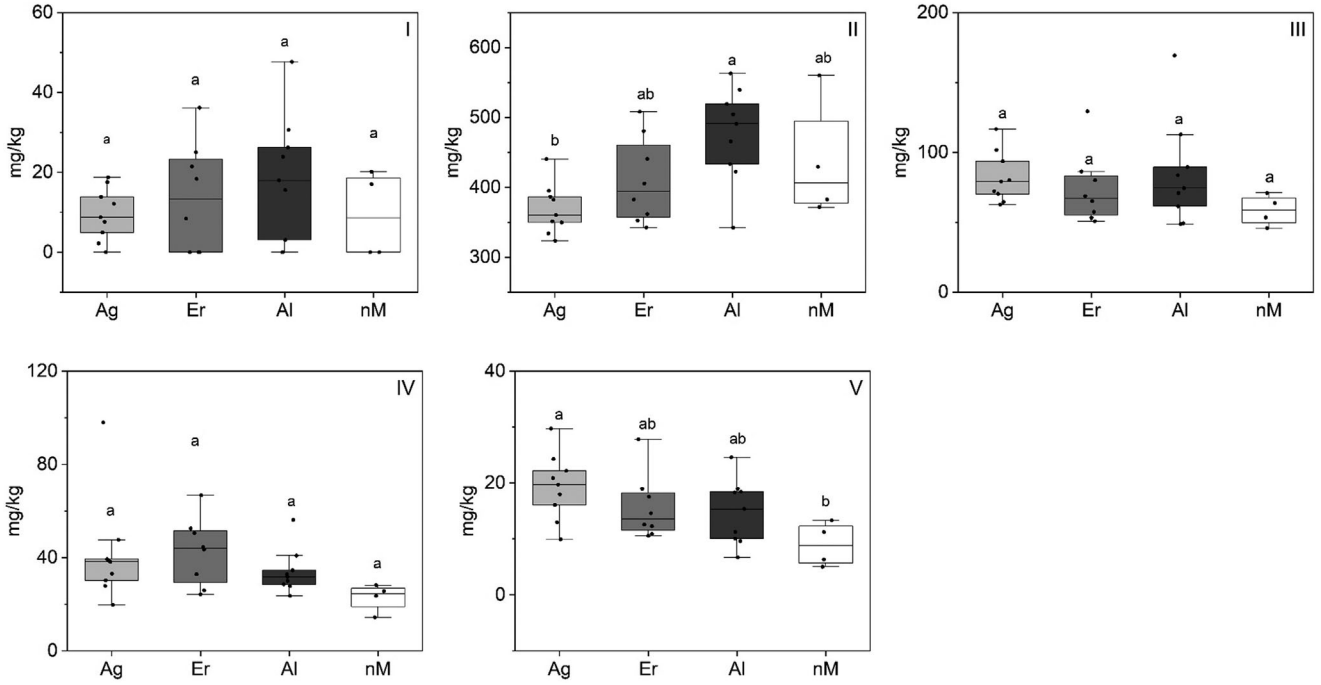


FIGURE 5 Tocochromanol composition of date seed oil (DSO) depending on geographic origin. *Allougoum* (Ag), *Errachidia* (Er), *Alnif* (Al), and non-Moroccan (nM); I: α-tocopherol, II: α-tocotrienol, III: β-tocotrienol, IV: γ-tocotrienol, and V: δ-tocotrienol. Statistical difference was identified by means of a one-way ANOVA test, which was then followed by a Tukey post hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval.

oil). Contents of further phytosterols were less than 200 mg kg^{-1} oil (Figure 6 and Table S5).

The predominance of β-sitosterol is very common in many vegetable oils, such as soybean, corn, sesame, palm, and corn oils.[15] On the other hand, Δ5-avenasterol was the second most phytosterol in all samples analyzed, except for the samples of the fruits "Tarzawa small", "Khalt 2", and "Deglet", where campesterol was the second most phytosterol, with Δ5-avenasterol being the third most. The presence of Δ5-avenasterol in DSO was interesting, as this compound was described as an antioxidant and anti-polymerization agent in frying oils. [42] It seemed to be an erroneous perception that cholesterol is only found in animal products because most plants contain relatively small quantities of cholesterol.[43–45] However, cholesterol was already been reported in camelina oil and other vegetable oils.[46–48] Especially, Sonawane et al. [49] made a significant contribution by elucidating the complete plant cholesterol biosynthesis pathway, comprising 10 enzymatic steps starting from 2,3-oxidosqualene.[49]

To the best of our knowledge, Besbes et al. [50] were the first and only ones to report the phytosterol composition of DSO before.[50] They reported contents of 3000 and 3500 mg kg^{-1} oil, respectively, in "Allig" and "Deglet Nour" seeds from Tunisia, which are comparable to the present results. However, those authors identified only 6 phytosterols in both samples, with a predominance of β-sitosterol, campesterol, Δ5-avenasterol, and stigmasterol, whereas in the present study, 12 different phytosterols in *P. dactylifera* seed oil were identified. The present study seems to be the first to identify 24-methylenecholesterol, campestanol, Δ5,23-stigmastadienol, sitostanol, Δ7-stigmastenol, and Δ7-avenasterol.

Regarding geographical origin, Figure 6 shows that present Moroccan DSO samples displayed slightly lower contents of stigmasterol, 24-methylenecholesterol, and cholesterol than samples of the non-Moroccan set. The mean values of those phytosterols in the non-Moroccan samples were as follows: 203 mg stigmasterol kg^{-1} oil, 70 mg 24-methylenecholesterol kg^{-1} oil, and 54 mg cholesterol kg^{-1} oil. Indeed, some differences were found among samples from the three Moroccan palm groves, especially in the content of β-sitosterol, stigmasterol, Δ5,23-stigmastadienol, and sitostanol. For example, samples from Er had a lower content of β-sitosterol (mean value of 2439 mg kg^{-1} oil) as compared to Ag and Al. Although Ag was characterized by its lower content of stigmasterol (mean value of 129 mg kg^{-1} oil), Al had a higher content of Δ5,23-stigmastadienol (mean value of 46 mg kg^{-1} oil) (Figure 6).

As noted above for DSO's tocochromanol composition, there was a similar difficulty in establishing clear relationships between phytosterol composition and the various geographical and climatic factors that might interfere. Very little research has been conducted on the relationship between phytosterol composition and altitude and climate. Some studies on olive oil and argan oil reported that the phytosterol composition of the oil changes depending on the geographical origin and climatic conditions. For olive oil, it was reported that the content of β-sitosterol increased and those of stigmasterol, campesterol, and Δ5-avenasterol decreased when the altitude was higher and climate became colder.[51,52] In argan oil, a more regular trend

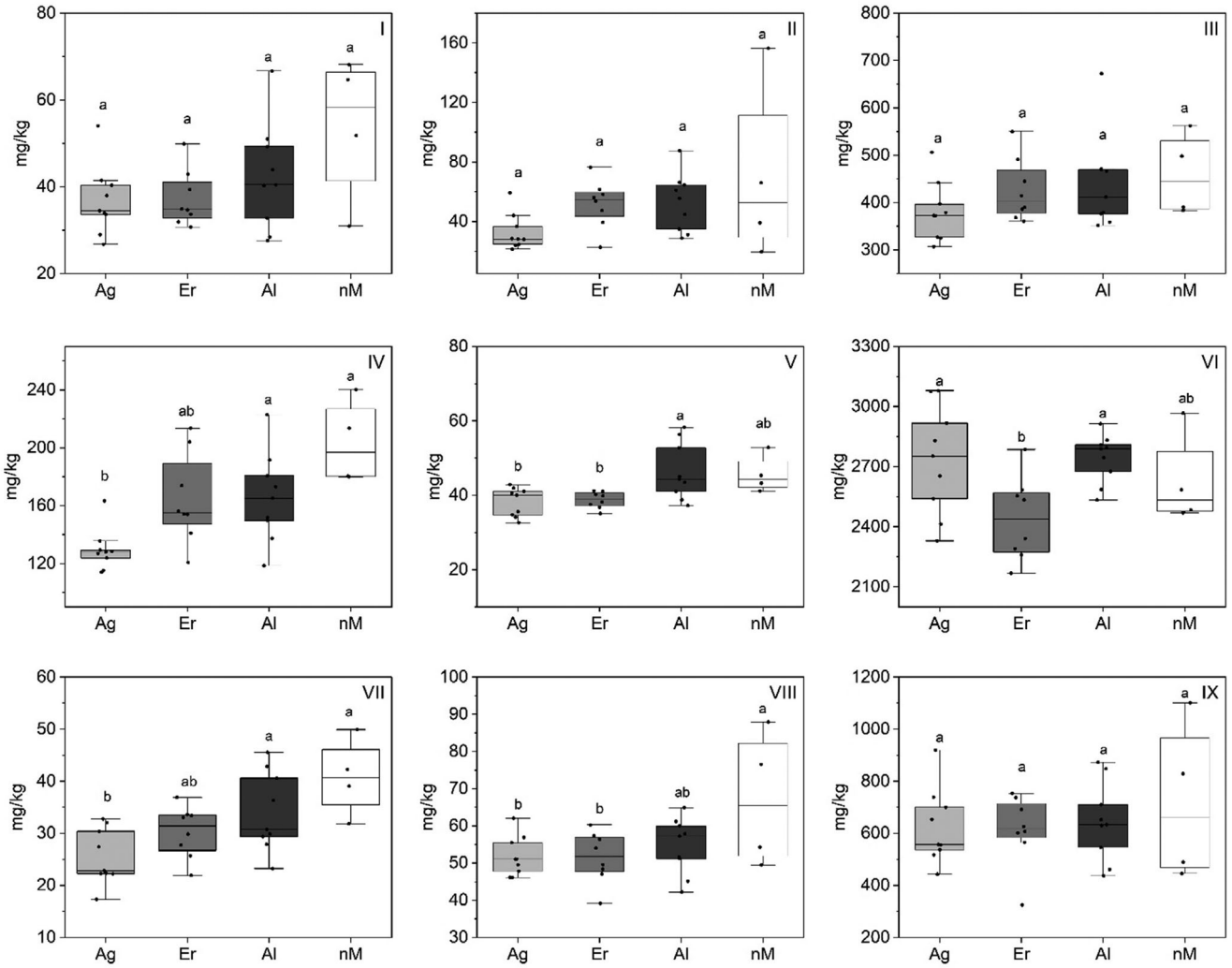


FIGURE 6 Sterol composition of date seed oil (DSO) depending on geographic origin. *Allougoum* (Ag), *Errachidia* (Er), *Alnif* (Al), and non-Moroccan (nM); I: cholesterol, II:24-methylencholesterol, III: campesterol, IV: stigmasterol, V: Δ5,23-stigmastadienol, VI: β-sitosterol, VII: sitostanol, VIII: Δ5,24-stigmastadienol, and IX: Δ5-avenasterol. Statistical difference was identified by means of a one-way ANOVA test, which was then followed by a Tukey post hoc test. Different letters indicated the existence of statistically significant differences with a 95% confidence interval.

was observed for all phytosterols. Kharbach et al. [53] reported that the contents of the main phytosterols, schottenol, spinasterol, and Δ7-avenasterol were higher in the region of Taroudant/Morocco, the highest altitude site, than in other areas of lower altitude.[53] In the present samples, contradictory results were found for DSO, as Ag, the lowest site, had a lower stigmasterol content. Moreover, in Er, characterized by a high altitude and comparatively cold weather, the content of β-sitosterol was lower than in Al and Ag. Thus, it appears that sterol variations depend on the geographic origin and climatic conditions.

3.5 | Multivariate analysis

Additionally, one aim was to get a first insight on the question, how geographical origin influences the composition of DSO using multivariate analysis. The objective was to identify any discernible trends or markers associated with specific origins, achieved through PCA and hierarchical clustering analysis (HCA), complemented by heatmaps. The samples from the three Morocco palm groves (Ag, Al, and Er) were first compared with the non-Moroccan samples and also against each other.

To ensure a thorough and reliable analysis, the study consolidated crucial attributes, including oil yield, FA, TAG, tocopherols, and phytosterols, into a unified dataset. Attributes ≤0.1 were not included in the multivariate analysis. Indeed, only attributes displaying statistically significant differences between origins, identified using *t*-tests or ANOVA with a significance level of $p < 0.05$, were included in the subsequent multivariate analysis. The resulting data matrix for the Moroccan versus non-Moroccan comparison contained 17 significant features (*t*-test, $p < 0.05$), whereas the matrix for the comparison among Moroccan samples contained 28 significant features (ANOVA, $p < 0.05$). For enhanced analysis and in order to make features more comparable,

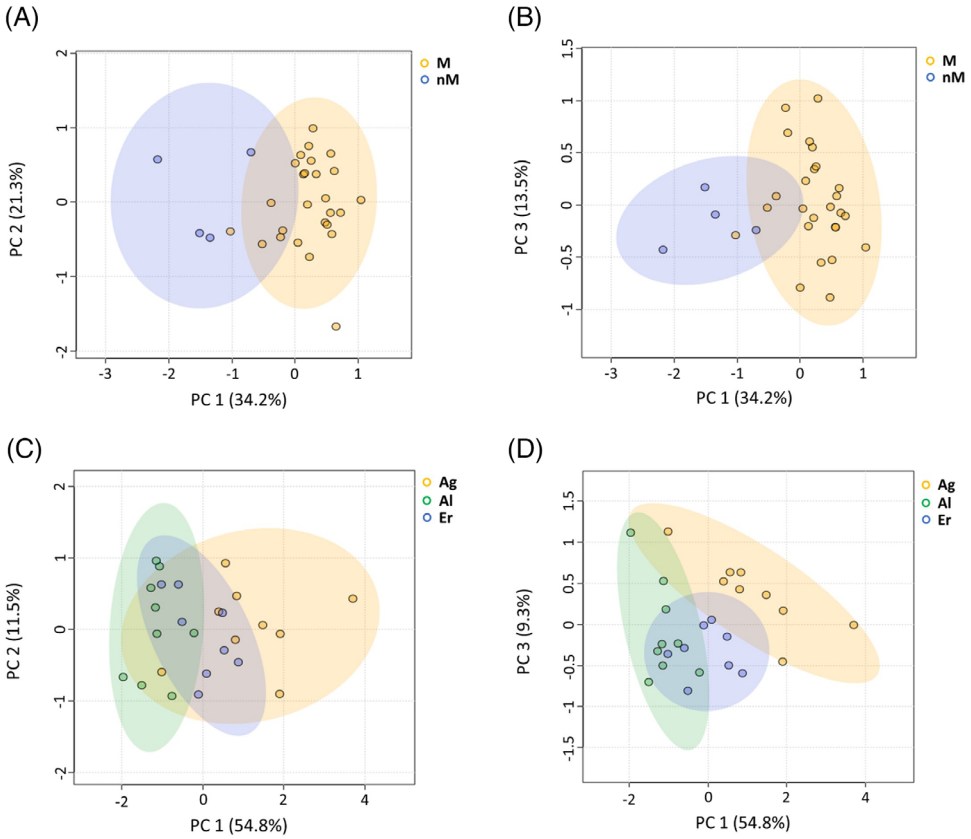


FIGURE 7 Score plots resulting from principal component analysis (PCA) based on the chemical composition of the examined date seed oil (DSO). Score plots A and B: comparison of Moroccan (M) samples against non-Moroccan (nM) samples in PC1/PC2 and PC1/PC3 planes, respectively. Score plots C and D: comparison of Moroccan samples (*Allougoum* (Ag), *Errachidia* (Er), and *Alnif* (Al)) against each other samples in PC1/PC2 plan and PC1/PC3 plan, respectively. The legend provides corresponding geographical origins, and the shaded area represents the 95% confidence intervals.

a generalized log transformation and Pareto scaling were applied to preprocess the data.[54]

In Figure 7A,B, the PCA score plot highlighted a distinct separation between Moroccan and non-Moroccan samples, emphasizing the influence of geographical origin on DSO chemical profiles. Further, within the Moroccan samples, the PCA plots (Figure 7C,D) revealed that PC1/PC3 gives more clustering than PC1/PC2, with no notable clustering between samples from Al and Er, as these two palm groves are very near each other (Figure 2). However, samples from Ag, situated in a southern, lower-altitude region nearer to the Atlantic Ocean, exhibited significant differentiation from the other origins (Figure 7D). This underscores the impact of geographical factors on DSO chemical compositions.

To further explore chemical alterations, HCA was employed using the same data matrices as PCA, visualizing alterations in chemical indices through a heatmap based on Euclidean distance and Ward's linkage (Figure 8A,B). The heatmaps supported the PCA results, highlighting a clearer separation between Moroccan and non-Moroccan samples, with some difficulty in classifying DSO samples by origin within Morocco.

When comparing Moroccan samples to non-Moroccan ones, higher abundance compounds were important in clustering samples according to geographical origin. As discussed earlier, the Moroccan samples were mainly characterized with a relatively higher concentration of St, Ol, Pa, PaOlSt, MyOlPa, PaOlPa, γ-tocotrienol, and δ-tocotrienol as compared with the non-Moroccan samples, and this is clearly shown in the biplot (Figure S1). Moreover, in the heatmap, differences in the concentration of those compounds on a color scale can be identified (Figure 8B). On the other hand, in the comparison within the Moroccan samples, those from Ag were notably clustered in the negative part of the PC1 (Figure 7D). The biplot (Figure S1B) further elucidates that minor TAG such as CaLaLa/CyLaMy, LaLaLa, LaLaMy, LaMyMy, and LaMyPa showed the strongest correlation with samples from Ag. This suggests that focusing on some TAG could be a promising approach to distinguishing the geographical origin of Moroccan DSO; however, larger sample set is required to confirm such an assumption. These findings emphasize also that while high-abundance compounds play a significant role, the less abundant compounds are equally crucial in classifying DSO as previously described for other oilseeds.[55,56]

Variations in date variety further complicated classification, a phenomenon also observed in PCA plots (Figure 7), where samples from the same provenance were widely dispersed within the cluster. This aligns with prior reports indicating variations in DSO chemical

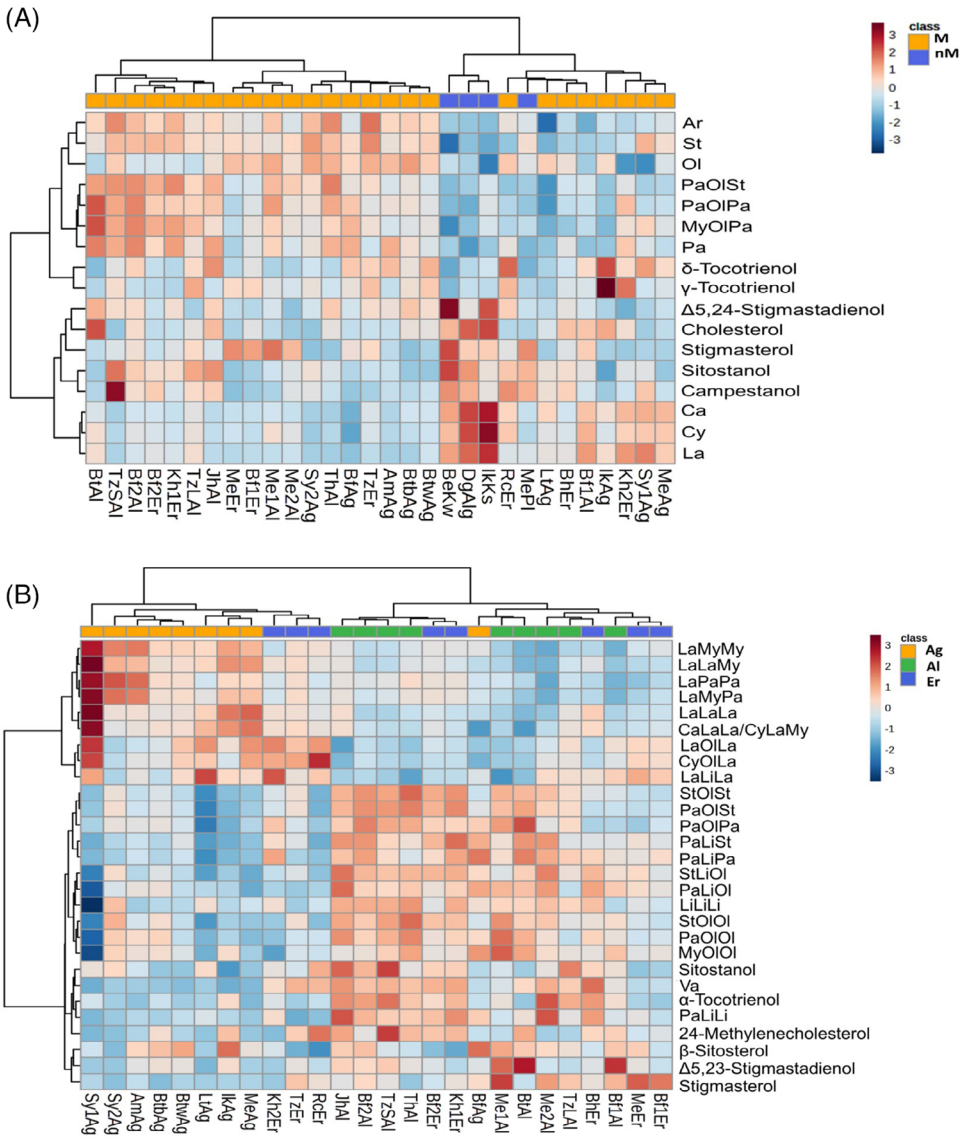


FIGURE 8 Clustering of date seed oil (DSO) samples by geographical origin using hierarchical cluster analysis (HCA) and a heatmap: (A) comparison of Moroccan (M) samples against non-Moroccan (nM) samples and (B) comparison of Moroccan samples (*Allougoum* (Ag), *Errachidia* (Er), and *Alnif* (Al)) against each other. The color scale on the right represents the original content of each variable, with blue and red colored cells indicating high and low content, respectively. The heatmaps display the data with the sample on the horizontal axis and the compounds on the vertical axis. The full sample names are available in Table S1 for reference. Ar, arachidic acid; Ca, capric acid; Cy, caprylic acid; La, lauric acid; Li, linoleic acid; My, myristic acid; Ol, oleic acid; Pa, palmitic acid; St, stearic acid; Va, vaccenic acid.

composition even among different varieties grown in the same region, suggesting the need to consider varieties in future geographical origin studies.

4 | CONCLUSION

This study aimed at investigating the chemical composition of DSO and its variations based on geographical origin, with a particular emphasis on Moroccan date seed samples. FA, TAG, tocochromanols, and phytosterols were analyzed. The results showed that the chemical composition of DSO varies notably based on the geographical origin of the date seeds. Employing advanced multivariate analysis techniques, including PCA and HCA with heatmaps, substantiated this influence of geographical origin on DSO composition. This underscored the necessity of considering both major and minor compounds, as well as date varieties, in studies exploring geographical effects on DSO. However, it is important to recognize the limitations of such an initial dataset. The sample size is relatively small, and there exists a slight imbalance in the distribution of samples among different regions and varieties, which could potentially introduce bias into the analysis. Furthermore, present analyses did not reveal any distinct patterns related to date

fruit quality or prices in DSO chemical composition, which is evident in both Figures 7 and 8. These findings suggest that, for large-scale DSO extraction, even low-quality date fruits can be considered viable sources. Future studies should strive to address these limitations by involving larger sample sizes from diverse production regions and different harvest years. A more extensive sample set will strengthen the robustness and generalizability of the conclusions drawn in this study. Additionally, investigating variety × environmental interactions on phytochemical composition and exploring the entire polar fraction of DSO via an untargeted metabolomics approach present intriguing prospects for future research.

AUTHOR CONTRIBUTIONS

Said El Harkaoui: Conceptualization; formal analysis; investigation; methodology; writing—original draft; writing—review and editing. Katharina N'Diaye: Methodology; investigation; writing—review and editing. Said Gharby and Maryam Al-Hilal: Conceptualization; resources; writing—review and editing. Zoubida Charrouf: Conceptualization; supervision; writing—review and editing; project administration. Sascha Rohn: Conceptualization; supervision; validation; writing—review and editing. Bertrand Matthäus: conceptualization; resources; validation; supervision; writing—review and editing; project administration; funding acquisition.

ACKNOWLEDGMENTS

This work was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) (support program: FKZ 2819DOKA03). The authors would like to thank Dr. Klaus Vosmann and Ms. Heike Wolf of the Max Rubner Institute, Detmold, Germany. Dr. Vosmann's expertise in the analysis of triacylglycerols and his guidance in the interpretation of the MS spectra were of great importance for our work, while the technical support of Ms. Heike Wolf proved invaluable throughout the project.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

REFERENCES

1. Elleuch, M., Besbes, S., Roiseux, O., Blecker, C., Deroanne, C., DRIRA, N.-E., & Attia, H. (2008). Date flesh: Chemical composition and characteristics of the dietary fibre. *Food Chemistry*, *111*, 676–682.
2. Bouhlali, E. D. T., Ramchoun, M., Alem, C., Ghafoor, K., Ennassir, J., & Zegzouti, Y. F. (2017). Functional composition and antioxidant activities of eight Moroccan date fruit varieties (*Phoenix dactylifera* L.). *Journal of the Saudi Society of Agricultural Sciences*, *16*, 257–264.
3. Fikry, M., Yusof, Y. A., Al-Awaadh, A. M., Rahman, R. A., Chin, N. L., Mousa, E., & Chang, L. S. (2019). Effect of the roasting conditions on the physicochemical, quality and sensory attributes of coffee-like powder and brew from defatted palm date seeds. *Foods*, *8*, 61.
4. Mrabet, A., Jiménez-Araujo, A., Guillén-Bejarano, R., Rodríguez-Arcos, R., & Sindic, M. (2020). Date seeds: A promising source of oil with functional properties. *Foods*, *9*, 787.
5. FAO 2023. Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL. Accessed 3/25/2023
6. Alkhoori, M. A., Kong, A. S.-Y., Aljaafari, M. N., Abushelaibi, A., Erin Lim, S.-H., Cheng, W.-H., Chong, C.-M., & Lai, K.-S. (2022). Biochemical composition and biological activities of date palm (*Phoenix dactylifera* L.) seeds: A review. *Biomolecules*, *12*, 1626.
7. Echegaray, N., Gullón, B., Pateiro, M., Amarowicz, R., Misihairabgwi, J. M., & Lorenzo, J. M. (2023). Date fruit and its by-products as promising source of bioactive components: A review. *Food Reviews International*, *39*, 1411–1432.
8. Farag, M. A., Otify, A., & Baky, M. H. (2021). *Phoenix dactylifera* L. date fruit by-products outgoing and potential novel trends of phytochemical, nutritive and medicinal merits. *Food*, *Reviews International*, *39*, 1–23.
9. Maqsood, S., Adiamo, O., Ahmad, M., & Mudgil, P. (2020). Bioactive compounds from date fruit and seed as potential nutraceutical and functional food ingredients. *Food Chemistry*, *308*, 125522.
10. Deutsche Gesellschaft für Fettwissenschaft (DGF). (2021). *Deutsche einheitsmethoden zur untersuchung von fetten & fettprodukten, tensiden und verwandten stoffen*. Wissenschaftliche Verlagsgesellschaft.
11. Ruiz-Samblás, C., González-Casado, A., Cuadros-Rodríguez, L., & García, F. P. R. (2010). Application of selected ion monitoring to the analysis of triacylglycerols in olive oil by high temperature-gas chromatography/mass spectrometry. *Talanta*, *82*, 255–260.
12. Pang, Z., Chong, J., Zhou, G., De Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P. É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*, *49*, W388–W396.
13. Ibourki, M., Azouguigh, F., Jadouali, S. M., Sakar, E. H., Bijla, L., Majourhat, K., Gharby, S., & Laknifli, A. (2021). Physical fruit traits, nutritional composition, and seed oil fatty acids profiling in the main date palm (*Phoenix dactylifera* L.) varieties grown in Morocco. *Journal of Food Quality*, *2021*, 1–12.
14. Bhatnagar, A. S., Kumar, P. K. P., Hemavathy, J., & Gopala Krishna, A. G. (2009). Fatty acid composition, oxidative stability, and radical scavenging activity of vegetable oil blends with coconut oil. *Journal of the American Oil Chemists Society*, *86*, 991–999.
15. Li, C., Yao, Y., Zhao, G., Cheng, W., Liu, H., Liu, C., Shi, Z., Chen, Y., & Wang, S. (2011). Comparison and analysis of fatty acids, sterols, and tocopherols in eight vegetable oils. *Journal of Agricultural and Food Chemistry*, *59*, 12493–12498.
16. Ayyildiz, H. F., Topkafa, M., Kara, H., & Sherazi, S. T. H. (2015). Evaluation of fatty acid composition, tocols profile, and oxidative stability of some fully refined edible oils. *International Journal of Food Properties*, *18*, 2064–2076.
17. Habib, H. M., Kamal, H., Ibrahim, W. H., & Dhaheri, A. S. A. (2013). Carotenoids, fat soluble vitamins and fatty acid profiles of 18 varieties of date seed oil. *Industrial Crops and Products*, *42*, 567–572.
18. Harkat, H., Bousba, R., Benincasa, C., Atrouz, K., Gültekin-Özgüven, M., Altuntaş, Ü., Demircan, E., Zahran, H. A., & Özçelik, B. (2022). Assessment of biochemical composition and antioxidant properties of algerian date palm (*Phoenix dactylifera* L.) seed oil. *Plants*, *11*, 381.
19. Holcapek, M., Lísa, M., Jandera, P., & Kabátová, N. (2005). Quantitation of triacylglycerols in plant oils using HPLC with APCI-MS, evaporative light-scattering, and UV detection. *Journal of Separation Science*, *28*, 1315–1333.
20. Jukić Špika, M., Perica, S., Žanetić, M., & Škevin, D. (2021). Virgin olive oil phenols, fatty acid composition and sensory profile: Can cultivar overpower environmental and ripening effect? *Antioxidants*, *10*, 689.
21. Ghnimi, S., Umer, S., Karim, A., & Kamal-Eldin, A. (2017). Date fruit (*Phoenix dactylifera* L.): An underutilized food seeking industrial valorization. *NFS Journal*, *6*, 1–10.

22. Nguyen, Q.-H., Talou, T., Evon, P., Cerny, M., & Merah, O. (2020). Fatty acid composition and oil content during coriander fruit development. *Food Chemistry*, *326*, 127034.
23. Telci, I., Sahin-Yaglioglu, A., Eser, F., Aksit, H., Demirtas, I., & Tekin, S. (2014). Comparison of seed oil composition of *Nigella sativa* L. and *N. damascena* L. during seed maturation stages. *Journal of the American Oil Chemists Society*, *91*, 1723–1729.
24. Sinha, P., Islam, M. A., Negi, M. S., & Tripathi, S. B. (2015). Changes in oil content and fatty acid composition in *Jatropha curcas* during seed development. *Industrial Crops and Products*, *77*, 508–510.
25. Rebey, I. B., Aidi Wannes, W., Kaab, S. B., Bourgou, S., Tounsi, M. S., Ksouri, R., & Fauconnier, M. L. (2019). Bioactive compounds and antioxidant activity of *Pimpinella anisum* L. accessions at different ripening stages. *Scientia Horticulturae*, *246*, 453–461.
26. Rebey, I. B., Iness, J.-K., Hamrouni-Sellami, I., Soumaya, B., Ferid, L., & Brahim, M. (2012). Effect of drought on the biochemical composition and antioxidant activities of cumin (*Cuminum cyminum* L.) seeds. *Industrial Crops and Products*, *36*, 238–245.
27. Laribi, B., Iness, B., Karima, K., Ali, S., Abdelaziz, M., & Brahim, M. (2009). Water deficit effects on caraway (*Carum carvi* L.) growth, essential oil and fatty acid composition. *Industrial Crops and Products*, *30*, 372–379.
28. Bettaieb, I., Knioua, S., Hamrouni, I., Limam, F., & Marzouk, B. (2011). Water-deficit impact on fatty acid and essential oil composition and antioxidant activities of cumin (*Cuminum cyminum* L.) aerial parts. *Journal of Agricultural and Food Chemistry*, *59*, 328–334.
29. Nouraei, S., Rahimmalek, M., Saeidi, G., & Bahreininejad, B. (2016). Variation in seed oil content and fatty acid composition of globe artichoke under different irrigation regimes. *Journal of the American Oil Chemists Society*, *93*, 953–962.
30. Besbes, S., Blecker, C., Deroanne, C., Lognay, G., Drira, N.-E., & Attia, H. (2004). Quality characteristics and oxidative stability of date seed oil during storage. *Food Science and Technology International*, *10*, 333–338.
31. Lieb, V. M., Kleiber, C., Metwali, E. M., Kadasa, N. M., Almaghrabi, O. A., Steingass, C. B., & Carle, R. (2020). Fatty acids and triacylglycerols in the seed oils of Saudi Arabian date (*Phoenix dactylifera* L.) palms. *International Journal of Food Science and Technology*, *55*, 1572–1577.
32. Wei, W., Sun, C., Jiang, W., Zhang, X., Hong, Y., Jin, Q., Tao, G., Wang, X., & Yang, Z. (2019). Triacylglycerols fingerprint of edible vegetable oils by ultra-performance liquid chromatography-Q-ToF-MS. *LWT*, *112*, 108261.
33. Qian, Y., Rudzińska, M., Grygier, A., & Przybylski, R. (2020). Determination of triacylglycerols by HTGC-FID as a sensitive tool for the identification of rapeseed and olive oil adulteration. *Molecules (Basel, Switzerland)*, *25*, 3881.
34. Chbani, M., El Harkaoui, S., Willenberg, I., & Matthäus, B. (2022). Review: Analytical extraction methods, physicochemical properties and chemical composition of cactus (*Opuntia ficus-indica*) seed oil and its biological activity. *Food Reviews International*, *39*, 4496–4512.
35. Hilali, M., Charrouf, Z., Soulhi, A. E. A., Hachimi, L., & Guillaume, D. (2005). Influence of origin and extraction method on argan oil physicochemical characteristics and composition. *Journal of Agricultural and Food Chemistry*, *53*, 2081–2087.
36. El harkaoui, S., Gharby, S., Kartah, B., El Monfalouti, H., El-sayed, M. E., Abdin, M., Salama, M. A., Charrouf, Z., & Matthäus, B. (2023). Lipid profile, volatile compounds and oxidative stability during the storage of Moroccan *Opuntia ficus-indica* seed oil. *Grasas y Aceites*, *74*, e486.
37. Juhaimi, F. A., Özcan, M. M., Adiamo, O. Q., Alsawmahi, O. N., Ghafoor, K., & Babiker, E. E. (2018). Effect of date varieties on physico-chemical properties, fatty acid composition, tocopherol contents, and phenolic compounds of some date seed and oils. *Journal of Food Processing and Preservation*, *42*, e13584.
38. Nehdi, I. A., Sbihi, H. M., Tan, C. P., Rashid, U., & Al-Resayes, S. I. (2018). Chemical composition of date palm (*Phoenix dactylifera* L.) seed oil from six Saudi Arabian cultivars. *Journal of Food Science*, *83*, 624–630.
39. Nehdi, I., Omri, S., Khalil, M. I., & Al-Resayes, S. I. (2010). Characteristics and chemical composition of date palm (*Phoenix canariensis*) seeds and seed oil. *Industrial Crops and Products*, *32*, 360–365.
40. Mohamed Mousa, Y., Gerasopoulos, D., Metzidakis, I., & Kiritsakis, A. (1996). Effect of altitude on fruit and oil quality characteristics of 'Mastoides' olives. *Journal of the Science of Food and Agriculture*, *71*, 345–350.
41. Elgadi, S., Ouhammou, A., Zine, H., Maata, N., Aitlhaj, A., El Allali, H., & El Antari, A. (2021). Discrimination of geographical origin of unroasted kernels argan oil (*Argania spinosa* (L.) Skeels) using tocopherols and chemometrics. *Journal of Food Quality*, *2021*, 1–9.
42. Frankel, E. N. (2005) *Lipid oxidation*. Woodhead Publishing Limited.
43. Behrman, E. J., & Gopalan, V. (2005). Cholesterol and plants. *Journal of Chemical Education*, *82*, 1791.
44. Oh, M.-J., So, H.-J., Hong, E.-S., Shin, J.-A., & Lee, K.-T. (2021). Presence of cholesterol in non-animal organisms: identification and quantification of cholesterol in crude seed oil from *Perilla frutescens* and dehydrated *Pyropia tenera*. *Molecules (Basel, Switzerland)*, *26*, 3767.
45. Moreau, R. A., Nyström, L., Whitaker, B. D., Winkler-Moser, J. K., Baer, D. J., Gebauer, S. K., & Hicks, K. B. (2018). Phytosterols and their derivatives: Structural diversity, distribution, metabolism, analysis, and health-promoting uses. *Progress in Lipid Research*, *70*, 35–61.
46. Shukla, V. K. S., Dutta, P. C., & Artz, W. E. (2002). *Camelina* oil and its unusual cholesterol content. *Journal of the American Oil Chemists Society*, *79*, 965–969.
47. Kirkhus, B., Lundon, A. R., Haugen, J.-E., Vogt, G., Borge, G. I. A., & Henriksen, B. I. F. (2013). Effects of environmental factors on edible oil quality of organically grown *Camelina* sativa. *Journal of Agricultural and Food Chemistry*, *61*, 3179–3185.
48. Liu, Y., Tu, X., Lin, L., Du, L., & Feng, X. (2022). Analysis of lipids in pitaya seed oil by ultra-performance liquid chromatography-time-of-flight tandem mass spectrometry. *Foods*, *11*, 2988.
49. Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., Unger, T., Malitsky, S., Arendt, P., Pauwels, L., Almekias-Siegl, E., Rogachev, I., Meir, S., Cárdenas, P. D., Masri, A., Petrikov, M., Schaller, H., Schaffer, A. A., Kamble, A., & Aharoni, A. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nature Plants*, *3*, 16205.
50. Besbes, S., Blecker, C., Deroanne, C., Bahloul, N., Lognay, G., Drira, N.-E., & Attia, H. (2004). Date seed oil: Phenolic, tocopherol and sterol profiles. *Journal of Food Lipids*, *11*, 251–265.
51. Noorali, M., Barzegar, M., & Sahari, M. A. (2014). Sterol and fatty acid compositions of olive oil as an indicator of cultivar and growing area. *Journal of the American Oil Chemists Society*, *91*, 1571–1581.
52. Piravi-Vanak, Z., Ghasemi, J. B., Ghavami, M., Ezzatpanah, H., & Zolfonoun, E. (2012). The influence of growing region on fatty acids and sterol composition of iranian olive oils by unsupervised clustering methods. *Journal of the American Oil Chemists' Society*, *89*, 371–378.
53. Kharbach, M., Kamal, R., Bousrabat, M., Alaoui Mansouri, M., Barra, I., Alaoui, K., Cherrah, Y., Ander Heyden, V. Y., & Bouklouze, A. (2017). Characterization and classification of PGI Moroccan argan oils based on their FTIR fingerprints and chemical composition. *Chemometrics and Intelligent Laboratory Systems*, *162*, 182–190.
54. Van Den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics [Electronic Resource]*, *7*, 142.
55. Zhang, L., Li, P., Sun, X., Wang, X., Xu, B., Wang, X., Ma, F., Zhang, Q., & Ding, X. (2014). Classification and adulteration detection of vegetable oils based on fatty acid profiles. *Journal of Agricultural and Food Chemistry*, *62*, 8745–8751.
56. Wang, X., Zeng, Q., Verardo, V., & Del Contreras, M. M. (2017). Fatty acid and sterol composition of tea seed oils: Their comparison by the "FancyTiles" approach. *Food Chemistry*, *233*, 302–310.

57. Huntington, J. L., Hegewisch, K. C., Daudert, B., Morton, C. G., Abatzoglou, J. T., McEvoy, D. J., & Erickson, T. (2017). *Climate engine: Cloud Computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding*. Bulletin of the American Meteorological Society.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: El Harkaoui, S., N'Diaye, K., Gharby, S., Al-Hilal, M., Charrouf, Z., Rohn, S., & Matthäus, B. (2024). Insights into date seed oil composition: Geographical variability and potential applications. *European Journal of Lipid Science & Technology*, *126*, e2400061. https://doi.org/10.1002/ejlt.202400061

Manuscript IV

Untargeted metabolomic analysis of date seed oil (Phoenix dactylifera L.) using UHPLC-ESI-QTOF-MS - evaluation of the geographical origin effect

Said El Harkaoui, Katharina N'Diaye, Zoubida Charrouf, Sascha Rohn, Stephan Drusch, Bertrand Matthäus

Food Chemistry: X 2025, 31,103162

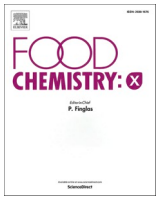
Accepted manuscript available via https://doi.org/10.1016/j.fochx.2025.103162

Food Chemistry: X 31 (2025) 103162

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Untargeted metabolomic analysis of date seed oil (*Phoenix dactylifera* L.) using UHPLC-ESI-QTOF-MS - evaluation of the geographical origin effect

Said El Harkaoui [a,b,c,*], Katharina N'Diaye [a], Zoubida Charrouf [d], Sascha Rohn [b], Stephan Drusch [c], Bertrand Matthäus [a]

[a] Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Schützenberg 12, 32756 Detmold, Germany
[b] Department of Food Chemistry and Analysis, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany
[c] Department of Food Technology and Food Material Science, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany
[d] Department of Chemistry, Faculty of Sciences, Mohammed V University in Rabat, Morocco

ARTICLE INFO

Keywords:
Phoenix dactylifera
date seed oil
metabolomic
geographical origin
authenticity
multivariate analysis
UHPLC-ESI-QTOF-MS

ABSTRACT

The chemical composition of edible oils is influenced by geographical origin, making traceability essential for ensuring authenticity and quality, particularly for unconventional (edible) oils like date seed oil (DSO). For economic, food safety, and research purposes, this study aimed to characterize the metabolomic profile of DSO and assess the impact of Moroccan origin to identify key chemical markers linked to regional differences. An untargeted metabolomic approach using UHPLC-QTOF-MS was applied to DSO samples from three Moroccan palm groves: *Allougoum*, *Alnif*, and *Errachidia*. PCA revealed modest clustering for *Allougoum* samples, while OPLS-DA enabled selection of 50 features contributing to differentiation. Among these, 25 metabolites were tentatively identified as geographical markers with three, n-(3-oxohexanoyl) homoserine lactone, indole-3-carboxaldehyde, and vanillin, confirmed using authentic standards. Hydroxy fatty acids were the most represented class, with four compounds tentatively annotated. This study provides the first comprehensive metabolomic profile of DSO and highlights potential markers for origin differentiation.

1. Introduction

The date palm (*Phoenix dactylifera* L.) is a crucial crop in arid and semi-arid regions, especially in North Africa and the Middle East, where its fruits serve as a major economic and nutritional resource. In Morocco, it is the most important arboriculturally in arid areas, playing a vital role in both traditional and commercial agriculture (Bouhlali, Alem, et al., 2017). While date fruits are commonly processed into products such as syrup, jam, and vinegar, their seeds, so far regarded as waste, are increasingly recognized as valuable by-products for their potential uses. For instance, roasted date seeds powder are used to prepare coffee-like beverages, and the seed can also be a potential (edible) oil source especially in arid areas (Farag et al., 2021; Mrabet et al., 2020). Date seed oil (DSO) is abundant in bioactive compounds, such as vitamin E vitamers, phenolic compounds, phytosterols, and others, making it an attractive candidate for functional applications in both the food and non-food sectors (Alkhoori et al., 2022; Echegaray et al., 2021; Farag et al., 2021; Maqsood et al., 2020; Mrabet et al., 2020).

Although edible oils, being recognized to possess lower chemical diversity, compared to other foods, their composition has been linked to terroir. This has been well documented for various edible oils, including extra-virgin olive oil, palm oil, or argan oil, among others (Lucini, Rocchetti, & Trevisan, 2020). Similarly, the chemical composition of DSO is influenced by multiple factors, including date variety and geographical origin, which will affect its biological activities as well as nutritional and functional properties (Abdul-Hamid et al., 2019). Thus, understanding how geographical origin impacts the composition of DSO is critical for quality control, authenticity verification, consumer trust, and economic valorization of date seeds as by-product (Caporale & Monteleone, 2001).

Previous studies mostly focused on major lipid classes of DSO and the influence of date variety, with limited attention to geographical effects (Al Juhaimi et al., 2018; Habib, Kamal, Ibrahim, & Dhaheri, 2013; Harkat et al., 2022; Lieb et al., 2020; Nehdi, Sbihi, Tan, Rashid, & Al-Resayes, 2018). Most sample collections were limited to specific regions, lacking a systematic comparison across different locations. When

* Corresponding author at: Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Schützenberg 12, 32756 Detmold, Germany.
E-mail address: Said.Elharkaoui@mri.bund.de (S. El Harkaoui).

https://doi.org/10.1016/j.fochx.2025.103162
Received 10 July 2025; Received in revised form 5 October 2025; Accepted 13 October 2025
Available online 14 October 2025

comparing DSO composition across studies from different countries, noticeable differences emerged, likely due to geographical variation. However, these observations were drawn from separate publications rather than from an individual study, explicitly designed to assess intra-country geographical effects (Al Juhaimi et al., 2018; Habib et al., 2013; Harkat et al., 2022; Lieb et al., 2020; Nehdi et al., 2018).

In a recent study, the main lipid composition of DSO from three Moroccan palm groves (*Allougoum*, *Alnif*, and *Errachidia*) were characterized, aiming at evaluating the effect of geographical location on DSO (El Harkaoui et al., 2024). Conventional lipid analyses, including fatty acids, triacylglycerols (TAG), tocochromanols, and phytosterols, were combined with multivariate statistical approaches such as principal component analysis (PCA) and heatmaps to identify compositional differences. While some trends were observed, differentiation between Moroccan palm groves remained challenging. Nevertheless, the *Allougoum* region exhibited a distinct clustering pattern, associated with variations in minor TAG. Especially minor compounds and patterns thereof provide relevance in authenticity assessment (Qian et al., 2020). These findings suggest that geographical origin may influence DSO composition, even when analyzing major lipid constituents.

While conventional lipid analysis provided valuable insights, it only captured a limited, portion of DSO's chemical diversity, excluding and even neglecting the other important parts of the metabolome, highlighting the need for a broader analytical approach. The previous study was a crucial first step in establishing authenticity markers, but its scope was inherently constrained (El Harkaoui et al., 2024). Moving towards an untargeted metabolomic approach allows for a more comprehensive investigation of DSO composition, potentially identifying novel markers that enhance geographical differentiation. A more wider and comprehensive profile, including the more polar or amphiphilic metabolites, holds significant potential for identifying reliable geographical markers (Hu, Zhang, Xing, Yu, & Chen, 2022).

Obviously, ultra-high-performance liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry (UHPLC-ESI-QTOF-MS) is a well-established technique for untargeted metabolomics, offering high sensitivity and resolution, as well as the capability to detect a diverse range of metabolites, including polar compounds that are often neglected in traditional lipid analysis. This approach has been widely employed in the geographical authentication of edible oils, particularly in olive oil (Ghisoni et al., 2019; Gil-Solsona et al., 2016; Kalogiouri, Aalizadeh, & Thomaidis, 2018; Mohamed et al., 2018; Willenberg, Parma, Bonte, & Matthäus, 2021), where secondary metabolites have been successfully used as markers. Given the effectiveness of metabolomic profiling in other oil types, applying this approach to DSO represents a logical extension of the previous work (El Harkaoui et al., 2024), with the potential to reveal novel compositional markers linked to geographical origin. However, a major challenge of untargeted metabolomics is the vast amount of data generated, necessitating advanced statistical and computational methods for effective interpretation. The high dimensionality of metabolomic datasets requires multivariate statistical techniques such as PCA, orthogonal partial least squares-discriminant analysis (OPLS-DA), and variable importance in projection (VIP) scores to extract meaningful patterns. These chemometric tools enhance classification accuracy by identifying key metabolites responsible for regional compositional differences, thus, improving the discrimination of samples based on origin (Mattoli, Gianni, & Burico, 2022; Yi et al., 2016). However, it remains unclear, if wheter DSO from different Moroccan origins, can be differentiated from each other, and wheter authentication is possible not only for economic reasons, but also with regard to food safety. A strategy with appropriate food analytical methodologies must be developed. This would enhance marketability of Moroccan DSO, but can also serve as another scientific example for food authenticity testing.

The main aim of the present study was to establish, for the first time, a comprehensive untargeted metabolomic profile of Moroccan DSO using UHPLC-ESI-QTOF-MS combined with advanced chemometric analysis, and to evaluate whether this profile varies across different palm groves within Morocco. Specifically, the study investigates compositional differences among samples from *Allougoum*, *Alnif*, and *Errachidia*, and seeks to identify key polar metabolites that may serve as geographical markers. This work builds on previous studies focusing on the analysis of major lipid constituent and contributes new insights into the authentication and traceability of Moroccan DSO, while also highlighting the potential of date seeds as valuable by-products for different functional applications. A high quality of such novel approaches is mandatorily to gain consumer trust.

2. Material and methods

2.1. Material

The sample set used consists of 26 date seed samples from three Moroccan provenances: *Allougoum* (Ag), *Alnif* (Al), and *Errachidia* (Er). The exact location of the sampling provenances is visualized in Fig. 1. These are the same samples used in a previous publication (El Harkaoui et al., 2024), where detailed information on sample codes, collection sites, and environmental parameters such as average temperature and annual rainfall is provided. The varieties of date seeds included were 'Amchaw', 'Berhi', 'Boufeggous', 'Bouhmroune', 'Bousthammi', 'Ikhlas', 'Iklan', 'Jihel', 'Khalt', (clone) 'Lmtrwah', 'Mejhoul', 'Racetmar', 'Sayer', 'Tahmout', and 'Tarzawa'. DSO was extracted using hexane (Promochem, Picograde quality, LGC Standards GmbH, Wesel, Germany) for 6 h in a Twisselmann apparatus, following the standard methods outlined by the *German Society for Fat Science* (DGF B—I 5 (12) and B-II 4a (09)) (DGF, 2021). Vanillin (99 %), indole-3-carboxaldehyde (97 %), and *n*-(3-oxohexanoyl) homoserine lactone (97 %) (Merck KGaA, Darmstadt, Germany) were used as reference compounds.

2.2. Sample preparation

The polar extract from the Moroccan DSO samples was obtained following an optimized method reported by (Willenberg et al., 2021) with some minor modifications. In brief, 1 g DSO was weighed in a 10 mL glass tube and 2 mL of the extraction solvent were added (methanol/H_2O, 80:20, (*v*/v)). The mixture was shaken vigorously for 1 min at 1500 min^{-1} (VXR basic Vibrax; IKA Werke GmbH & Co. KG, Staufen, Germany), then centrifuged (1550 ×*g*, 15 min) and 1.5 mL of the resulting supernatant were collected into a new 10 mL-glass tube. The remaining lower phase was extracted a second time in the same manner and 2 mL of the second supernatant were combined with the first extract. The combined extracts were evaporated to dryness at 40 °C with 1 mbar vacuum using a rotational vacuum concentrator (RVC 233 CDplus, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany). The residue was dissolved in 0.25 mL of methanol/H_2O (80:20, *v*/v), shaken vigorously for 1 min (1,500 min^{-1}) and transferred into a 1.5 mL-microreaction tube. After centrifugation for 5 min at 16,000 ×g, the supernatant was transferred into an HPLC vial with a glass insert. Once the samples were processed as described above and filled into vials, a pooled sample (quality control, QC) was prepared by combining equal volumes from each vial and vigorously mixing them into a homogenous pooled sample.

2.3. Instrumentation

The untargeted analysis of the polar extract was carried out using an UHPLC-ESI-QTOF-MS approach. For the UHPLC (Ultimate 3000 Series, Thermo Fisher Scientific Inc., Waltham, MA, USA), samples were kept in the autosampler at 10 °C until injection of a 3 µL aliquot. Separation was carried out at 40 °C on a 1.7 µm × 150 mm × 2.1 mm Kinetex EVO C18-reversed phase column with a SecurityGuard ULTRA sub-2 µm pre-column (Phenomenex Ltd. Deutschland, Aschaffenburg, Germany). An aqueous solution of 0.1 % formic acid (98 % p.a., Honeywell Specialty

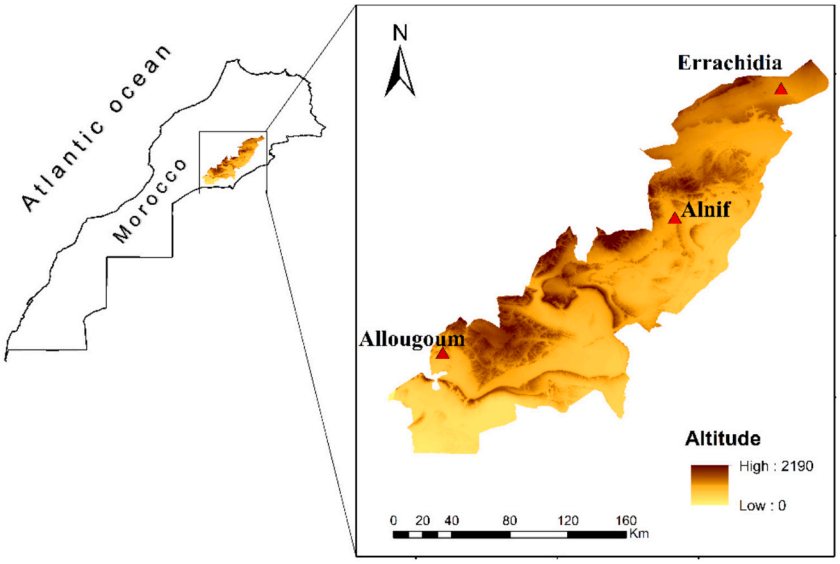


Fig. 1. Sampling provenances *Allougoum*, *Alnif*, and *Errachidia* (red triangles) with regard to location and altitude. Figure reproduced from (El Harkaoui et al., 2024), licenced under CC BY 4.0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Chemicals Seelze GmbH, Seelze, Germany) was used as eluent A, and methanol (Honeywell Specialty Chemicals Seelze GmbH, Seelze, Germany) acidified with 0.1 % formic acid was used as eluent B. The elution began with 20 % eluent B and a flow rate of 0.3 mL/min. A gradient was then used, starting with a linear increase from 20 % to 65 % B over 1 to 2.5 min, followed by a linear increase from 65 % to 99 % B over 2.5 to 8 min, and an isocratic hold at 99 % B from 8 to 11 min. Finally, the system was re-equilibrated to the initial conditions (20 % eluent B) in less than a minute, and the system was left for 3 min until the next start.

The mass spectrometric detection was done after electrospray ionization (positive and negative modes) using a QTOF-MS (impact HD, Bruker Daltonics GmbH & Co. KG, Bremen, Germany). The data were acquired over an *m/z* ratio ranging from 50 to 1000 Da using the following TOF parameters: capillary voltage 4500 V, drying gas temperature 250 °C, dry gas flow 10 L/min, nebulizing gas pressure 3 bar and plate offset -500 V. Tune parameter were: Funnel 1 RF 150 Vpp; funnel 2 RF 200 Vpp, isCID energy 0 eV, hexapole RF 100 Vpp, ion energy 4 eV, low mass 50 *m/z*, collision RF 500 Vpp, transfer time 100 µs, pre-pulse storage 5 µS. MS/MS parameters: bbCID mode, collision energy MS mode 7 eV, MS/MS mode 40 eV. The Compass 2023b for otof series package with otofControl 6.3 (Bruker Daltonics GmbH & Co. KG, Bremen, Germany) was used for data acquisition. The system was calibrated using sodium formate solution (10 mM) injected at 0.18 µL/min. During the first 30 s of every run. Samples were injected randomly in triplicates, with QC and blank samples analyzed every ten injections to monitor instrument performance.

2.4. Data processing

The mass calibration processing and initial inspection was performed using DataAnalysis 6.1 (Bruker Daltonics GmbH & Co. KG, Bremen, Germany). For further data processing (peak picking, retention time alignment, and normalization), the raw data set of the analyzed samples was converted into *.abf files using Analysis Base File Converter (Reifycs Inc., Tokyo, Japan) to make them compatible with the software MS-Dial version 5.5 (RIKEN Center, Kanagawa, Japan). The data collection was done within the retention range 1 min to 12 min and a mass range of 50–1000 Da. For peak detection, a minimum peak height of 5000 (negative mode) and 10,000 (positive mode), and the selected adduct were $[M + H]^+$ and $[M\text{-}H]^-$ for the positive and negative mode, respectively. The QC sample in the middle of the batch was selected as an alignment sample with a retention time tolerance of 0.5 min and a MS tolerance of 0.015 Da. The complete applied MS-Dial parameters, including those used for data collection, peak detection and alignment are reported in Table S1. MS-Dial integrated LOWESS drift normalization was performed on the data set to reduce systematic variations and ensure reproducibility. The resulting feature list which correspond to a distinct compound with the associated *m/z* value and retention time was exported to a .csv file (Microsoft, Redmont, WA, USA) for further steps. Following, the data was filtered, log-transformed, and Pareto-scaled before undergoing multivariate analysis in MetaboAnalyst 5.0 (Pang et al., 2021). Subsequently, a selection of potential markers for discriminating the geographical origin of DSO oil was compiled based on

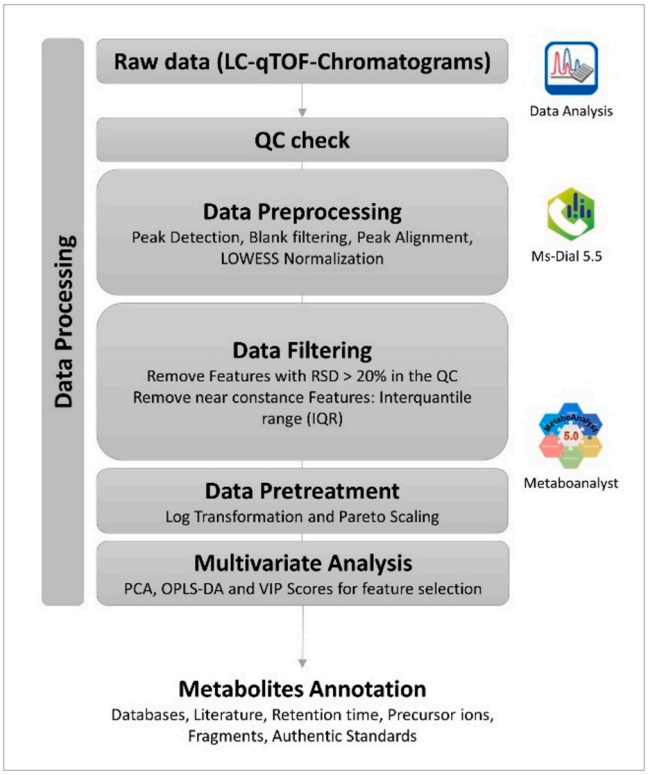


Fig. 2. Workflow from data processing to metabolites annotation.

the processed data. The overall data processing workflow is summarized in Fig. 2.

2.5. Screening and identification

Significant features were selected, excluding those resulting from in-source fragmentation. Molecular formulas were generated using Bruker's *Smart Formula* tool in DataAnalysis 6.1, applying a mass deviation threshold of <5 ppm and an isotope distribution deviation of <20 %. These formulas, along with exact masses, were then searched and tentatively annotated using online metabolite databases, including LipidMaps (https://www.lipidmaps.org/, last accessed 22/11/2024), Mass Bank of North America (https://mona.fiehnlab.ucdavis.edu/, last accessed 22/11/2024), and Foodb (https://foodb.ca/, last accessed 22/11/2024). Annotations from different databases were crosschecked, and, when available, MS/MS spectra with acquisition parameters similar to ours were considered for fragment comparison. When applicable, authentic standards were also used to confirm the annotated features. To ensure reliable identification, features were classified based on the confidence level criteria proposed by Schymanski et al. (2014) and Sumner et al. (2007). In the present study, annotation levels were assigned as follows: Level 1 (confirmed structure with an authentic standard), Level 2 (probable structure, single candidate based on comparison with databases MS/MS spectra), Level 3 (tentative candidate (s)), Level 4 (unequivocal molecular formula), and Level 5 (unknown, exact mass).

3. Results and discussion

3.1. Data elaboration

To obtain comprehensive information on the metabolites present in Moroccan DSO, the polar fraction of the oil samples was extracted and analyzed using UHPLC-ESI-QTOF-MS. This study aimed to characterize the metabolomic profile of DSO for the first time and to compare these findings with previously reported chemical composition data, assessing the effect of geographical origin on the metabolite profile (El Harkaoui et al., 2024). Additionally, marker identification and structure elucidation were key objectives.

Analyses were conducted in both positive and negative electrospray ionization modes to maximize metabolite detection. Overall, more features were observed in the positive mode than in the negative mode, as illustrated in the base peak chromatograms (BPC) of a representative DSO sample (Fig. 3).

The BPCs of ten quality control (QC) samples were overlaid (Fig. S1) and inspected for peak intensity and retention time stability. The high overlap in BPCs across QC samples in both modes indicates the stability of the analyzing method. Following data acquisition, log transformation and Pareto Scaling were applied, and an initial principal component analysis (PCA) was performed. This step ensured the quality of normalization by assessing the clustering of QC samples and the separation of blanks. As expected, QC samples clustered closely with DSO samples, while blanks were distinctly separated (Fig. S2). These findings align with recommended metabolomics guidelines (Broadhurst et al., 2018; Mattoli et al., 2022), confirming the robustness of the analytical method and data pre-processing.

To refine the dataset, statistical data filtering was implemented. Features with a relative standard deviation (RSD) exceeding 20 % in QC samples were removed. Additionally, variables with low variance across experimental conditions were excluded based on interquartile range analysis. Such filtering enhances data quality by eliminating variables unlikely to contribute meaningfully to statistical models (Mattoli et al., 2022). Following data filtration, 446 features remained in the negative mode and 1654 features in the positive mode. The predominance of features in positive mode could be attributed to the ionization efficiency of certain metabolites or more related to the instrumental settings favouring positive ion formation. This trend was consistent with previous metabolomic studies of oils, which frequently reported a higher number of detected features in positive mode (Dou et al., 2025; Gil-Solsona et al., 2016; Hu et al., 2022; Willenberg et al., 2021). The filtered datasets were subsequently subjected to multivariate analysis with log transformation and Pareto scaling, as recommended for metabolomics data (Di Guida et al., 2016; Van den Berg et al., 2006).

3.2. Multivariate analysis

A new PCA was conducted without QC samples to characterize the differentiation among DSO samples from three geographical regions: *Allougoum*, *Alnif*, and *Errachidia* (Fig. 4). As an unsupervised method, PCA enables dimensionality reduction while offering insights into clustering patterns. Overall, the first eight principal components explained over 80 % of total variance in both modes as shown by the scree plots (Fig. S3). PC1 accounted for 65 % of total variance in negative mode and 49 % in positive mode, however, the variance captured by PC1 was not related to geographical origin. The optimal projection for distinguishing the three regions was achieved in the PC2–PC4 plane for negative mode and the PC2–PC3 plane for positive mode (Fig. 4). The PCA score plot revealed no clear separation between samples from *Alnif* and *Errachidia*, which was anticipated given their geographical proximity (Fig. 4) and likely similarities in soil type, environmental conditions, and cultivations practices affecting metabolite composition. However, samples from *Allougoum*, located further

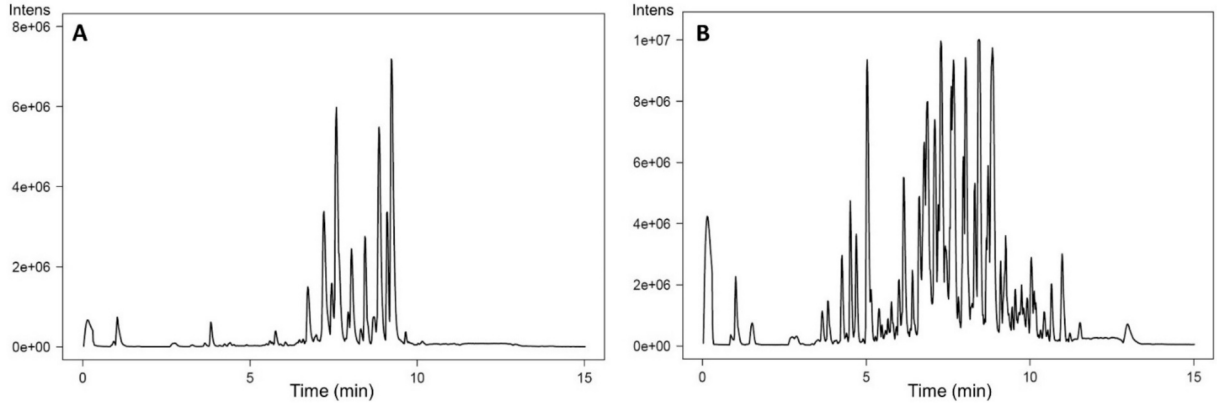


Fig. 3. A&B. Comparison of the base peak chromatograms (BPCs) of the polar extract derived from date seed oil depending on the ionization mode. (A) negative ion mode and (B) positive ion mode, obtained using UHPLC-ESI-QTOF-MS. These chromatograms are representative of a typical sample and illustrate the higher signal complexity and feature richness observed especially for the positive ionization mode.

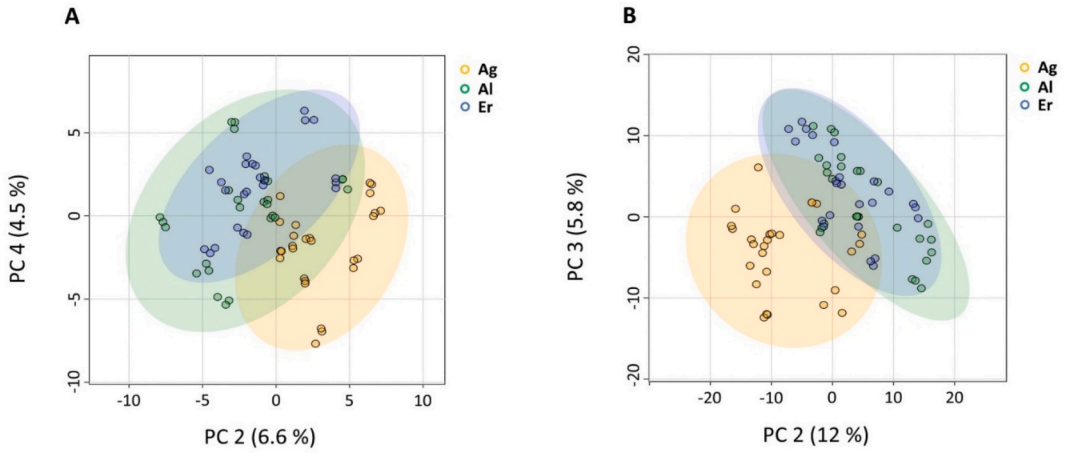


Fig. 4. A&B. PCA Score plot visualization in negative (A) and positive (B) ionization modes comparing samples from the three palm groves: *Allougoum* (Ag), *Alnif* (Al), and *Errachidia* (Er).

south at a lower altitude and closer to the Atlantic Ocean, exhibited a modest clustering trend distinct from the other two locations (Fig. 1 and Fig. 4). This trend was consistent across both ionization modes (Fig. 4) and aligns with previous findings based on chemical composition, where samples from *Allougoum* demonstrated a clustering as compared to the samples from *Alnif* and *Errachidia* (El Harkaoui et al., 2024).

Given this observed trend, orthogonal partial least squares discriminant analysis (OPLS-DA) was applied to identify markers distinguishing *Allougoum* samples from the rest. In this analysis, samples from *Alnif* and *Errachidia* were combined into a single group (AlEr) for comparison. According to established criteria, a valid OPLS-DA model should have Q2 > 0.5, R2Y > 0.7, and a difference between R2Y and Q2 not exceeding 0.2–0.3 (Eriksson et al., 2003). The generated OPLS-DA model exhibited a Q2 value exceeding 0.5 and an R2Y value greater than 0.7, indicating model validity.

To identify key variables contributing to the classification, variable importance projection (VIP) scores were generated for both positive and negative mode datasets. The VIP score summarizes a variable's contribution to the model and is calculated as a weighted sum of the squared correlations between the OPLS-DA components and the original variables (Mohamed et al., 2018). Features were ranked based on their VIP scores, and only the top 50 were selected for further investigation to attempt their identification (Fig. 5 and Table S2 and S3). In addition to the VIP scores, the S-plot from the OPLS-DA model was also used to visualize the relationship between the covariance and correlation of each variable, aiding in the selection of reliable markers. The top 50 features also appeared as significant contributors in the corresponding S-plot (Fig. 6); their distribution in the top-right and bottom-left quadrants confirmed their strong influence in driving the separation between sample groups.

3.3. Compound identification and interpretation

Among the top 50 features, features identified with confidence levels ranging from Level 4 to Level 1 are reported in Tables 1 and 2.

The distribution of the tentatively identified metabolites is visualized

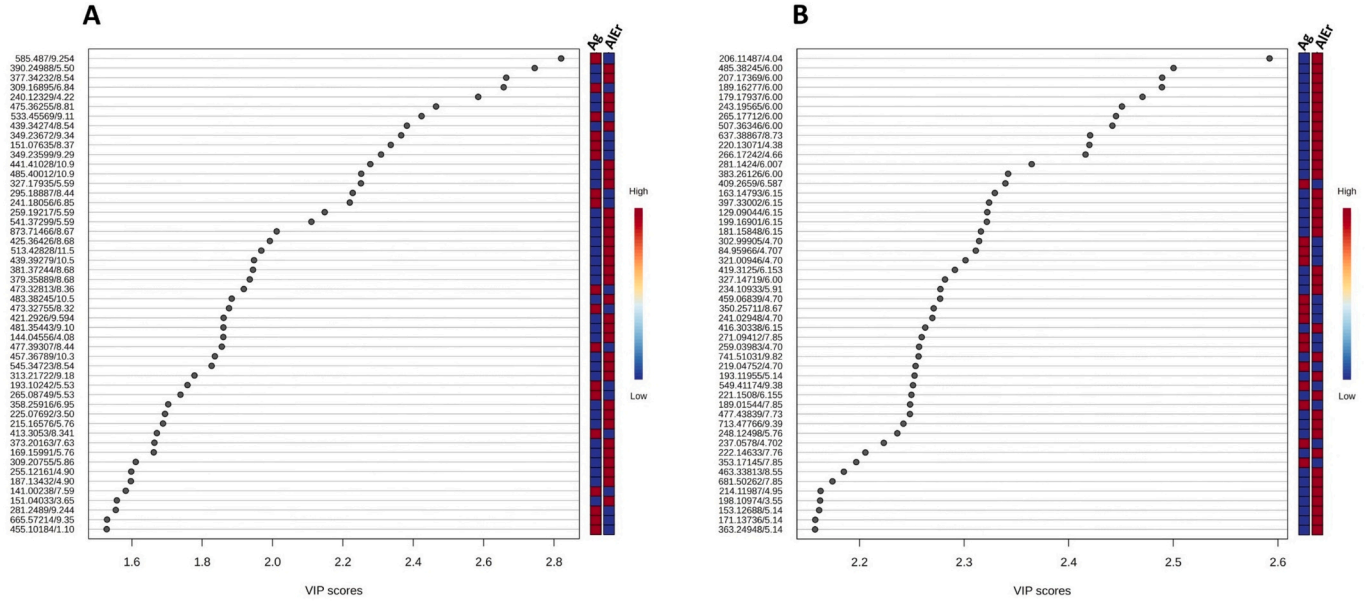


Fig. 5. A&B. Top 50 discriminating features for geographical origin, ranked by their VIP scores, shown for negative ion mode (A) and positive ion mode (B). Samples from *Allougoum* (Ag) are compared with the combined group of *Alnif* and *Errachidia* (AlEr). The color bars represent the median intensity of each feature across the sample groups. Tables S2 and S3 indicate the same top-ranking features with the corresponding ordering numbers (from 1 to 50) and the exact VIP score for each feature.

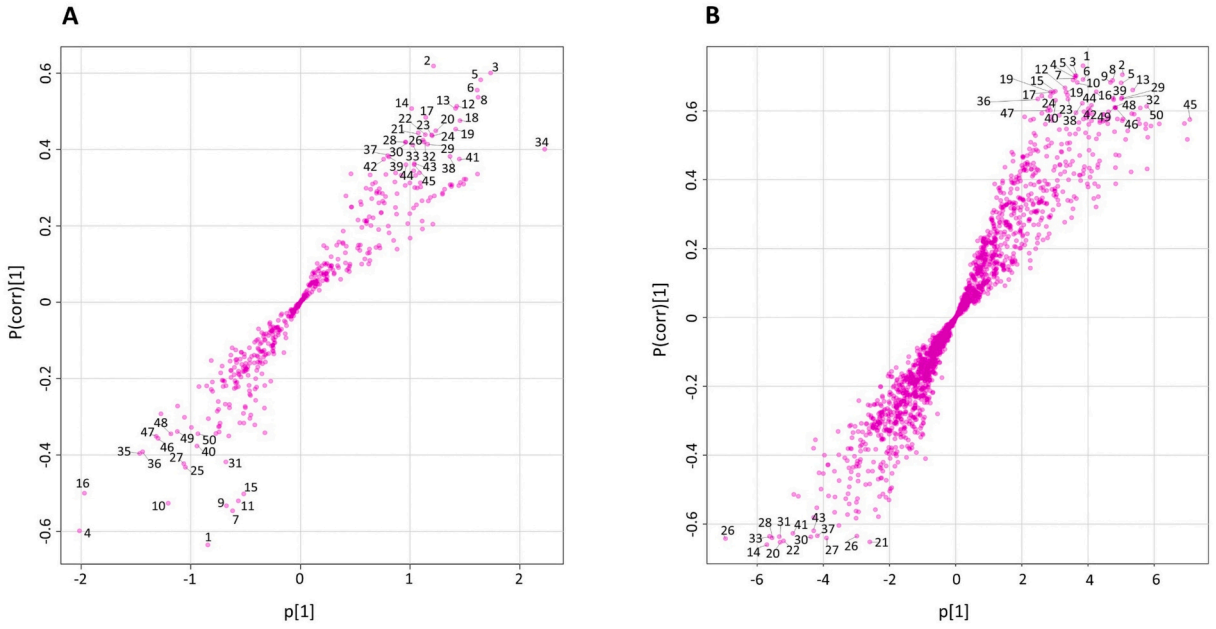


Fig. 6. A&B. S-plots for negative ion mode (A) and positive ion mode (B). The numbered markers (1 to 50) correspond to the top 50 discriminating features ranked by VIP scores, as reported in Fig. 5 and listed in Tables S2 and S3.

Table 1
Tentatively identified markers contributing to the geographical origin discrimination (negative mode).

No.	VIP score	MS (*m/z*)	RT (min)	Adducts	Formula	Δ (ppm)	Compound	Class (Subclass)	Annotation confidence
1	2.820	586.49427	9.25	$[M-H]^-$	$C_{38}H_{66}O_4$	3.1	octacosyl ferulate	cinnamic acids and derivatives (hydroxycinnamic acids and derivatives)	Level 3
2	2.745	391.25715	5.05	$[M-H]^-$	$C_{19}H_{37}NO_7$	−1.0	–	–	Level 4
5	2.585	241.13056	4.22	$[M-H]^-$	$C_{12}H_{19}NO_4$	3.5	n-(3-oxohexanoyl) homoserine lactone	carboxylic acids and derivatives (amino acids and derivatives)	Level 1
6	2.465	476.36982	8.82	$[M-H]^-$	$C_{26}H_{52}O_7$	−3.1	1-O-α-D-glucopyranosyl-1,2-eicosandiol	fatty acyls (fatty acyls glycosides)	Level 3
8	2.382	440.35001	8.55	$[M-H]^-$	$C_{26}H_{48}O_5$	−0.4	momordol	fatty acyls (fatty alcohols)	Level 3
9	2.366	350.24399	9.34	$[M-H]^-$	$C_{16}H_{34}N_2O_6$	5.0	–	–	Level 4
13	2.252	486.40739	10.91	$[M-H]^-$	$C_{32}H_{54}O_3$	−0.9	–	–	Level 4
16	2.219	242.18783	6.85	$[M-H]^-$	$C_{14}H_{26}O_3$	−1.5	hydroxy tetradecenoic acid	fatty acyls (hydroxy fatty acids)	Level 3
17	2.148	260.19944	5.59	$[M-H]^-$	$C_{14}H_{28}O_4$	2.6	dihydroxy tetradecanoic acid	fatty acyls (hydroxy fatty acids)	Level 3
20	1.992	426.37153	8.68	$[M-H]^-$	$C_{26}H_{50}O_4$	0.2	–	–	Level 4
21	1.968	514.43555	11.59	$[M-H]^-$	$C_{34}H_{58}O_3$	−5.9	–	–	Level 4
25	1.918	474.33540	8.37	$[M-H]^-$	$C_{29}H_{46}O_5$	0.7	–	–	Level 4
29	1.860	482.36170	9.10	$[M-H]^-$	$C_{28}H_{50}O_6$	2.0	certonardosterol M	steroids and steroid derivatives (bile acids, alcohols and derivatives)	Level 3
30	1.860	145.05283	4.08	$[M-H]^-$	C_9H_7NO	0.5	indole-3-carboxaldehyde	indoles and derivatives (Indoles)	Level 1
32	1.836	458.37516	10.33	$[M-H]^-$	$C_{30}H_{50}O_3$	−3.0	–	–	Level 4
36	1.738	266.09476	5.53	$[M-H]^-$	$C_{17}H_{14}O_3$	1.8	7-methoxy-2-methylisoflavone	isoflavonoids (*o*-methylated isoflavonoids)	Level 3
37	1.703	359.26643	6.96	$[M-H]^-$	$C_{19}H_{37}NO_5$	−3.0	–	–	Level 4
39	1.689	216.17303	5.77	$[M-H]^-$	$C_{12}H_{24}O_3$	2.2	12-hydroxydodecanoic	fatty Acyls (hydroxy fatty acids)	Level 2
47	1.557	152.04760	3.65	$[M-H]^-$	$C_8H_8O_3$	1.68	vanillin	phenols	Level 1

The class of the compounds were attributed via ClassyFire and LipidMaps; No: the order of the compounds in the list of 50 top ranked features reported in Table S2 and Fig. 4. The MS/MS spectra for compounds identified at Level 1 are provided in Figs. S4–S6, and the main MS/MS fragments for compounds identified at Level 2 are summarized in Table S4.

in box plots (Fig. 7), showing their presence across all DSO samples. This suggests that differentiation between geographical origins is driven more by concentration differences than by the presence or absence of specific metabolites, which was expected, as all samples belong to the same type of oil (DSO). The concentration of individual metabolites varied, with some being higher in *Allougoum* and others in the combined *Alnif-Errachidia* group. Although precise discrimination thresholds for practical authenticity testing cannot be defined yet, the observed differences in relative intensities across the regions, as illustrated by box plots, suggest preliminary ranges that could guide future targeted

Table 2
Tentatively identified markers contributing to the geographical origin discrimination (positive mode).

No.	VIP score	MS (m/z)	RT (min)	Adducts	Formula	Δ (ppm)	Compounds	Class (Subclass)	Annotation confidence
2	2.500	484.37499	6.00	$[M+H]^+$	$C_{28}H_{52}O_6$	−4.0	–	–	Level 4
3	2.489	206.16689	6.00	$[M+H]^+$	$C_{14}H_{22}O$	−3.5	–	–	Level 4
16	2.324	396.32427	6.15	$[M+H]^+$	$C_{24}H_{44}O_4$	0.8	dihydroxy tetracosadienoic acid	fatty Acyls (hydroxy fatty acids)	Level 3
18	2.322	198.16224	6.15	$[M+H]^+$	$C_{12}H_{22}O_2$	−1.3	δ-dodecalactone	lactones (delta valerolactones)	Level 2
19	2.316	180.1513	6.15	$[M+H]^+$	$C_{12}H_{20}O$	0.6	2,4-dodecadienal	fatty acyls (fatty aldehydes)	Level 2
25	2.277	233.10307	5.91	$[M+H]^+$	$C_{10}H_{17}O_6$	0.0	–	–	Level 4

The class of the compounds were attributed via ClassyFire and LipidMaps; No: the order of the compounds in the list of 50 top ranked features reported in Table S3 and Fig. 4. The MS/MS spectra for compounds identified at Level 1 are provided in Figs. S4–S6, and the main MS/MS fragments for compounds identified at Level 2 are summarized in Table S4.

investigations.

The box plots and PCA plots (Figs. 4 and 7) also revealed a wide dispersion of samples from the same provenance within the cluster. As all samples were collected at the same maturity stage (*Tamr* stage) and processed under identical conditions, including oil extraction, and storage, this intra-regional variability may also reflect the genetic background of the date varieties included in the sampling. A similar pattern was also reported in a previous publication (El Harkaoui et al., 2024). However, disentangling these effects requires a larger dataset with a broader representation of varieties within each region to allow robust statistical evaluation. It should be noted that in Morocco, literature reports more than 200 date varieties, which highlights the challenge of fully assessing varietal effects (Bouhlali, Ramchoun, et al., 2017).

Among the markers, the compound with the highest VIP score was octacosyl ferulate, a hydroxycinnamic acid derivative. This compound has not been previously reported in oils, but its chemical structure suggests that it may result from the esterification of ferulic acid, previously detected in DSO (Hamza et al., 2021; Harkat et al., 2022), with octacosanol, a policosanol commonly found in vegetable oils (Jung et al., 2011). The presence of related ferulate-based compounds, such as γ-oryzanol in vegetable oils (Cuevas et al., 2017), further supports this tentative structure suggestion. The variation of octacosyl ferulate across geographical origins could be influenced by the concentrations of both ferulic acid and octacosanol. However, confirming this compound's identity requires analysis using an authentic standard under identical experimental conditions. Additionally, n-(3-oxohexanoyl) homoserine lactone, indole-3-carboxaldehyde, and vanillin were confirmed by a comparison with authentic standards, with their extracted ion chromatograms (EICs) and corresponding MS/MS spectra presented in Fig. S4-S6. Indole-3-carboxaldehyde, commonly found in cruciferous vegetables (Palladino et al., 2024), was tentatively identified but not confirmed in corn oil, (Alberdi-Cedeño et al., 2017). Vanillin, another key marker, was previously identified as the second most abundant phenolic compound in Algerian DSO, with variations depending on variety (Harkat et al., 2022). Phenolic compounds are well-known markers for geographical origin, as demonstrated in studies on olive oils (Ghisoni et al., 2019; Kalogiouri et al., 2018; Mohamed et al., 2018; Olmo-García et al., 2019). In the present study, vanillin was the only phenolic compound among the top 50 features.

Furthermore, hydroxy fatty acids, a subclass of fatty acyls, were identified as geographical markers for DSO. Compounds such as hydroxy tetradecenoic acid, dihydroxy tetradecanoic acid, 12-hydroxydodecanoic acid, and dihydroxy tetracosadienoic acid were among the most prominent markers. A metabolomic study on Spanish extra virgin olive oil tentatively identified 9,13-dihydroxy-11-octadecenoic acid as one of the significant markers for origin discrimination (Gil-Solsona et al., 2016), further underscoring the potential of hydroxy fatty acids as geographical markers. Additionally relevance of hydroxy fatty acids in general as authenticity markers has been also emphasized in a recent study (Koch et al., 2022). Hydroxy fatty acids are widespread in plant genera such as *Apocynaceae*, *Asteraceae*, *Brassicaceae*, *Coriariaceae*, *Euphorbiaceae*, *Fabaceae*, *Malpighiaceae*, and *Papaveraceae*. The biosynthesis of hydroxy fatty acids can be attributed to desaturases enzymes, which, beyond desaturation, can also catalyze hydroxylation due to mutations that alter substrate interactions (Cahoon & Li-Beisson, 2020). For example, a variant of the FAD2 enzyme converts oleic acid (D9–18:1) into ricinoleic acid (12-OH-D9–18:1), the main component of castor oil, by adding a hydroxyl (-OH) group instead of a double bond forming linoleic fatty acid (Schmid, 2021). In the present study, a C24 dihydroxy fatty acid (dihydroxy tetracosadienoic acid) was tentatively identified, whose elemental formula corresponds to wuhanic acid. However, due to the absence of an authentic standard and reference MS/MS spectra in the literature, it was assigned a Level 3 identification. Evidence supporting the biosynthesis of wuhanic acids comes from studies involving the transgenic expression of candidate biosynthetic enzymes in *Arabidopsis* spp. These enzymes were identified from the transcriptome of developing seeds of Chinese violet cress (*Orychophragmus violaceus*). The findings revealed that the C-18 hydroxyl group is introduced by a FAD2-type hydroxylase, while the C-7 hydroxyl group originates from a divergent FAE1-encoded 3-ketoacyl-CoA synthase (KCS), operating through a unique biosynthetic mechanism referred to as "discontinuous elongation" (Li et al., 2018). This may suggest that fatty acids and hydroxy fatty acids may share common enzymatic pathways. Following this hypothesis, the hydroxy fatty acids identified in the present study correspond to C12, C14, and C24, which aligns, in terms of carbon chain length, with fatty acids identified in a previous study (El Harkaoui et al., 2024). Specifically, lauric acid (C12:0), myristic acid (C14:0), and lignoceric acid (C24:0) were previously reported, suggesting that the hydroxy fatty acids may originate enzymatically, as discussed earlier. Environmental factors influencing enzyme activity in fatty acid synthesis have been well documented (Ghaffari et al., 2023; Porokhovinova et al., 2022; Sidorov & Tsydendambaev, 2014), and such factors may also affect the variation of hydroxy fatty acids. For example, fluctuations in ricinoleic acid in castor oil have been observed in response to irrigation and water deficit (Ramanjaneyulu et al., 2013), supporting the hypothesis that the variation of hydroxy fatty acids in DSO could be influenced by environmental conditions.

Interpreting the chemical metabolites responsible for the geographical variation of DSO remains challenging, especially as the present study is the first to investigate its metabolomic profile. While fats and oils are generally less influenced by geographical origin compared to other foods, impact of the origin has been demonstrated in various types of oil (Lucini et al., 2020). However, in the context of untargeted metabolomics, the relationship between geographical location, defined ecophysiological factors, and the resulting metabolite profile is rarely discussed in detail, making comparative analysis difficult. This challenge is even more pronounced for less-studied oils like DSO, where data is still limited.

On the other hand, previous metabolomic studies already

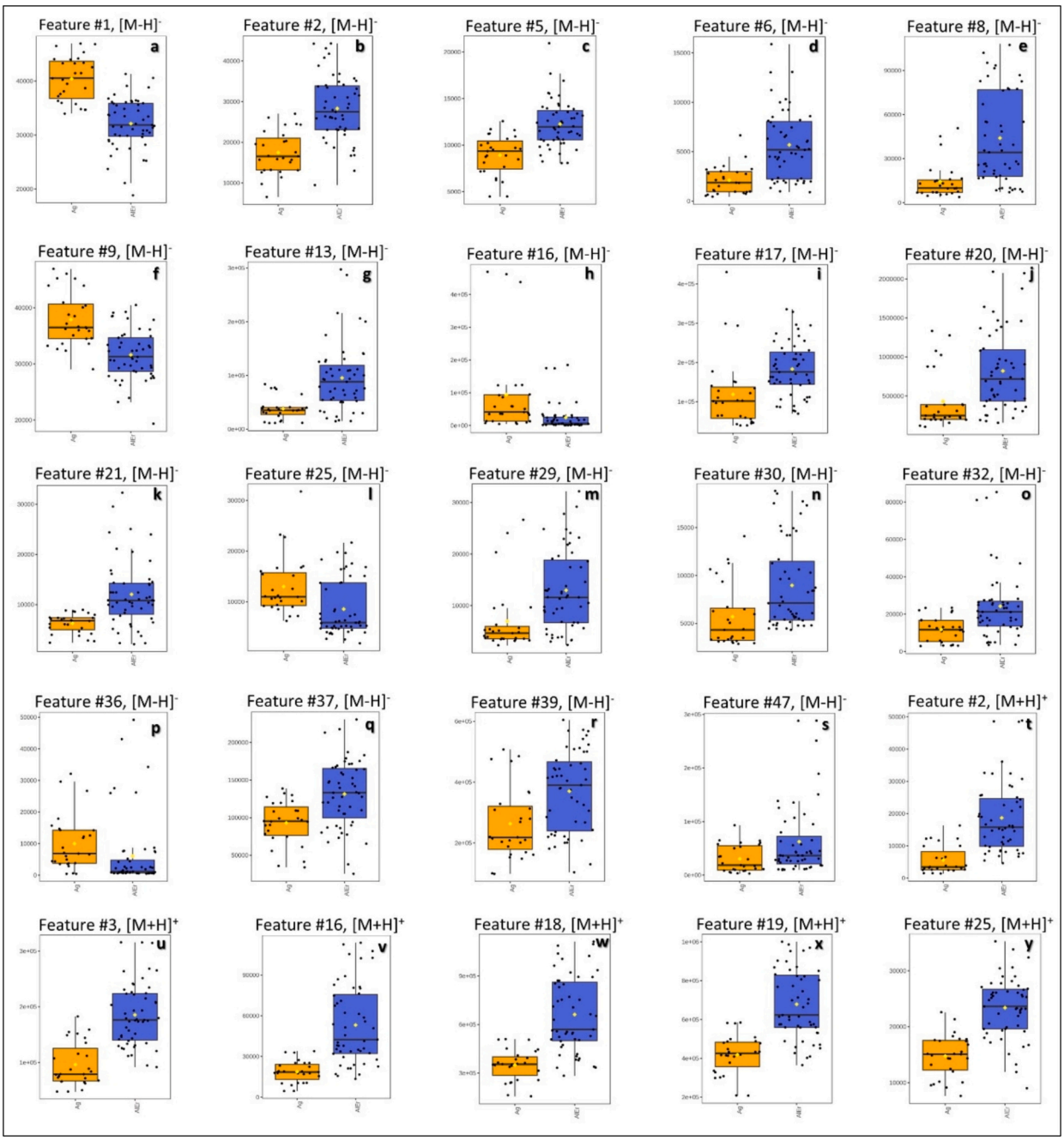


Fig. 7. a-y. Variation of the 25 tentatively identified metabolites (listed in Tables 1 and 2) detected in both negative and positive ionization modes. These 25 features are part of the top 50 most significant features (ranked 1–50) based on their VIP scores. Each subfigure is labeled with the corresponding feature number (feature #number), matching the markers in Tables 1 and 2, and also indicates the ionization mode: negative ($[M + H]^-$) or positive ($[M + H]^+$). Samples from *Allougoum* (Ag, orange) are compared with the combined *Alnif* and *Errachidia* group (AlEr, blue).

characterized date palm fruits, seeds, and pollen, but they did not specifically address the influence of geographical origin. When comparing the molecular formula reported in these studies, it was found that compounds with the formulas $C_{18}H_{30}O_3$ (tentatively identified as hydroxyoctadecatrienoic acid), $C_7H_6O_4$ (tentatively identified as gentisic acid), C10H10O4 (tentatively identified as ferulic acid), and $C_9H_{10}O_4$ (tentatively identified as homovanillic acid) occurred also in the present dataset (Alsuhaymi et al., 2023; Harkat et al., 2022; M AbouZeid et al., 2022). However, these compounds did not rank among the top 50 features in the present analysis (according to the VIP scores), implying that they are not relevant for the classification of a geographical origin for these samples.

4. Conclusion

The present study investigated the effect of geographical location on the metabolomic profile of Moroccan date seed oil (DSO) using an untargeted approach with UHPLC-QTOF-MS. Samples from three Moroccan palm groves (*Allougoum*, *Alnif*, and *Errachidia*) were analyzed, revealing a modest clustering trend linked to geographical origin. These results support the potential of untargeted metabolomics, combined with chemometric modelling, as a powerful tool for origin verification of DSO. Among the top 50 discriminating features identified, 25 metabolites from various chemical classes were tentatively annotated at different confidence levels. Notably, these compounds are reported for

the first time in DSO, expanding current knowledge of its chemical complexity. All identified metabolites were present across all samples, indicating that origin-related differentiation is primarily driven by differences in concentration rather than the presence or absence of unique metabolites. The ability to distinguish DSO based on geographical origin adds value to this edible, unconventional and maybe highly valuable oil and supports, as a perspective, the development of region-specific labels. This authenticity aspect will become even more important as DSO continues to gain attention as a source of (edible) oil in arid regions, especially when certain attributes are more strongly associated with oils from specific areas. On the other hand, although the most pronounced variation was observed in samples from *Allougoum*, it should be acknowledged that the overall differences between regions were moderate. This may reflect a relatively stable compositional profile of Moroccan DSO across the studied palm groves. However, further research is needed since this study included only samples from a single crop year and within one country. Future studies should include multi-year sampling and comparisons with DSO from other major date-producing countries, such as Saudi Arabia and the UAE, to better understand how geographical origin influences DSO composition and to validate this initial, promising approach.

CRediT authorship contribution statement

Said El Harkaoui: Writing – original draft, Visualization, Project administration, Investigation, Formal analysis, Conceptualization. **Katharina N'Diaye:** Writing – review & editing, Methodology. **Zoubida Charrouf:** Writing – review & editing, Supervision, Conceptualization. **Sascha Rohn:** Writing – review & editing, Validation, Supervision. **Stephan Drusch:** Writing – review & editing, Validation, Supervision. **Bertrand Matthäus:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funds of the Federal Ministry of Agriculture, Food and Regional Identity BMLEH based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) [support program: FKZ 2819DOKA03]. The authors gratefully acknowledge Prof. Said Gharby (Polydisciplinary Faculty of Taroudant, Ibn Zohr University, Agadir, Morocco) for his valuable support and assistance with sample collection. We also thank Mr. Ralph Schuster (Max Rubner Institute, Detmold, Germany) for his technical support with the UHPLC-ESI-QTOF-MS analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2025.103162.

Data availability

Data will be made available on request.

References

Abdul-Hamid, N. A., Abas, F., Ismail, I. S., Tham, C. L., Maulidiani, M., Mediani, A., … Zolkeflee, N. K. Z. (2019). Metabolites and biological activities of *Phoenix dactylifera* L. pulp and seeds: A comparative MS and NMR based metabolomics approach. *Phytochemistry Letters, 31*, 20–32. https://doi.org/10.1016/j.phytol.2019.03.004

AbouZeid, M., Afifi, A. E. H., Salama, A., Hussein, R., Youssef, F., El-Ahmady, S. H., & Mohamed Ammar, N. (2022). Comprehensive metabolite profiling of Phoenix rupicola pulp and seeds using UPLC-ESI-MS/MS and evaluation of their estrogenic activity in ovariectomized rat model. *Food Research International (Ottawa, Ont.), 157*, Article 111308. https://doi.org/10.1016/j.foodres.2022.111308

Al Juhaimi, F., Özcan, M. M., Adiamo, O. Q., Alsawmahi, O. N., Ghafoor, K., & [K.], & Babiker, E. E.. (2018). Effect of date varieties on physico-chemical properties, fatty acid composition, tocopherol contents, and phenolic compounds of some date seed and oils. *Journal of Food Processing and Preservation, 42*(4), Article e13584. https://doi.org/10.1111/jfpp.13584

Alberdi-Cedeño, J., Ibargoitia, M. L., & Guillén, M. D. (2017). Bioactive compounds detected for the first time in corn oil: Cyclic dipeptides and other nitrogenated compounds. *Journal of Food Composition and Analysis, 62*, 197–204. https://doi.org/10.1016/j.jfca.2017.06.005

Alkhoori, M. A., Kong, A. S.-Y., Aljaafari, M. N., Abushelaibi, A., Erin Lim, S.-H., Cheng, W.-H., … Lai, K.-S. (2022). Biochemical composition and biological activities of date palm (*Phoenix dactylifera* L.) seeds: A review. *Biomolecules, 12*(11). https://doi.org/10.3390/biom12111626

Alsuhaymi, S., Singh, U., Al-Younis, I., Kharbatia, N. M., Haneef, A., Chandra, K., … Jaremko, M. (2023). Untargeted metabolomics analysis of four date palm (*Phoenix dactylifera* L.) cultivars using MS and NMR. *Natural Products and Bioprospecting, 13*(1), 44. https://doi.org/10.1007/s13659-023-00406-y

Bouhlali, E. D. T., Alem, C., Ennassir, J., Benlyas, M., Mbark, A. N., & Zegzouti, Y. F. (2017). Phytochemical compositions and antioxidant capacity of three date (*Phoenix dactylifera* L.) seeds varieties grown in the South East Morocco. *Journal of the Saudi Society of. Agricultural Sciences, 16*(4), 350–357. https://doi.org/10.1016/j.jssas.2015.11.002

Bouhlali, E. D. T., Ramchoun, M., Alem, C., Ghafoor, K., Kashif, Ennassir, J., & Zegzouti, Y. F. (2017). Functional composition and antioxidant activities of eight Moroccan date fruit varieties (*Phoenix dactylifera* L.). *Journal of the Saudi Society of Agricultural Sciences, 16*(3), 257–264. https://doi.org/10.1016/j.jssas.2015.08.005

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics, 14*(6), 72. https://doi.org/10.1007/s11306-018-1367-3

Cahoon, E. B., & Li-Beisson, Y. (2020). Plant unusual fatty acids: Learning from the less common. *Current Opinion in Plant Biology, 55*, 66–73. https://doi.org/10.1016/j.pbi.2020.03.007

Caporale, G., & Monteleone, E. (2001). Effect of expectations induced by information on origin and its guarantee on the acceptability of a traditional food: olive oil. *Sciences des Aliments, 21*(3), 243–254. https://doi.org/10.3166/sda.21.243-254

Cuevas, M. S., Souza, P. T., Da Costa Rodrigues, C. E., & Meirelles, A. J. A. (2017). Quantification and Determination of Composition of Steryl Ferulates in Refined Rice Bran Oils Using an UPLC-MS Method. *Journal of the American Oil Chemists' Society, 94*(3), 375–385. https://doi.org/10.1007/s11746-017-2955-5

DGF. (2021). *Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten, Tensiden und verwandten Stoffen*. Stuttgart, Germany: Wissenschaftliche Verlagsgesellschaft.

Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., … Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: A comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics, 12*(5), 93. https://doi.org/10.1007/s11306-016-1030-9

Dou, X., N'Diaye, K., El Harkaoui, S., Willenberg, I., Ma, F., Zhang, L., … Matthäus, B. (2025). Authentication of Virgin Olive Oil Based on Untargeted Metabolomics and Chemical Markers. *European Journal of Lipid Science and Technology, 127*(1), Article e202400126. https://doi.org/10.1002/ejlt.202400126

Echegaray, N., Gullón, B., Pateiro, M., Amarowicz, R., Misihairabgwi, J. M., & Lorenzo, J. M. (2021). Date Fruit and Its By-products as Promising Source of Bioactive Components: A Review. *Food Reviews International, 39*(3), 1–22. https://doi.org/10.1080/87559129.2021.1934003

El Harkaoui, S., N'Diaye, K., Gharby, S., Al-Hilal, M., Charrouf, Z., Rohn, S., & Matthäus, B. (2024). Insights into date seed oil composition: Geographical variability and potential applications. *European Journal of Lipid Science and Technology, 126*(7), Article 2400061. https://doi.org/10.1002/ejlt.202400061

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives, 111*(10), 1361–1375. https://doi.org/10.1289/ehp.5758

Farag, M. A., Otify, A., & Baky, M. H. (2021). *Phoenix Dactylifera* L. Date Fruit By-products Outgoing and Potential Novel Trends of Phytochemical, Nutritive and Medicinal Merits. *Food Reviews International*, 1–23. https://doi.org/10.1080/87559129.2021.1918148

Ghaffari, M., Gholizadeh, A., Rauf, S., & Shariati, F. (2023). Drought-stress induced changes of fatty acid composition affecting sunflower grain yield and oil quality. *Food Science & Nutrition*. https://doi.org/10.1002/fsn3.3690. Advance online publication.

Ghisoni, S., Lucini, L., Angilletta, F., Rocchetti, G., Farinelli, D., Tombesi, S., & Trevisan, M. (2019). Discrimination of extra-virgin-olive oils from different cultivars and geographical origins by untargeted metabolomics. *Food Research International (Ottawa, Ont.), 121*, 746–753. https://doi.org/10.1016/j.foodres.2018.12.052

Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, R., Ibáñez, M., … Hernández, F. J. (2016). Metabolomic approach for Extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-of-

flight mass spectrometry. *Food Control, 70*, 350–359. https://doi.org/10.1016/j.foodcont.2016.06.008

Habib, H. M., Kamal, H., Ibrahim, W. H., & Dhaheri, A. S. A. (2013). Carotenoids, fat soluble vitamins and fatty acid profiles of 18 varieties of date seed oil. *Industrial Crops and Products, 42*, 567–572. https://doi.org/10.1016/j.indcrop.2012.06.039

Hamza, H., Elfalleh, W., & Nagaz, K. (2021). Date Palm Seed Oil (*Phoenix dactylifera* L.) Green Extraction: Physicochemical Properties, Antioxidant Activities, and Phenolic and Fatty Acid Profiles. *Journal of Food Quality, 2021*(1), 1–9. https://doi.org/10.1155/2021/2394220

Harkat, H., Bousba, R., Benincasa, C., Atrouz, K., Gültekin-Özgüven, M., Altuntaş, Ü., … Özçelik, B. (2022). Assessment of biochemical composition and antioxidant properties of algerian date palm (*Phoenix dactylifera* L.) Seed Oil. *Plants, 11*(3), 381. https://doi.org/10.3390/plants11030381

Hu, Q., Zhang, J., Xing, R., Yu, N., & Chen, Y. (2022). Integration of lipidomics and metabolomics for the authentication of camellia oil by ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry coupled with chemometrics. *Food Chemistry, 373*(Pt B), Article 131534. https://doi.org/10.1016/j.foodchem.2021.131534

Jung, D. M., Lee, M. J., Yoon, S. H., & Jung, M. Y. (2011). A gas chromatography-tandem quadrupole mass spectrometric analysis of policosanols in commercial vegetable oils. *Journal of Food Science, 76*(6), C891–C899. https://doi.org/10.1111/j.1750-3841.2011.02232.x

Kalogiouri, N. P., Aalizadeh, R., & Thomaidis, N. S. (2018). Application of an advanced and wide scope non-target screening workflow with LC-ESI-QTOF-MS and chemometrics for the classification of the Greek olive oil varieties. *Food Chemistry, 256*, 53–61. https://doi.org/10.1016/j.foodchem.2018.02.101

Koch, E., Wiebel, M., Löwen, A., Willenberg, I., & Schebb, N. H. (2022). Characterization of the Oxylipin Pattern and Other Fatty Acid Oxidation Products in Freshly Pressed and Stored Plant Oils. *Journal of Agricultural and Food Chemistry, 70*(40), 12935–12945. https://doi.org/10.1021/acs.jafc.2c04987

Li, X., Teitgen, A. M., Shirani, A., Ling, J., Busta, L., Cahoon, R. E., … Cahoon, E. B. (2018). Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. *Nature Plants, 4*(9), 711–720. https://doi.org/10.1038/s41477-018-0225-7

Lieb, V. M., Kleiber, C., Metwali, E. M., Kadasa, N. M., Almaghrabi, O. A., Steingass, C. B., & Carle, R. (2020). Fatty acids and triacylglycerols in the seed oils of Saudi Arabian date (*Phoenix dactylifera* L.) palms. *International Journal of Food Science & Technology, 55*(4), 1572–1577. https://doi.org/10.1111/ijfs.14383

Lucini, L., Rocchetti, G., & Trevisan, M. (2020). Extending the concept of terroir from grapes to other agricultural commodities: an overview. *Current Opinion in Food Science, 31*, 88–95. https://doi.org/10.1016/j.cofs.2020.03.007

Maqsood, S., Adiamo, O., Ahmad, M., & Mudgil, P. (2020). Bioactive compounds from date fruit and seed as potential nutraceutical and functional food ingredients. *Food Chemistry, 308*, Article 125522. https://doi.org/10.1016/j.foodchem.2019.125522

Mattoli, L., Gianni, M., & Burico, M. (2022). Mass spectrometry-based metabolomic analysis as a tool for quality control of natural complex products. *Mass Spectrometry Reviews, e21773*. https://doi.org/10.1002/mas.21773

Mohamed, M. B., Rocchetti, G., Montesano, D., Ali, S. B., Guasmi, F., Grati-Kamoun, N., & Lucini, L. (2018). Discrimination of Tunisian and Italian extra-virgin olive oils according to their phenolic and sterolic fingerprints. *Food Research International (Ottawa, Ont.), 106*, 920–927. https://doi.org/10.1016/j.foodres.2018.02.010

Mrabet, A., Jiménez-Araujo, A., Guillén-Bejarano, R., Rodríguez-Arcos, R., & Sindic, M. (2020). Date seeds: A promising source of oil with functional properties. *Foods, 9*(6), 787. https://doi.org/10.3390/foods9060787

Nehdi, I. A., Sbihi, H. M., Tan, C. P., Rashid, U., & Al-Resayes, S. I. (2018). Chemical composition of date palm (*Phoenix dactylifera* L.) seed oil from six saudi arabian cultivars. *Journal of Food Science, 83*(3), 624–630. https://doi.org/10.1111/1750-3841.14033

Olmo-García, L., Wendt, K., Kessler, N., Bajoub, A., Fernández-Gutiérrez, A., Baessmann, C., & Carrasco-Pancorbo, A. (2019). Exploring the capability of LC-MS and GC-MS multi-class methods to discriminate virgin olive oils from different geographical indications and to identify potential Origin markers. *European Journal of Lipid Science and Technology, 121*(3), Article 1800336. https://doi.org/10.1002/ejlt.201800336

Palladino, P., Attanasio, L., Scarano, S., Degano, I., & Minunni, M. (2024). Colorimetric determination of indole-3-carbaldehyde by reaction with carbidopa and formation of aldazine in ethanolic extract of cabbage. *Food Chemistry Advances, 4*, Article 100643. https://doi.org/10.1016/j.focha.2024.100643

Pang, Z., Chong, J., Zhou, G., Lima Morais, D. A., Chang, L., Barrette, M., … Xia, J. (2021). Metaboanalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research, 49*(W1), W388–W396. https://doi.org/10.1093/nar/gkab382

Porokhovinova, E. A., Matveeva, T. V., Khafizova, G. V., Bemova, V. D., Doubovskaya, A. G., Kishlyan, N. V., … Gavrilova, V. A. (2022). Fatty acid composition of oil crops: Genetics and genetic engineering. *Genetic Resources and Crop Evolution, 69*(6), 2029–2045. https://doi.org/10.1007/s10722-022-01391-w

Qian, Y., Rudzińska, M., Grygier, A., & Przybylski, R. (2020). Determination of triacylglycerols by HTGC-FID as a sensitive tool for the identification of rapeseed and olive oil adulteration. *Molecules, 25*(17), 3881. https://doi.org/10.3390/molecules25173881

Ramanjaneyulu, A. V., Reddy, A. V., & Madhavi, A. (2013). The impact of sowing date and irrigation regime on castor (*Ricinus communis* L.) seed yield, oil quality characteristics and fatty acid composition during post rainy season in South India. *Industrial Crops and Products, 44*, 25–31. https://doi.org/10.1016/j.indcrop.2012.10.008

Schmid, K. M. (2021). Lipid metabolism in plants. In N. D. Ridgway, & R. S. McLeod (Eds.), *Biochemistry of Lipids, Lipoproteins and Membranes* ((Seventh Edition), pp. 121–159). Elsevier. https://doi.org/10.1016/B978-0-12-824048-9.00011-0.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science & Technology, 48*(4), 2097–2098. https://doi.org/10.1021/es5002105

Sidorov, R. A., & Tsydendambaev, V. D. (2014). Biosynthesis of fatty oils in higher plants. *Russian Journal of Plant Physiology, 61*(1), 1–18. https://doi.org/10.1134/S1021443714010130

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., … Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics, 3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

Van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142. https://doi.org/10.1186/1471-2164-7-142

Willenberg, I., Parma, A., Bonte, A., & Matthäus, B. (2021). Development of chemometric models based on a LC-qToF-MS approach to verify the geographic origin of virgin olive oil. *Foods, 10*(2), 479. https://doi.org/10.3390/foods10020479

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta, 914*, 17–34. https://doi.org/10.1016/j.aca.2016.02.001

Manuscript V

Adulteration detection in cactus seed oil: Integrating analytical chemistry and machine learning approaches

Said El Harkaoui, Cristina Ortiz Cruz, Aaron Roggenland, Micha Schneider, Sascha Rohn, Stephan Drusch, Bertrand Matthäus

Current Research in Food Science, 2025, 10, 100986

Accepted manuscript available via https://doi.org/10.1016/j.crfs.2025.100986

Note:

As part of this manuscript, three associated data publications have been made available open access through the OpenAgrar repository:
https://doi.org/10.25826/Data20240930-105507-0
https://doi.org/10.25826/Data20240930-114743-0
https://doi.org/10.25826/Data20240930-113217-0

Additionally, the full Python script is openly accessible on GitHub:
https://github.com/kida4bmel/oil-adulteration

Current Research in Food Science 10 (2025) 100986

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Adulteration detection in cactus seed oil: Integrating analytical chemistry and machine learning approaches

Said El Harkaoui [a,b,c,*], Cristina Ortiz Cruz [d,g], Aaron Roggenland [e,g], Micha Schneider [f,g], Sascha Rohn [b], Stephan Drusch [c], Bertrand Matthäus [a]

[a] Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Schützenberg 12, 32756, Detmold, Germany
[b] Department of Food Chemistry and Analysis, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany
[c] Department of Food Technology and Food Material Science, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Berlin, Germany
[d] Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Zentralabteilung, Haid-und-Neu-Str. 9, 76131, Karlsruhe, Germany
[e] Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Zentralabteilung, Schützenberg 12, 32756, Detmold, Germany
[f] Johann Heinrich von Thünen Institute - Federal Research Institute for Rural Areas, Forestry and Fisheries, Bundesallee 50, 38116, Braunschweig, Germany
[g] BMEL Project KIDA, AI consultancy, Germany

ARTICLE INFO

Handling Editor: Dr. Maria Corradini

Keywords:
Cactus seed oil
Authenticity
Machine learning
Conditional generative adversarial network
Monte-Carlo
Random Forest
Neural network

ABSTRACT

Economically motivated adulteration threatens both consumer rights and market integrity, particularly with high-value cold-pressed oils like cactus seed oil (CO). This study proposes a machine learning model that integrates analytical measurements, data simulations, and classification techniques to detect adulteration of CO with refined sunflower oil (SO) and determine the detectable limit of adulteration without measuring a huge number of different mixtures. First, pure CO and SO samples were analyzed for their fatty acid, triacylglycerol, and tocochromanol content using HPLC or GC. The resulting oil composition data served as the foundation for further simulations. Monte Carlo (MC) simulations outperformed Conditional Tabular Generative Adversarial Networks (CTGAN) in simulating realistic oil compositions, with MC yielding lower Kullback-Leibler Divergence values compared to CTGAN. The MC-simulated data were then used to simulate larger datasets, a critical step for training and testing two classification models: Random Forest (RF) and Neural Networks (NN), as robust training cannot be achieved with small sample sizes. Both models achieved good classification accuracies, with RF achieving higher accuracy than NN, reaching 94% on simulated datasets and 90% on real-world samples with detectable adulteration levels as low as 1%. RF also offers better interpretability and is computational less demanding as compared to NN which makes it advantageous for authenticity verification in this study. Therefore, combining MC simulation with RF as a robust method for detecting CO adulteration is proposed. The proposed method, coded in Python and available as open-source, offers a flexible framework for continuous adaptation with new data.

1. Introduction

Economically motivated adulteration poses a significant threat to consumer rights and market integrity, with cold-pressed vegetable oils being particularly susceptible to fraudulent practices (Dou et al., 2023; Yuan et al., 2020). Among these oils, expensive varieties such as cactus seed oil (CO) are more prone to adulteration. Detecting and preventing such fraudulent practices is essential not only for maintaining consumer trust and safeguarding fair trade practices but also for protecting public health and supporting robust regulatory frameworks.

Initially, CO is extracted from the seeds of the resilient cactus plants of the genus *Opuntia*. This oil provides a rich profile of essential fatty acids, antioxidants, and vitamins, making it a potential source of edible oil, particularly in arid regions where the cactus thrives (Nounah et al., 2024; Chbani et al., 2023). CO is characterized by its high levels of linoleic acid and abundant gamma-tocopherol content (Nounah et al., 2024). Additionally, its polyphenol composition includes significant amounts of vanillin, syringaldehyde, and furaldehyde (Chbani et al., 2020). The olfactometric analysis further identifies hexanal, 2-methyl propanal, acetaldehyde, acetic acid, acetoin, and 2,3-butanedione as

* Corresponding author. Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Department for Safety and Quality of Cereals, Schützenberg 12, 32756, Detmold, Germany.
E-mail address: Said.Elharkaoui@mri.bund.de (S. El Harkaoui).

https://doi.org/10.1016/j.crfs.2025.100986
Received 14 November 2024; Received in revised form 17 January 2025; Accepted 21 January 2025
Available online 22 January 2025

the most abundant aroma-active compounds, contributing to CO's characteristic flavor (Nounah et al., 2020). These components among others contribute to CO's diverse biological activities, which include *in vivo* and *in vitro* antioxidant effects, as well as antimicrobial, antidiabetic, lipid-lowering, anticancer, anti-inflammatory, and anti-ulcer properties (Barba et al., 2017; Ramadan and Tamer, 2021; Chbani et al., 2023; Al-Naqeb et al., 2021). Despite its edible nature, the primary application of CO currently lies in the cosmetic industry, where it is valued for its moisturizing, anti-aging, and skin-nourishing properties (Ramadan and Tamer, 2021; Chbani et al., 2023). The oil is marketed for its hydration potential, ability to improve skin elasticity, and to reduce skin redness and pigmentation. It also contains vitamin K1, which helps reduce dark circles and spider veins (Ramadan and Tamer, 2021).

In Morocco, CO production is a labor-intensive process, predominantly carried out by cooperatives in regions where the *Opuntia* cactus (mainly *Opuntia ficus-indica*) thrives. The extraction of oil from the small seeds requires significant quantities of raw material, contributing to the oil's high pricing (approx. 600 €/L in bulk) which underscores the economic motivation for adulteration (Chbani et al., 2020). The adulteration of CO can not only affect the health-promoting properties of the oil but also erode consumer's trust in CO and other products sold by cooperatives, highlighting the importance of ensuring its authenticity. Recent studies that have characterized CO from the perspectives of extraction, chemical composition, and potential applications, concluded that investigation of its potential adulteration remains an important, yet underexplored, research area (Nounah et al., 2024; Chbani et al., 2020, 2023).

Adulteration of high-value oils often involves diluting the expensive oil with inexpensive oils rather than outright substitution (Azadmard-Damirchi and Torbati, 2015). Commonly available and inexpensive refined oils are often chosen as potential adulterants (Dou et al., 2023). Adulterants that closely mimic organoleptic properties, such as smell and color of the expensive oil are preferred. Additionally, the market availability and the chemical composition of the adulterant are also considered. Based on this fact, refined sunflower oil (SO), with its similar visual appearance and fatty acid composition, is a likely candidate for adulterating CO. For detecting adulteration, one approach would be to focuses on analyzing a specific marker, such as trans fatty acids or stigmastadienes, which are often employed in detecting virgin-refined oil mixtures (Aued-Pimentel et al., 2013; Jabeur et al., 2014). While this approach have proven useful in certain cases, they are not always sufficient especially for lower levels of adulteration (Jabeur et al., 2014). Another key limitation of marker-based detection is that counterfeiters can specifically target these markers to avoid detection. An alternative approach involves analyzing multiple aspects of the oil's chemical composition, such as its fatty acids, triacylglycerols, and tocopherols. This multi-parameter approach strengthens detection by making it harder for counterfeiters to manipulate the oil's composition to avoid detection. Fatty acids, triacylglycerols, and tocopherols are frequently analyzed using chromatographic techniques such as gas chromatography (GC) or high-performance liquid chromatography (HPLC), which are both reliable and widely accepted methods in quality control laboratories. Chromatography techniques are appreciated in terms of robustness, sensitivity, precision, and accuracy, which are key parameters also useful for authenticity studies (Shi et al., 2022; Xing et al., 2019; Ilić et al., 2022; Esteki et al., 2018).

Supporting analytical approaches with chemometrics has proven to be a powerful tool for detecting adulteration in vegetable oils (Esteki et al., 2018). Chemometrics applies mathematics and statistical methods to process acquired data, particularly useful when results cannot be attained through the analysis of a single chemical marker but require the generation of multivariate data sets (i.e., analysis of the entire fatty acid composition) (Kamal and Karoui, 2015; Sudhakar et al., 2023). Distinguishing pure oils from adulterated ones presents a multivariate classification task, where classification models are employed to discriminate between pure and adulterated oils. Various classification models in combination with chromatographical techniques have been applied to detect adulteration in vegetable oils, each demonstrating its performance in different scenarios (Sudhakar et al., 2023; Zhang et al., 2017a). However, the size of the dataset is critical for building a robust classification model, and insufficient data may lead to overfitting, reducing the model's reliability (Qiu et al., 2018; Zhang et al., 2022; Shorten and Khoshgoftaar, 2019).

To generate an adequate dataset for training classification models, traditional approaches involve preparing and analysing a large number of pure oil samples and mixtures with varying levels of adulteration. This process, which includes physically combining and shaking the samples, is time-consuming and resource-intensive, presenting a significant challenge. Further, this challenge underscores the need for innovative approaches that optimize data generation and enhance model robustness without relying solely on extensive experimental work. To address this, data augmentation techniques have been introduced. Data augmentation methods involves the creation of new data by either slightly modifying original data or simulating new synthetic data based on artificial intelligence and statistical modeling (Gracia Moisés et al., 2023; Shorten and Khoshgoftaar, 2019; Georgouli et al., 2018; Wang et al., 2024). By incorporating simulated data, the dataset size can be expanded, enabling the exploration of a broader range of oil compositions.

To address the issue of CO adulteration, this study aimed at developing a machine learning model that integrates analytical measurements, data simulations, and classification techniques to detect CO adulterated with SO and determine the minimum detectable level of adulteration, without the need to analyze an extensive number of physical mixtures. First, oil samples, including pure CO and SO, were analyzed for their fatty acid, triacylglycerols, and tocochromanol using liquid and gas chromatography techniques (HPLC and GC). This analysis provided a database of oil compositions, serving as the foundation for subsequent simulations. Specifically, two simulation techniques, Monte Carlo simulation (MC), and Conditional Tabular Generative Adversarial Networks (CTGAN), were then employed to simulate oil compositions. By comparing the generated data from both models to the real samples, the model that produced the most accurate simulations should be identified. Using the selected model, oil mixtures were simulated by varying the proportions of CO and SO. The generated data from these simulations was then used to train two classification models: Random Forest (RF) and Neural Network (NN). At the end the trained models were evaluated for their accuracy in detecting adulteration on real-world samples. To ensure the model's long-term validity and applicability beyond the current dataset, the entire methodology was coded in Python and will be made publicly available. This approach enables continuous updates, allowing the model to remain a reliable resource for detecting CO adulteration as its market evolves.

2. Material and methods

To provide a comprehensive overview of the study, a workflow summary is presented in Fig. 1.

The following subsections describe each part of the methodology in detail, starting with the sample collection.

2.1. Material

The cold-pressed CO samples, in total 27 (labeled CO1 … CO27), were purchased from partner cooperatives across various regions of Morocco. The fruits of the cactus (*Opuntia ficus-indica*) were peeled, and the seeds were manually collected, washed, and sun-dried within the cooperatives. Oil extraction was performed using a screw press. The cooperatives were carefully selected from different locations in Morocco to ensure diverse provenance, and the oil quality was guaranteed by the cooperatives themselves. SO samples, in total 10 (labeled SO1 … SO10), were purchased from local Moroccan markets, representing the two

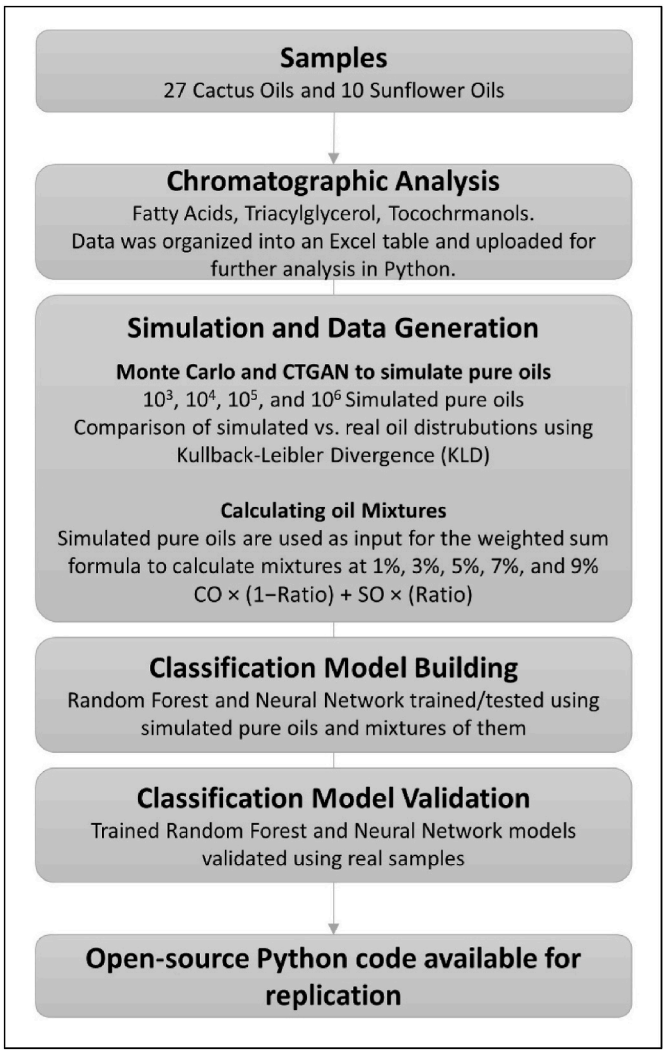


Fig. 1. Flow chart summarizing the methodology.

main brands available in the Moroccan market (Lesieur Cristal Group (HuilOR) and Huileries du Souss Belhassan (Zohor)). Adulterated oils were prepared in brown glass bottles at concentrations of 1%, 3%, 5%, 7%, and 9%, with each sample weighing 3 g. For practical reasons related to the nature of the oil matrix, mass-based measurements were chosen when preparing the mixtures. The oils were precisely measured using an analytical balance (AUW220, SHIMADZU, Japan) with an accuracy of 0.1 mg. The samples were then subjected to overnight shaking using a Shaker Mixer (Turbula T2F, Switzerland) for preparing a homogeneous blend. The prepared samples were stored at −18 °C until further analysis.

2.2. Determination of the fatty acid composition

Gas chromatography was employed to determine the fatty acid (FA) composition, following standard methods DGF C-VI 10a (00) and C-VI 11d (19) (DGF, 2021). Initially, a drop of oil was dissolved in 1 mL *n*-heptane (for liquid chromatography; LiChrosolv®, Supelco®, Merck KGaA, Darmstadt, Germany) and mixed with 50 μL sodium methylate (30% solution in methanol; Merck KGaA, Darmstadt, Germany). The mixture was agitated for 1 min at room temperature (22 °C), and then 100 μL of Millipore water was added. After centrifuging at 1550×*g* for 5 min, the lower aqueous phase was removed, and 50 μL 1M hydrochloric acid (min. 25.0%, p.a; Chemsolute®) was added along with some drops of the indicator methyl orange (ACS reagent, dye content 85%; Merck KGaA, Darmstadt, Germany). Following brief mixing, the lower aqueous phase was removed. Next, 20 mg sodium hydrogen sulphate (EMSURE® ACS, ISO, Reag. Ph Eur, Merck KGaA, Darmstadt, Germany) were added, and after centrifuging at 1550×*g* for 5 min, the top *n*-heptane phase was transferred to a vial and injected into an HP5890 gas chromatograph (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The chromatograph was equipped with a CP-Sil 88 capillary column (100 m × 0.25 mm × 0.25 μm; Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The temperature was gradually increased from 150 °C to 250 °C at a rate of 1.5 °C/min and was maintained at 250 °C for 5 min. The injector and detector were set to 260 °C, with the carrier gas (H_2) flow rate at 1.7 mL/min and a 1:50 split ratio. The detector was operated using 40 mL/min hydrogen, 400 mL/min air, and 40 mL/min nitrogen and had an injection volume of 1 μL. The fatty acid methyl esters (FAME) were identified by comparing their retention times with a standard mix (Supelco®37 Component FAME Mix; Merck KGaA, Darmstadt, Germany) and their composition was quantified as a percentage of the total FA.

2.3. Determination of the triacylglycerol composition

The triacylglycerol (TAG) composition was assessed using gas chromatography following standard method DGF C-VI 14 (08) (DGF, 2021). The analysis was conducted on an Agilent 6890 gas chromatograph equipped with a flame ionization detector (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). Fifty milligrams of oil were dissolved in 10 mL isooctane (EMSURE®, ACS, Reag. Ph Eur; Merck KGaA, Darmstadt, Germany), and 1 μL of this solution was injected into an RTX®65 TG column (30 m × 0.25 mm × 0.1 μm; Restek Corp., Bellefonte, PA, USA). The oven temperature was held at 300 °C for 1 min, then increased from 300 °C to 360 °C at a rate of 2 °C/min and held at 360 °C for 10 min. The injector and detector were set to 370 °C, and the carrier gas (H2) flow rate was 1 mL/min with a 1:40 split ratio. The detector operated using 40 mL/min hydrogen, 450 mL/min air, and 45 mL/min nitrogen. According to the DGF method, TAG were identified by comparing their retention times with those of sesame oil, which were previously established for TAG composition analysis. Commercial TAG standards, PaOlPa, PaOlOl, StOlOl, OlOlOl (99%, Merck KGaA, Darmstadt, Germany), were used for the identification process, as well ('Pa' represents palmitic acid, 'St' represents stearic acid, and 'Ol' represents oleic acid). The composition of each TAG was calculated as the percentage of its peak area relative to the total peak area of all detected peaks.

2.4. Determination of the tocochromanol composition

The tocochromanol content and composition were determined following standard method DGF F-II 4a (00) (DGF, 2021). Initially, 150 mg of oil were dissolved in 1 mL of *n*-heptane (for liquid chromatography LiChrosolv®, Supelco®, Merck KGaA, Darmstadt, Germany) and filtered using a 1.0 μm filter (Whatmann®, Maidstone, UK) followed by a 0.45 μm filter (Restek Corp., Bellefonte, Pennsylvania, USA). The filtered solution was then transferred into a vial and afterwards injected into the HPLC-FLD. The HPLC system configuration comprised a pump (L-7100 LaChrom Elite, Merck KGaA, Darmstadt, Germany), an autosampler (L-2200 LaChrom Elite, Merck KGaA, Darmstadt, Germany), a fluorescence detector (L-2485 LaChrom Elite, Merck KGaA, Darmstadt, Germany), and the interface box (Knauer Interface Box IF2, Berlin, Germany). To perform an isocratic separation, a diol phase column (25 cm × 4 mm × 5 μm, LiChroCART® 250-4, Merck KGaA, Darmstadt, Germany) was utilized. The mobile phase comprised *n*-heptane/tert. butyl methyl ether (for liquid chromatography LiChrosolv®, Supelco®, Merck KGaA, Darmstadt, Germany) at a flow rate of 1.3 mL/min. The injection volume for all samples was 20 μL, and the analysis time took 66 min. The fluorescence detector was set to an excitation wavelength of 295 nm and an emission wavelength of 330 nm. The tocochromanols

were identified using α-, β-, δ-, and γ-tocopherol reference standards (chromatographic purity 97.6–99.6%, Merck KGaA, Darmstadt, Germany) and quantified through external calibration with standard solutions (0.25–40 μg/mL). The tocochromanol content was quantified in mg/kg of oil, and these values were used to calculate the proportion of each tocochromanol relative to the total tocochromanol content (%).

2.5. Statistical methods

The analysis of FA, TAG, and tocochromanols of the samples were performed in two analytical replicates, with the results reported as means and standard deviations, to ensure that the measurements are reliable and not affected by external factors. For the subsequent sections on simulations and classification models, the mean values were used as robust estimates of the chemical values. Attributes below the limit of quantification or those not detected in one of the samples were assigned a value of zero.

2.5.1. Simulation models and data generation

In order to capture the higher variety of oils and to increase the database for more robust classification models, data were generated using two different methods: MC and CTGAN. The decision to use both CTGAN and MC simulations was made to compare a traditional statistical method with a deep learning approach in the generation of synthetic data adapted to the aims of the study. MC methods provide a well-proven approach for simulating data based on distributions. In contrast, CTGAN provides a more flexible and powerful tool with a little assumption for capturing complex relationships within the data. To our knowledge CTGAN has not been used before in the area of authenticity testing. By using both methods, their respective strengths should be evaluated and the most effective technique for our study should be identified to ensure that the generated samples closely approximate the statistical properties of the original data. The simulation process was conducted in two parts: first, the pure oils, CO and SO, were simulated using both MC and CTGAN. The simulated dataset that more closely matched the estimated real (observed) distribution was then used to calculate the synthetic mixtures, which were subsequently applied in the classification task.

Monte Carlo (MC): are a group of algorithms with different applications, such as modeling real-world scenarios and estimating possible outcomes. In the context of artificial intelligence, MC methods can be used to generate artificial data by randomly sampling data points from the probabilistic distribution of real data (Metropolis and Ulam, 1949). Assuming that each provided real CO and SO samples (observed samples) follow a normal distribution, the mean values for each of the features for CO and SO were computed, respectively, yielding a vector with 37 mean values for CO and SO. The mean values provide the central point of the distribution and ensure that the generated samples will be allocated around it. In order to capture the relations between the features, the covariance matrix, which provides the variance of the features (diagonal of the matrix), and the covariance between pairs of features (off-diagonal values), was computed. The covariance matrixes for SO and CO are reported in Fig. S1 and Fig. S2, respectively. With the covariance matrix, the variability and the relationships among the features for CO and SO in the real data is preserved in the generated samples. Thus, the samples were drawn from a multivariate normal distribution with 37 mean values of the features and a 37 × 37 covariance matrix capturing the relationship between the 37 features for SO and CO separately.

Conditional Tabular Generative Adversarial Networks model (CTGAN): is a generative deep-learning model that can learn the distribution of tabular data (without distribution assumption) and generate new data points out of that distribution (Goodfellow et al., 2014). These networks consist of two opposite networks, the generator and the discriminator. The generator is trained on generating data similar to the original data and the discriminator tries to differentiate between fake data points generated by the generator and real data. Both networks get trained simultaneously, while only the discriminator sees the real data. The discriminator then tries to distinguish between fake data and real data. After every training of an epoch, the weights of both networks get adjusted with backpropagation. The loss function of the generator determines how wrong the discriminator labeled the data and the loss function of the discriminator determines how right the data got classified. CTGAN, as part of a broader family of deep learning-based synthetic data generators, is particularly suited for handling single-table tabular data (Xu et al., 2019). In this study, the CTGAN model was trained using 27 samples of CO and 10 samples of SO, with training conducted over various numbers of epochs (100, 500, 1,000, 4,000, 10, 000, and 50,000).

For comparing MC and CTGAN, between 1,000 and 1,000,000 synthetic data points (per class) were generated using both methods. The accuracy of the simulations was evaluated using the Kullback-Leibler Divergence (KLD) (Kullback and Leibler, 1951). KLD offers a holistic metric by measuring the divergence between two probability distributions, providing a comprehensive assessment of the model's overall performance. It is defined in the formula below (Formula 1):

$$D_{KL}(N_0 \,||N_1) = \frac{1}{2}\left(trace\left(\Sigma_1^{-1}\Sigma_0\right) + (\mu_1 - \mu_0)^T \Sigma_1^{-1}(\mu_1 - \mu_0) - k + \log\left(\frac{\det\Sigma_1}{\det\Sigma_0}\right)\right) \quad (1)$$

Σ_1 is the non-singular covariance matrix of the observed data with dimension k

Σ_0 is the non-singular covariance matrix of the simulated data with dimension k

μ_1 are the mean values of the observed data

μ_0 are the mean values of the simulated data

The KLD is a positive value that is 0, when the two distributions are identical. The larger the KLD value, the more distant are the two distributions. Since the covariance matrices for CO and SO are singular (i.e., not invertible), small noise was added to these matrices to make them invertible for the KLD computation.

According to the KLD values, only the simulations from the more accurate model (MC or CTGAN) will be used to calculate mixtures of CO adulterated with SO at varying ratios 99:1%, 97:3%, 95:5%, 93:7%, and 91:9% using the weighted sum formula (Formula 2).

$$Calculated\ Mixture = CO \times (1 - Ratio) + SO \times (Ratio) \quad (2)$$

where CO is the value of a certain feature of cactus oil and SO the value of the same feature of sunflower oil and "Ratio" the level of sunflower oil adulteration between 1% and 9%.

2.5.2. Classification models

Random Forest and Neural Networks were selected as the two widely used classification models, with simulated pure oils and their mixtures serving as the training data. The two models were further tested on simulated data and on small real-world dataset.

Random Forest (RF): is a supervised machine learning algorithm built from an ensemble of Decision Trees. The trees classify the data set samples using a flowchart structure with yes/no questions that divide the tree nodes in a binary fashion. Usually, RF is preferred to Decision Trees because it is more robust and less prone to overfitting. In each tree and at each node, the features of the data are evaluated using a metric to determine which feature best splits the node into purer subsets i.e., nodes with the maximum possible number of samples belonging to a single class. The final predicted class for a sample is the class with the highest average probability among all the trees (Ho, 1995; Breiman, 2001). To enhance performance, RF can be fine-tuned through hyperparameter optimization. In this study, the following hyperparameters were tuned:

- Maximal depth the trees can reach (max_depth).

- How data will be selected to build the trees (bootrstrap).
- Number of features to consider for the best split (max_features).
- Minimum number of samples required to split a node (min_samples_split).
- Minimum number of samples to consider a node leaf or final node (min_samples_leaf).
- Number of trees of the RF (n_estimators).

For hyperparameter tuning the tool GridSearchCV from Scikit-learn was used. This approach takes a set or grid of hyperparameters and tests every possible combination of them to find out the best values. GridSearchCV also performs cross-validation, allowing the model being trained/tested on different subsets of the data. The output of the search is a set of hyperparameters for which the model achieved the highest performance in terms of accuracy.

Neural Network (NN): is a machine learning models inspired by the structure and function of the biological brain. The fundamental unit of a NN is the artificial neuron, which processes and transmits information. Each neuron receives input signals, combines them into a single output value, and applies a activation function to determine the neuron's response. The influence of each incoming signal is determined by a weight, which is adjusted during the training process. Neurons are organized into layers, typically with no connections within the same layer. Information flows through the NN in one direction, from the first layer (input layer), through multiple hidden layers, to the final output layer (Bishop, 2006; Schmidhuber, 2015). In this study, a fully connected NN was used. The network's input size matches the number of features (37), and the output layer corresponds to the number of classification classes (7). The following hyperparameters were tuned:

- Number of hidden layers: the hidden layers are the layers between input and output layer.
- Dimension for each hidden layer: number of neurons in each hidden layer.
- Batch size: number of data points fit to the NN in each iteration during the training.

The hyperparameter tuning was performed with the optuna framework. The framework tunes the hyperparameters based on the target metric, for which the accuracy was chosen. Optuna creates a trial with different values of hyperparameters. The values can be specified as a certain value space so the values are not chosen randomly. The defined values are reported as footnote in Table S2. Then the optuna randomly searches the value spaces of the hyperparameters to find the combination of hyperparameters with the best accuracy.

2.5.3. Coding and libraries

Machine learning models for oil simulation and classification were built using Python, with specific libraries addressing various methodological needs. MC simulations were performed with numpy (Harris et al., 2020), CTGAN with the CTGAN library (Xu et al., 2019), RF with Scikit-learn (Pedregosa et al., 2011), and the NN using the Pytorch Lightning library (Paszke et al., 2019). Hyperparameter tuning for NN was done using optuna (Akiba et al., 2019). Exploratory data analysis (EDA) was conducted in Jupyter notebooks using numpy, pandas, matplotlib, and PyTorch. GPU support was utilized for NN training, while CPUs were used for RF. The open-source script is freely available on the following GitHub (https://github.com/kida4bmel/oil-adulteration). The increase of dataset size impacts model performance and processing time, so that the availability of resources is a relevant factor when selecting the settings.

3. Results and discussion

3.1. Chemical composition of the samples

Regarding the chemical composition, the CO and SO samples were analyzed in terms of FA, TAG, and tocochromanol. The complete results are reported in the following data publications for CO (El Harkaoui et al., 2024a, 2024c, 2024b) and in the supplementary data B for SO.

For CO, the primary FAs identified were palmitic acid (Pl, 11.4–12.9%), oleic acid (Ol, 13.0–23.0%), and linoleic acid (Li, 54.6–65.4%), which together comprise approximately 90% of CO's total FA composition. These values are consistent with previous studies on Moroccan CO, such as those by Taoufik et al. (2015) which reported similar ranges for the major FA: palmitic acid (11.75–12.3%), oleic acid (18.2–22.6%), and linoleic acid (60.2–64.6%). The stable and characteristic FA profile of CO is further supported by multiple studies on Moroccan samples (El Harkaoui et al., 2023; Ettalibi et al., 2020, 2021; Gharby et al., 2015, 2021; Nounah et al., 2021, 2024). This consistency suggests a stable and characteristic FA profile for CO, which remains largely unaffected by processing.

For SO, the FA composition revealed palmitic acid (5.6–6.9%), oleic acid (28.5–32.7%), and linoleic acid (54.6–59.1%), in the same predominance order as CO. Similar to CO, these three FAs account for approximately 90% of SO's total FA composition. The results for SO are within the ranges specified for SO by the Codex Alimentarius Standard for named Vegetable Oils (Codex, 1999). The major FA in both oils, linoleic, oleic, palmitic, and stearic acids constitute the main monomers of TAGs such as PaLiLi, LiLiLi, and LiLiOl. FA which accounts for 95–98% of the oil composition (Yara-Varón et al., 2017), serve as a chemical fingerprint that remains relatively stable even after processing and storage, making them reliable indicators of oil authenticity.

Both CO and SO contain α-tocopherol and γ-tocopherol as their predominant tocopherol isomers, with differences in their distribution. In CO, γ-tocopherol typically makes up the majority, ranging from 92 to 97%, while α-tocopherol is found in smaller amounts (up to 4%). In SO, α-tocopherol tends to be more prominent, around 96–97%, with γ-tocopherol present in lower proportions (up to 4%). Detailed tocochromanol composition for both oils can be found in the following data publication (El Harkaoui et al., 2024b) and in the supplementary data B for SO. The tocopherol profile for the analyzed CO seed oils is consistent with previously published results on Moroccan samples, showing similar patterns in tocopherol isomer dominance, although with slight variations in the reported ranges (Nounah et al., 2021, 2024; El Harkaoui et al., 2023; Gharby et al., 2015, 2020, 2021; Taoufik et al., 2015). For SO, the tocopherol content aligns with the ranges specified by the Codex Alimentarius Standard for named Vegetable Oils (Codex, 1999). While tocopherols offer additional insights for authenticity detection, their profiles can be influenced by factors such as light exposure, thermal degradation, mold growth, and refining processes. This sensitivity underscores the need for caution when relying solely on tocopherols for adulteration detection (Jee, 2002). The examination of multiple variables together with machine learning models (which leverages multiple features for decision-making), should enhance robustness and make it more difficult for counterfeiters to manipulate the composition of the oil to avoid detection.

3.2. Simulation of the data

Two simulation models, MC and CTGAN, were employed to simulate data, with the objective of comparing their effectiveness in producing samples that closely mirror the distribution of the original data. While box plots are commonly used to evaluate simulation results by visually comparing the distribution of raw and simulated data (Zhang et al., 2022; Cui et al., 2024), they are less practical in the context of multivariate simulations due to the large number of variables involved, which would require a high number of box plots. To address this, the KLD for

multivariate normal distributions was utilized. The results are shown in Table 1.

The MC method demonstrated consistent performance across all tested data set sizes, maintaining rather low KLD values for each sample size, especially from 10,000 samples onwards. For CO, KLD values ranged between 2.3 and 3.1, while for SO, the values remained between 1.2 and 1.4. An improvement is noticeable from 1,000 to 10,000 samples reducing KLD value from 3.1 to 2.5 and 1.4 to 1.2, respectively. In contrast, the KLD values of CTGAN were several times higher and were heavily influenced by the number of training epochs. At lower epochs (100–1,000), CTGAN's KLD values were significantly higher than those of MC, indicating a less accurate simulation of the original data. For example, with 1,000 simulations, KLD values for SO were as high as 2,531 at 100 epochs. However, as the number of epochs increased, CTGAN's ability to mimic the original data improved slightly. At 50,000 epochs, CTGAN produced the lowest KLD values across all scenarios, with 89.5 for CO and 68.2 for SO at 10,000 simulations. But even with 50,000 epochs, all KLD are several times higher than for MC. The size of the data set had minimal impact on the performance of MC, further reinforcing its stability and reliability across various simulation scales. For CTGAN, while larger data sets did contribute to a slight reduction in KLD values, the effect was less pronounced compared to the impact of increasing the number of epochs. This suggests that for CTGAN, the duration of training is more critical than the volume of data to be simulated. However, longer training durations are computationally expensive, and if 50,000 epochs still result in high KLD values, it may not be practical to extend the training further.

The MC method performed well in this study scenario probably because it is based on a straightforward statistical approach. By sampling from the distributions (mean and covariance) of the original data, it simulates data that naturally aligns with the statistical properties of the real-world samples. This leads to a lower KLD. CTGAN while powerful, require large and diverse datasets to effectively learn the complex relationships in the data. With only 27 CO and 10 SO samples, the CTGAN model likely lacked sufficient data to fully capture these relationships, leading to higher KLD. This could be a possible explanation why the MC performed better than the CTGAN with the sample set that we have in the study.

MC simulation has been successfully applied in other studies that used comparable original sample sizes for the simulation of vegetable oils, further supporting our findings (Zhang et al., 2014, 2017a, 2017b). Although these studies did not compare MC with other simulation models nor detailed the specific implementation, the consistent results suggest the robustness of the MC approach in similar scenarios. In contrast, to the best of our knowledge, GAN has not yet been applied to the chemical composition of oils using conventional tabular datasets for simulating vegetable oil compositions, making direct comparisons with similar studies difficult. However, other GAN variants have been successfully employed with hyperspectral images to generate sufficient synthetic data for training models aimed at different objectives, such as predicting oil content in maize (using deep convolutional GAN) (Zhang et al., 2022) or polyunsaturated FA content in meat (using autoencoder-assisted GAN) (Cui et al., 2024). After multiple iterations, these GAN models were able to generate synthetic data that closely resembled experimental data (Zhang et al., 2022; Cui et al., 2024). Thus, the performance of GAN models depends significantly on the specific scenario, data type, and study objectives. Indeed its application as data simulation in food authenticity is still in its early stages (Deng et al., 2024).

The artificial pure oils simulated using MC were then used to calculate different levels of adulteration artificially using the weighted sum formula (Formula 2) as explained in the method section. Table 2 shows the number of the data simulated for each class which will be used for the training of the classification models.

3.3. Classification models

For the classification task, RF and NN models were employed, utilizing data simulated by MC and the weighted sum formula (Table 2). The data was split into 80% for training and 20% for testing. Hyperparameter tuning was performed for both classification models. Tuning the hyperparameters is a critical step before proceeding to the classification and it helps to reduce the overfitting of the classification model, which may happen when default hyperparameters are used (A Ilemobayo et al., 2024). The tuning of the hyperparameters was done for both NN and RF as described in section 2.5.2. The optimized hyperparameters for RF and NN are reported in Table S1 and Table S2, respectively.

After training the models on simulated data, their performance was first assessed using a separate set of simulated test data (section 3.3.1) and was validated later using small real-world dataset (section 3.3.2). Model performance was primarily evaluated using accuracy, which represents the proportion of correctly classified oil samples out of all predictions made by the model. In addition to accuracy as a global performance metric, also confusion matrices and classification reports were examined, which provide additional insights into model performance at the class level through metrics such as precision, recall, and F-score.

3.3.1. Evaluation of the classification model performance using simulated data

Table 3 presents the classification accuracies of the RF and NN models for various quantities of simulated samples. The table highlights

Table 2
Number of simulated samples for each class.

Classes	Number of simulated samples
CO (100% pure CO)	1,000/10,000/100,000/1,000,000
SO (100% pure SO)	1,000/10,000/100,000/1,000,000
99:1 (Mixture of 99% CO and 1% SO)	1,000/10,000/100,000/1,000,000
97:3 (Mixture of 97% CO and 3% SO)	1,000/10,000/100,000/1,000,000
95:5 (Mixture of 95% CO and 5% SO)	1,000/10,000/100,000/1,000,000
93:7 (Mixture of 93% CO and 7% SO)	1,000/10,000/100,000/1,000,000
91:9 (Mixture of 91% CO and 9% SO)	1,000/10,000/100,000/1,000,000

Table 1
Kullback-Leibler Divergence values for data sets of different sizes simulated with Monte Carlo (MC) and Conditional Tabular Generative Adversarial Network (CTGAN) [values in 10^3].

Simulation per oil		MC	CTGAN by Epochs					
			100	500	1,000	4,000	10,000	50,000
1,000	CO	3.1	974.3	972.7	898.4	272.5	187.2	105.1
	SO	1.4	2,531	2,266.2	1,717.4	595.6	209.0	76.8
10,000	CO	2.5	991.5	949.9	898.6	277.2	192.7	89.5
	SO	1.2	2,482.9	2,303.5	1,676.2	603.6	207.1	68.2
100,000	CO	2.3	997.7	952.3	895.1	277.8	194.3	93.1
	SO	1.3	2,470.9	2,282.6	1,681.7	460.7	208.5	68.3
1,000,000	CO	2.3	1,000.7	950.2	899.8	272.3	195.3	92.9
	SO	1.3	2,471	2,279.1	1,683.9	512.4	207.8	68.2

Table 3
Performances of Random Forest (RF) and Neural Network (NN) on simulated data (classification accuracies %).

Number of simulated samples per class	RF	NN
1,000	83.7%	92.0%
10,000	90.5%	92.0%
100,000	92.5%	93.0%
1,000,000	94.0%	93.0%

how the performance of each model scales as the amount of available data for training/testing grows, which is crucial for understanding their potential.

The RF model showed an improvement in classification accuracy as the number of simulated samples used to train/test the model increases. Starting with 1,000 samples per class, RF achieves an accuracy of 83.7%. This improves to 90.5% with 10,000 samples per class, and continue to rise, reaching 94% with 1,000,000 samples per class. The trend suggests that RF benefits from larger training datasets, which likely allows the model to capture more complex patterns and variability in the data. Another relevant enhancing factor is the use of hyperparameter tuning, which allows determining which hyperparameters are optimal for each dataset size. However, the accuracy increases from 100,000 samples per class (92.5%) to 1,000,000 samples per class (94%) were relatively modest, suggesting that the relevant patterns were possible captured by the algorithm. It is a quite natural behavior that algorithms don't show a linear improvement with increasing number of observations, but may increase faster at the beginning and slowly later (Domingos, 2012).

In contrast, the NN model showed a pattern which were less related to the sample size. At smaller dataset sizes, NN starts with a high accuracy of 92% with 1,000 samples, outperforming RF at this initial stage. However, as the dataset size increases, NN's performance stabilizes, at accuracy values of 92% for 10,000, and 93% for 100,000 and 1,000,000 samples per class. While further tuning of NN parameters for 100,000 samples onwards could potentially improve its accuracy, it becomes highly complex and computationally demanding, especially with larger datasets. Therefore, the observed performance difference between RF and NN can be mainly attributed to the intrinsic characteristics of the models but also to some extent of the tuning applied.

For a more detailed visualization of the results, Fig. 2 shows the confusion matrices for the two classification models, RF and NN. Each column contains the confusion matrices for each size of the simulated data sets (1,000, 10,000, 100,000, and 1,000,000 samples/per class). The confusion matrix shows in detail the number of correct classified samples per class, which can be found on the diagonal, and the number of misclassified samples per class, located at the off-diagonals.

Both models consistently differentiate CO from SO across all dataset sizes, indicating that the fundamental distinction between the oils is well captured. However, the most frequent misclassifications occur between the 99:1 adulteration class and pure CO. This is not surprising, as the presence of only 1% SO in a mixture is difficult to distinguish from pure CO and may also fall within measurement error. Other misclassifications tend to occur between classes with minimal differences adulteration, such as 97:3 and 95:5 or 93:7 and 91:9. In other words, differences of ±2% SO will not be always recognized by both classification models. This is also supported by the values of precision, recall, and f1-score for which lower values for all the metrics where found for the challenging classifications (Fig. S3). Nevertheless, adulteration levels of 3% and higher are reliably identified by both models, and there are no misclassifications within a margin of ±4%. When comparing the models as the simulated dataset size increases, NN demonstrates reasonable classification performance at the 1,000 simulation, with more correct classifications in the confusion matrix (more samples in the diagonal) compared to RF at this early stage. However, as also indicated in Table 3, RF begins to gain in performance as the dataset size increases, showing improved accuracy and better class-level metrics, particularly for larger datasets. Even with 1% adulteration, the misclassifications remain reasonable for RF, as it classified around 14% of pure CO as 1% adulteration and around 13% of the 1% adulteration as pure CO, which is acceptable level of error.

3.3.2. *Evaluation of the classification model performance using real-world samples*

Having developed and tested two classification model's RF and NN on simulated data, now the focus has shifted to assessing their performance using a smaller, real-world dataset of adulterated samples. This serves as an important validation step to ensure that the models, which

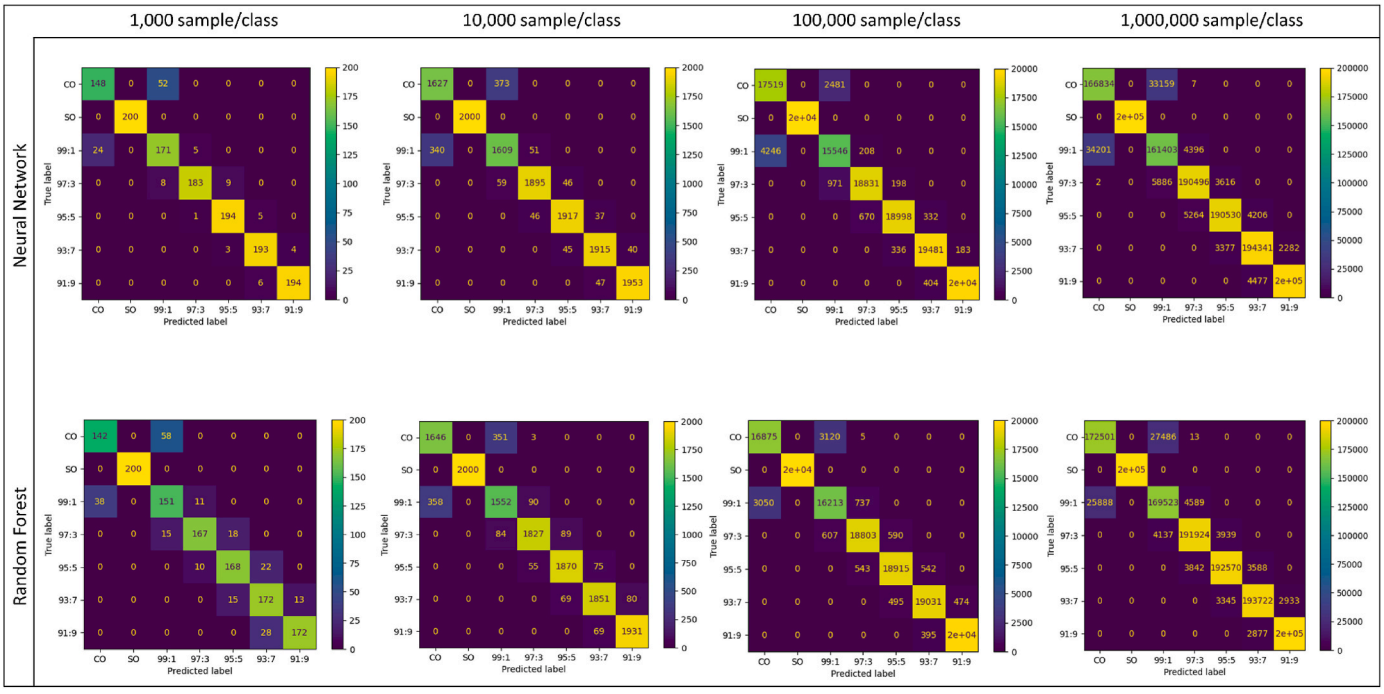


Fig. 2. Confusion matrices showing the performances of Random Forest and Neural Network using test sets of different numbers of simulated samples.

were trained and tested on multivariate normal distributions (MC simulated data), can be effectively applied in real-world scenarios. A total of 33 adulterated CO samples were prepared, as detailed in the methods section. These represent approximately 2.4% (33 out of 27 CO × 10 SO × 5 adulteration levels) of all possible combinations. Table S3 provides details on the number of adulterated CO samples prepared for each level of adulteration.

When evaluating the models on real-world samples (Table 4), the observed accuracies were consistently lower than those achieved with the simulated data. However, in every case, they remained close or above 70%, which is reasonable considering the number of classes (seven), the possible variation of adulterations which may be not always exact at the assumed levels, the potential unobserved heterogeneities in real-world conditions and the small sample size of specific oils.

Considering the performances of RF, the differences between simulated and real-world data become smaller when the size of the simulated datasets becomes larger. The smallest distance of 4% is reached for the RF trained on one million samples of simulated data. This suggests that with larger datasets not only the performances of RF become better, but also the capacity of the model to generalize is enhanced, so that the risk of overfitting is reduced. The combination of both hyperparameter tuning and also using larger datasets for training yields enhances robustness for RF models. Considering the performances of NN, the accuracies were in most cases lower than the RF. The only exception was found for a smaller dataset (1,000 samples), where NN achieved a slight accuracy advantage over RF (with ca. 0.5% points higher). However, when the models were trained on larger datasets (10,000 and above), RF clearly outperformed NN. Therefore, the NN seems to be less well generalizable to real-world conditions as the RF. This behavior is often observed, when a model focuses too much on certain given data points than learning the overall pattern.

Further insights into model performance are provided by the confusion matrices in Fig. 3 and the classification metrics in Fig. S4. Both models correctly classified pure oils, but minor misclassifications occur in the other classes, primarily between classes that were close in adulteration level. These misclassifications followed the same pattern as observed with the simulated data, with both models struggling more with classes that differ by only small amounts of adulteration. Notably, models trained with larger datasets tend to correctly classify more samples, particularly in the case of RF. As the number of collected pure oil samples was limited, the same CO and SO real-world samples used for estimating the underlying distributions for MC simulations, were also used as the basis for generating real-world adulterated samples. It is not expected that this has a large impact on the validation since the real-world samples were considered as random realizations of the underlying distribution. However, a positive effect on the model performance cannot be entirely excluded. Therefore, further independent evaluations are recommended to confirm the promising results reported here, providing a more robust validation.

When comparing the models in terms of practical application, both performed well; however, RF exhibited greater robustness, especially with larger datasets, and consistently achieved better classification accuracy, on both simulated and real-world samples. Additionally, KLD values (Table 1) indicate that datasets with 10,000 or more samples provided a closer fit to observed data than those with 1,000 samples, implying that the advantage of NN at 1,000 samples was not as relevant as the performance in the settings with more samples.

From the perspective of CO authenticity testing and possibly oil authenticity in general model interpretability is an important consideration. RF offers an advantage here, as its decision trees provide less complex, traceable decision-making processes, which can be crucial for regulatory purposes where transparency is required. In contrast, NN operates more like a "black box", which could make it less practical in situations requiring interpretability (Roβbach, 2018). Furthermore, RF models are computationally less demanding, requiring less time and fewer hardware resources for training and optimization compared to NN, which can be demanding when working with large datasets which seem to be needed to capture the observed data sufficiently. Although NN shows stable performance and has the ability to capture complex patterns, the comparable and often better accuracy of RF in our case, combined with its interpretability and computational efficiency, suggests that RF is the more suitable model for this particular scenario.

Table 4
Performance of trained Random Forest and Neural Network tested on real-world samples after being trained/tested on different numbers of simulated samples (classification accuracies %).

Number of simulations used to train/test the model	RF	NN
1,000	68.5%	69.0%
10,000	74.2%	73.0%
100,000	88.5%	81.0%
1,000,000	90.0%	74.0%

4. Conclusion

This study explored the integration of data simulation and machine learning techniques to address the challenge of detecting adulterated cactus seed oil (CO) in order to evaluate the limit of adulteration that can be detected without measuring a huge number of different mixtures. From a methodological perspective, Monte Carlo simulation proved to simulate more realistic oil composition data, in our case, compared to Conditional Tabular Generative Adversarial Network, enhancing model performance and reducing the need for large-scale sample collection. While Neural Networks showed potential in capturing complex patterns, Random Forests interpretability and practicality make it more suitable for this setting. By employing a Monte Carlo simulation with Random Forest classification, we achieved high accuracy (90%) in detecting adulteration of CO with refined sunflower oil, with detectable adulteration levels as low as 1%. As a proof-of-concept, the approach has exemplarily demonstrated its potential specifically for an adulteration of Moroccan CO with SO. It could be feasibly implemented as a routine protocol in quality control laboratories in order to detect adulterated CO and estimate the level of adulteration. The open-access Python-based methodology ensures that the models can be continuously updated with new data, making it adaptable for future studies. This methodology has significant implications for protecting the authenticity of CO, reinforcing consumer trust, and supporting Moroccan cooperatives. However, further evaluations in real-world settings will be necessary to refine its applicability. Future research should explore the scalability and robustness of the method for broader CO adulteration scenarios. This includes addressing mixtures involving various refined oils, incorporating larger and more diverse datasets, and enhancing the model's ability to handle missing values. From a methodological perspective, the proposed approach can be extended to other types of oils, as long as suitable data are utilized and the model is properly trained and tuned.

CRediT authorship contribution statement

Said El Harkaoui: Conceptualization, Methodology, Investigation, Writing – original draft, Visualization, Project administration. **Cristina Ortiz Cruz:** Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Aaron Roggenland:** Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Micha Schneider:** Methodology, Data curation, Formal analysis, Writing – review & editing. **Sascha Rohn:** Validation, Data curation, Writing – review & editing, Supervision. **Stephan Drusch:** Validation, Data curation, Writing – review & editing, Supervision. **Bertrand Matthäus:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

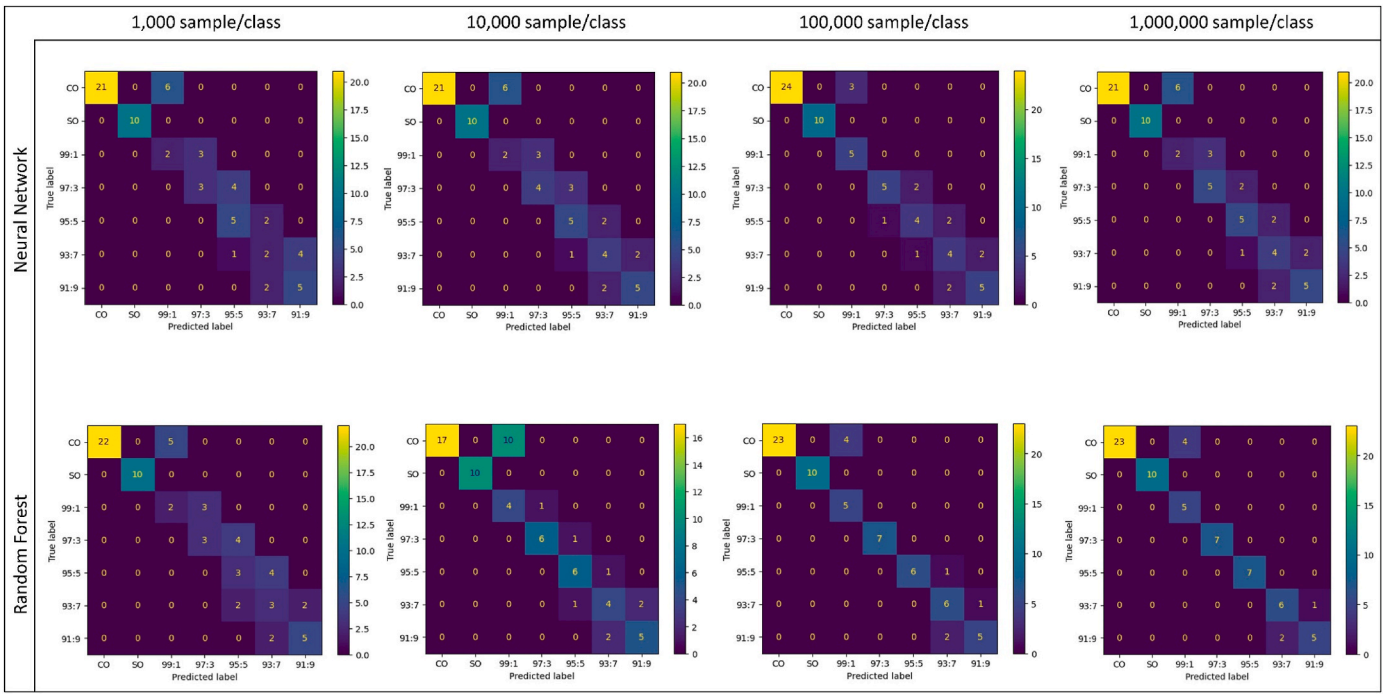


Fig. 3. Confusion matrices of trained Random Forest and Neural Network models, tested on real-world samples after being trained/tested on different numbers of simulated samples.

Funding

This work was supported by funds of the Federal Ministry of Food and Agriculture BMEL based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) [support program: FKZ 2819DOKA03 and FKZ 28KIDA007].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Prof. Zoubida Charrouf, Prof. Hanae El Monfalouti, Prof. Badr Eddine Kartah from Mohammed V University in Rabat, Morocco, as well as Prof. Said Gharby from Ibn Zohr University in Agadir, Morocco, for their help and assistance in sample collection and ensuring smooth communication with the cooperatives.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2025.100986.

Data availability

The data publications containing the data used in this article are referenced in Section 3.1 and are available in the OpenAgrar repository, along with their corresponding DOIs. The Python script that implements the entire methodology is openly accessible at the following link: https://github.com/kida4bmel/oil-adulteration.

References

A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., E Odezuligbo, I., E Edu, O., 2024. Hyperparameter tuning in machine learning: a comprehensive review. Journal of Engineering Research and Reports 26 (6), 388–395. https://doi.org/10.9734/jerr/2024/v26i61188.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: a next-generation hyperparameter optimization framework. In: In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631.

Al-Naqeb, G., Fiori, L., Ciolli, M., Aprea, E., 2021. Prickly pear seed oil extraction, chemical characterization and potential health benefits. Molecules 26 (16), 5018. https://doi.org/10.3390/molecules26165018.

Aued-Pimentel, S., Da Silva, S.A., Takemoto, E., Cano, C.B., 2013. Stigmastadiene and specific extitntion (270 nm) to evaluate the presence of refined oils in virgin olive oil commercialized in Brazil. Food Sci. Technol. 33 (3), 479–484. https://doi.org/10.1590/S0101-20612013005000067.

Azadmard-Damirchi, S., Torbati, M., 2015. Adulterations in some edible oils and fats and their detection methods. Journal of Food Quality & Hazards Control 2, 38–44.

Barba, F.J., Putnik, P., Bursać Kovačević, D., Poojary, M.M., Roohinejad, S., Lorenzo, J. M., Koubaa, M., 2017. Impact of conventional and non-conventional processing on prickly pear (Opuntia spp.) and their derived products: from preservation of beverages to valorization of by-products. Trends Food Sci. Technol. 67, 260–270. https://doi.org/10.1016/j.tifs.2017.07.012.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Chbani, M., El Harkaoui, S., Willenberg, I., Matthäus, B., 2023. Review: analytical extraction methods, physicochemical properties and chemical composition of cactus (Opuntia ficus-indica) seed oil and its biological activity. Food Rev. Int. 39 (7), 4496–4512. https://doi.org/10.1080/87559129.2022.2027437.

Chbani, M., Matthäus, B., Charrouf, Z., El Monfalouti, H., Kartah, B., Gharby, S., Willenberg, I., 2020. Characterization of phenolic compounds extracted from cold pressed cactus (Opuntia ficus-indica L.) seed oil and the effect of roasting on their composition. Foods 9 (8), 1098. https://doi.org/10.3390/foods9081098.

Codex, 1999. Codex Alimentarius: standard for named vegetable oils Codex; CXS 210-1999. Adopted in 1999, Amended in 2023. https://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/. (Accessed 8 November 2024).

Cui, J., Li, K., Lv, Y., Liu, S., Cai, Z., Luo, R., Zhang, Z., Wang, S., 2024. Development of a new hyperspectral imaging technology with autoencoder-assisted generative adversarial network for predicting the content of polyunsaturated fatty acids in red meat. Comput. Electron. Agric. 220, 108842. https://doi.org/10.1016/j.compag.2024.108842.

Deng, Z., Wang, T., Zheng, Y., Zhang, W., Yun, Y.-H., 2024. Deep learning in food authenticity: recent advances and future trends. Trends Food Sci. Technol. 144, 104344. https://doi.org/10.1016/j.tifs.2024.104344.

DGF, 2021. Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten. Tensiden und verwandten Stoffen. Wissenschaftliche Verlagsgesellschaft, Stuttgart, Germany.

Dou, X., Zhang, L., Chen, Z., Wang, X., Ma, F., Yu, L., Mao, J., Li, P., 2023. Establishment and evaluation of multiple adulteration detection of camellia oil by mixture design. Food Chem. 406, 135050. https://doi.org/10.1016/j.foodchem.2022.135050.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr Eddine, Gharby, S., Charrouf, Z., Matthäus, B., 2024a. Dataset: fatty acid composition of 27 cold-pressed cactus seed oils from the Moroccan market. https://doi.org/10.25826/Data20240930-105507-0.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr Eddine, Gharby, S., Charrouf, Z., Matthäus, B., 2024b. Dataset: tocochromanol composition of 27 cold-pressed cactus seed oils from the Moroccan market. https://doi.org/10.25826/Data20240930-113217-0.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr Eddine, Gharby, S., Charrouf, Z., Matthäus, B., 2024c. Dataset: triacylglycerol composition of 27 cold-pressed cactus seed oils from the Moroccan market. https://doi.org/10.25826/Data20240930-114743-0.

El Harkaoui, S., Gharby, S., Kartah, B., El Monfalouti, H., El-sayed, M.E., Abdin, M., Salama, M.A., Charrouf, Z., Matthäus, B., 2023. Lipid profile, volatile compounds and oxidative stability during the storage of Moroccan Opuntia ficus-indica seed oil. Grasas Aceites 74 (1), e486. https://doi.org/10.3989/gya.1129212.

Esteki, M., Simal-Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E., Vander Heyden, Y., 2018. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control 93, 165–182. https://doi.org/10.1016/j.foodcont.2018.06.015.

Ettalibi, F., El Antari, A., Gadhi, C., Harrak, H., 2020. Oxidative stability at different storage conditions and adulteration detection of prickly pear seeds oil. J. Food Qual. 2020 (1), 1–12. https://doi.org/10.1155/2020/8837090.

Ettalibi, F., El Antari, A., Hamouda, A., Gadhi, C., Harrak, H., 2021. Comparative assessment of physical and chemical characteristics of prickly pear seed oil from Opuntia ficus-indica and Opuntia megacantha varieties. J. Food Qual. 2021 (1), 1–8. https://doi.org/10.1155/2021/3098608.

Georgouli, K., Osorio, M.T., Del Martinez Rincon, J., Koidis, A., 2018. Data augmentation in food science: synthesising spectroscopic data of vegetable oils for performance enhancement. J. Chemometr. 32 (6), e3004. https://doi.org/10.1002/cem.3004.

Gharby, S., Guillaume, D., Nounah, I., Harhar, H., Hajib, A., Matthäus, B., Charrouf, Z., 2021. Shelf-life of Moroccan prickly pear (Opuntia ficus-indica) and argan (Argania spinosa) oils: a comparative study. Grasas Aceites 72 (1), e397. https://doi.org/10.3989/gya.1147192.

Gharby, S., Harhar, H., Charrouf, Z., Bouzobaa, Z., Boujghagh, M., Zine, S., 2015. Physicochemical composition and oxidative stability of opuntia FICUS-indica seed oil from Morocco. Acta Hortic. 1067, 83–88. https://doi.org/10.17660/actahortic.2015.1067.11.

Gharby, S., Ravi, H.K., Guillaume, D., Abert Vian, M., Chemat, F., Charrouf, Z., 2020. 2-methyloxolane as alternative solvent for lipid extraction and its effect on the cactus (Opuntia ficus-indica L.) seed oil fractions. OCL 27, 27. https://doi.org/10.1051/ocl/2020021.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Gracia Moisés, A., Vitoria Pascual, I., Imas González, J.J., Ruiz Zamarreño, C., 2023. Data augmentation techniques for machine learning applied to optical spectroscopy datasets in agrifood applications: a comprehensive review. Sensors 23 (20), 8562. https://doi.org/10.3390/s23208562.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585 (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Ho, T.K., 1995. Random decision forests. Proceedings of 3rd international conference on document analysis and recognition 1, 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

Ilić, M., Pastor, K., Romanić, R., Vujić, Đ., Ačanski, M., 2022. A new challenge in food authenticity: application of a novel mathematical model for rapid quantification of vegetable oil blends by gas chromatography – mass spectrometry (GC-MS). Anal. Lett. 55 (17), 2752–2763. https://doi.org/10.1080/00032719.2022.2069795.

Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., Bouaziz, M., 2014. Detection of Chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. J. Agric. Food Chem. 62 (21), 4893–4904. https://doi.org/10.1021/jf500571n.

Jee, M., 2002. Oils and Fats Authentication. Blackwell Publishing, Oxford.

Kamal, M., Karoui, R., 2015. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: a review. Trends Food Sci. Technol. 46 (1), 27–48. https://doi.org/10.1016/j.tifs.2015.07.007.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86. https://doi.org/10.1214/aoms/1177729694.

Metropolis, N., Ulam, S., 1949. The Monte Carlo method. J. Am. Stat. Assoc. 44 (247), 335. https://doi.org/10.2307/2280232.

Nounah, I., Chbani, M., Matthäus, B., Charrouf, Z., Hajib, A., Willenberg, I., 2020. Profile of volatile aroma-active compounds of cactus seed oil (Opuntia ficus-indica) from different locations in Morocco and their fate during seed roasting. Foods 9 (9), 1280. https://doi.org/10.3390/foods9091280.

Nounah, I., El Harkaoui, S., Hajib, A., Gharby, S., Harhar, H., Bouyahya, A., Caprioli, G., Maggi, F., Matthäus, B., Charrouf, Z., 2024. Effect of seed's geographical origin on cactus oil physico-chemical characteristics, oxidative stability, and antioxidant activity. Food Chem. X 22, 101445. https://doi.org/10.1016/j.fochx.2024.101445.

Nounah, I., Gharby, S., Hajib, A., Harhar, H., Matthäus, B., Charrouf, Z., 2021. Effect of seeds roasting time on physicochemical properties, oxidative stability, and antioxidant activity of cactus (Opuntia ficus-indica L.) seed oil. J. Food Process. Preserv. 45 (9), e15747. https://doi.org/10.1111/jfpp.15747.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., Zhang, C., 2018. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. Appl. Sci. 8 (2), 212. https://doi.org/10.3390/app8020212.

Ramadan, M.F., Tamer, E.M.A., 2021. Sascha Rohn. *Opuntia Spp.: Chemistry, Bioactivity and Industrial Applications*. Springer International Publishing, Cham.

Roßbach, P., 2018. Neural Networks vs. Random Forests–Does it Always Have to Be Deep Learning. Frankfurt School of Finance and Management.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural networks the official journal of the International Neural Network Society 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Shi, T., Wu, G., Jin, Q., Wang, X., 2022. Camellia oil adulteration detection using fatty acid ratios and tocopherol compositions with chemometrics. Food Control 133, 108565. https://doi.org/10.1016/j.foodcont.2021.108565.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48. https://doi.org/10.1186/s40537-019-0197-0.

Sudhakar, A., Chakraborty, S.K., Mahanti, N.K., Varghese, C., 2023. Advanced techniques in edible oil authentication: a systematic review and critical analysis. Crit. Rev. Food Sci. Nutr. 63 (7), 873–901. https://doi.org/10.1080/10408398.2021.1956424.

Taoufik, F., Zine, S., El Hadek, M., Idrissi Hassani, L.M., Gharby, S., Harhar, H., Matthäus, B., 2015. Oil content and main constituents of cactus seed oils Opuntia Ficus Indica of different origin in Morocco. Mediterr. J. Nutr. Metabol. 8 (2), 85–92. https://doi.org/10.3233/MNM-150036.

Wang, Y., Gu, H.-W., Yin, X.-L., Geng, T., Long, W., Fu, H., She, Y., 2024. Deep leaning in food safety and authenticity detection: an integrative review and future prospects. Trends Food Sci. Technol. 146, 104396. https://doi.org/10.1016/j.tifs.2024.104396.

Xing, C., Yuan, X., Wu, X., Shao, X., Yuan, J., Yan, W., 2019. Chemometric classification and quantification of sesame oil adulterated with other vegetable oils based on fatty acids composition by gas chromatography. Lebensm. Wiss. Technol. 108, 437–445. https://doi.org/10.1016/j.lwt.2019.03.085.

Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K., 2019. Modeling tabular data using conditional GAN. Adv. Neural Inf. Process. Syst. 32.

Yara-Varón, E., Li, Y., Balcells, M., Canela-Garayoa, R., Fabiano-Tixier, A.-S., Chemat, F., 2017. Vegetable oils as alternative solvents for green oleo-extraction, purification and formulation of food and natural products. Molecules 22 (9), 1474. https://doi.org/10.3390/molecules22091474.

Yuan, Z., Zhang, L., Du, Wang, Jiang, J., Harrington, P.d.B., Mao, J., Zhang, Q., Li, P., 2020. Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis. Lebensm. Wiss. Technol. 125, 109247. https://doi.org/10.1016/j.lwt.2020.109247.

Zhang, L., Huang, X., Li, P., Na, W., Jiang, J., Mao, J., Ding, X., Zhang, Q., 2017a. Multivariate adulteration detection for sesame oil. Chemometr. Intell. Lab. Syst. 161, 147–150. https://doi.org/10.1016/j.chemolab.2016.11.009.

Zhang, L., Li, P., Sun, X., Wang, X., Xu, B., Wang, X., Ma, F., Zhang, Q., Ding, X., 2014. Classification and adulteration detection of vegetable oils based on fatty acid profiles. J. Agric. Food Chem. 62 (34), 8745–8751. https://doi.org/10.1021/jf501097c.

Zhang, L., Wang, Y., Wei, Y., An, D., 2022. Near-infrared hyperspectral imaging technology combined with deep convolutional generative adversarial network to predict oil content of single maize kernel. Food Chem. 370, 131047. https://doi.org/10.1016/j.foodchem.2021.131047.

Zhang, L., Yuan, Z., Li, P., Wang, X., Mao, J., Zhang, Q., Hu, C., 2017b. Targeted multivariate adulteration detection based on fatty acid profiles and Monte Carlo one-class partial least squares. Chemometr. Intell. Lab. Syst. 169, 94–99. https://doi.org/10.1016/j.chemolab.2017.09.002.

3- General discussion

The seeds used as oil sources in this study, *Balanites aegyptiaca*, *Opuntia ficus-indica* (cactus), and *Phoenix dactylifera* (date palm), differ markedly in their morphology and mechanical properties. These differences are illustrated in Figure 1. Among them, cactus seeds are the smallest (2–4 mm) and possess a relatively hard outer shell. Balanites seeds resemble almonds in texture, with a softer kernel that can be crushed by hand. In contrast, date seeds are extremely hard, making mechanical processing more challenging. All oils were initially intended to be extracted using a Komet single-screw cold press (IBG Monforts Oekotec GmbH & Co. KG, Mönchengladbach, Germany), as cold-pressed oils were the focus of this project.

Figure 1. seed and oils investigated in this dissertation

Using this press, oil from both Balanites and cactus seeds was successfully extracted. However, due to the hardness of date seeds, it was not possible to press them with the available equipment. Therefore, for DSO, solvent extraction was employed instead. With the mechanical press, the extraction yield for BKO was approximately 25%, and for CO, around 4%. According to reports from local cooperatives, a similar cold-pressing yield (~4%) is obtained for DSO. It is important to note that pressing does not extract the total oil content of the seed. Hence, for comparative purposes, we also assessed oil yields using Twisselmann solvent extraction, which generally provides higher recovery. Under these conditions, BKO showed a yield of approximately 44%, CO about 11%, and DSO a mean value of 9%. These results highlight BKO's significant advantage in terms of oil yield, especially considering that it comes from a desert tree. Although the oil yield of DSO is lower, it is still valuable, particularly since date seeds are typically an agricultural by-product or waste material. In terms of appearance, the three oils share a greenish-yellow to golden-yellow color, which aligns well with the visual expectations for edible oils. Color is a key attribute in consumer preference and plays an important role in product perception. Ensuring that oils have an appealing, natural color can contribute positively to marketability and consumer satisfaction (Espejel et al., 2008)

1.1.Chemical composition of the investigated oils

The following section gives an overview of the chemical composition of the investigated oils in terms of the major lipids, fatty acid (FA), triacylglycerol (TAG), tocochromanol, and phytosterol. To give a general overview of the chemical composition across the individual manuscript (I-III), for each oil sample group, the mean values for FA, TAG, tocochromanols, and phytosterols were calculated and summarized in Figure 2. This approach facilitates a straightforward comparison, while detailed compositions are reported in the individual manuscripts (I-III).

The measured data are compared to benchmark values defined in the Codex Alimentarius (Codex, 1999), which specifies typical compositional ranges for 23 common vegetable oils, as well as modified and fractionated oils. These benchmarks serve as a useful reference point to assess how the unconventional oils studied here relate to established commercial oils.

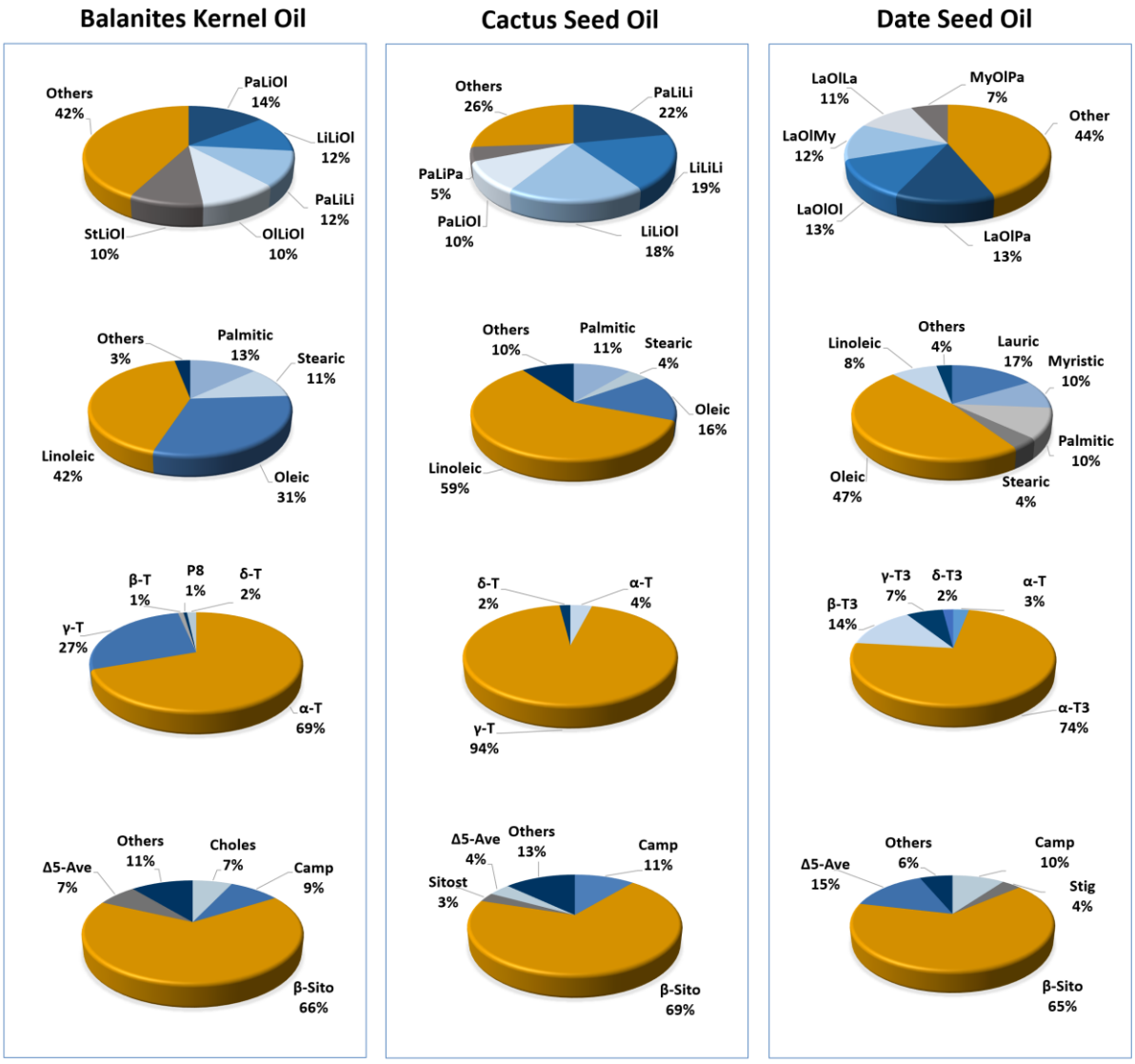


Figure 2. Overview of the chemical composition of BKO, CO, and DSO based on the mean percentage value of different variables taken from manuscript I-III. La: lauric acid, My: myristic acid, Pa: palmitic acid, St: searic acid, Ol: oleic acid, Li: linoleic acid; T: tocopherol, T3: tocotrienol; P8: plastochromanol-8, Choles: Cholesterol, Camp: Campesterol, Stig: Stigmasterol, β-Sito: β-Sitosterol, Δ5-Ave: Δ5-Avenasterol, Δ7-Ave: Δ7-Avenasterol, Sitost: Sitostanol

1.1.1. Fatty acid and triacylglycerol composition

Fatty acids represent the core fraction of vegetable oils, determining both their nutritional properties and applications. In both BKO and CO, linoleic acid was the predominant fatty acid, followed by oleic, palmitic, and stearic acids. The predominance of these four fatty acids in the same order makes CO and BKO comparable in fatty acid composition to sunflower, soybean, safflower, walnut, corn, and grapeseed oils, as reported in the Codex Alimentarius (Codex, 1999). A distinctive feature of CO is its relatively high content of vaccenic acid (~5%), a naturally occurring isomer of oleic acid. Although vaccenic acid exists in some vegetable oils, it is generally considered a rare fatty acid (Avato & Tava, 2022). This makes it a potential marker for the authenticity of CO in cases of complete substitution. However, if CO is adulterated with small amounts of another oil, only a dilution effect would be observed, making vaccenic acid less reliable as a marker in such cases. In terms of degree of unsaturation, BKO and CO are mainly unsaturated oils, with unsaturated fatty acids (UFAs) accounting for over 70% in BKO and more than 80% in CO. This is attributed to the high abundance of Li and Ol.

On the other hand, DSO shows a more diverse FA profile, including oleic (Ol), lauric (La), myristic (My), palmitic (Pa), linoleic (Li), and stearic (St). The UFA/SFA ratio in DSO ranges between 0.9 and 1.5, meaning that the levels of SFAs and UFAs are roughly equal. Comparing DSO to Codex benchmarks, its FA profile appears unique compared to all the other oils reported in the Codex. Although La (14.0–24.2%) and My (9.0–12.6%) are the second and third most abundant fatty acids in DSO, their levels are lower than in coconut oil or palm kernel oil, where these fatty acids dominate. DSO and palm oil are similar in terms of Ol (DSO: 40.8–50.2%; palm oil: 36.0–44.0%) and Li (DSO: 7.1–10.7%; palm oil: 9.0–12.0%), but their saturated fatty acid compositions differ considerably. In palm oil, the SFA fraction is mainly Pa (39.3–47.5%), whereas in DSO it consists of a more diverse mix of La, My, Pa, and St. This diversity contributes to DSO's ability to remain liquid at room temperature.

As FA are primarily stored as TAG, analyzing TAG profiles provides deeper insights into the oils structure and functional behavior. TAG molecules consist of three fatty acids esterified to a glycerol backbone, and their composition reflects both the types and distribution of fatty acids present in the oil. The distribution is not random and it is often plant-specific, with a tendency for unsaturated fatty acids to occupy the sn-2 position (Indelicato et al., 2017; Ngo-Duy et al., 2009; Yao et al., 2019). This positional specificity can influence both the metabolic fate and the physicochemical properties of the oil. In the investigated samples, the dominant TAGs by carbon number were C54 for BKO and CO, and C48 for DSO. In BKO, the main TAGs were PaLiOl, LiLiOl, PaLiLi, OlLiOl, and StLiOl. In CO, the identified TAGs included PaLiLi, LiLiLi, LiLiOl, PaLiOl, and PaLiPa which align with the FA content. Due to the high concentration of UFAs in both BKO and CO, nearly all their TAG molecules contained at least two unsaturated fatty acids. Disaturated TAGs were minor, and trisaturated TAGs were not detected. DSO showed a broader diversity in both FA and TAGs, with 34 quantifiable TAGs (compared to 13 in BKO and 11 in CO). The main TAGs in DSO were LaOlPa, LaOlOl, LaOlMy, LaOlLa, MyOlPa, MyOlOl, and PaOlOl. DSO also contained a

low concentration of medium-chain TAG (LaLaLa), which are hydrolysis more quickly and are beneficial rapid source of energy (Lauridsen, 2017).

1.1.2. Tocochromanol composition

Tocopherols and tocotrienols, collectively known as tocochromanols (or tocols), are lipid-soluble antioxidants naturally present in vegetable oils. Structurally, they share a chromanol ring, but differ in their side chains. Tocopherols have a saturated carbon tail, while tocotrienols possess an unsaturated isoprenoid chain with three double bonds (Bora et al., 2022).

The total tocochromanol content varied among the three studied oils: BKO (552–828 mg/kg), CO (500–680 mg/kg) DSO (424–760 mg/kg). Despite comparable total levels, the composition of individual tocochromanol homologues differed significantly. BKO was dominated by α-tocopherol (324–607 mg/kg) and γ-tocopherol (120–226 mg/kg), together accounting for over 90% of its total tocochromanol content. Minor components included β-tocopherol (2–17 mg/kg), δ-tocopherol (3–14 mg/kg), and plastochromanol-8 (2–12 mg/kg), with lower levels of γ- and β-tocotrienols. CO, however, was characterized by a high amount of γ-tocopherol (445–655 mg/kg), with minor contributions from α- (10–30 mg/kg) and δ-tocopherol (5–11 mg/kg). When comparing to Codex benchmark values, it can be noted that α-tocopherol and γ-tocopherol are also the main isomers found in many vegetable oils, especially those with high UFA content (Codex, 1999). On the other hand, the analyzed DSO represent a different profile as compared to BKO and CO. DSO tocochromanol was predominated with rather tocotrienol especially α-tocotrienol (324–563 mg/kg), which was its major constituent, followed by β- (46–169 mg/kg) and γ-tocotrienol (39 mg/kg), making tocotrienols the predominant group. α-Tocopherol was present in minor amounts (0–48 mg/kg), and no γ- or δ-tocopherol were detected. Tocotrienols, particularly γ and δ forms, have shown stronger antioxidant activity than tocopherols, especially at high temperatures (Deng et al., 2023; Trela & Szymańska, 2019). Only a few edible oils contain such high levels of tocotrienols. Palm oil and rice bran oil are known examples where tocotrienols make up the majority of the tocochromanol fraction.

It is worth noting that many previous studies have reported the presence of β- and γ-tocopherols in DSO, based on analyses using normal phase HPLC, the same technique employed in our study (Al Juhaimi et al., 2018; Besbes et al., 2004; Harkat et al., 2022; Nehdi et al., 2010; Nehdi et al., 2018). However, we did not detect these two tocopherols in any of the DSO samples analyzed (30 samples analyzed). This may indicate that they are either genuinely absent from our samples or present at levels below our limit of detection. It is therefore possible that the DSO samples analyzed in our study truly lack these tocopherols, while those in other studies may have contained them. Interestingly, we consistently observed two additional peaks at approximately the same retention times. When overlaid with authentic β- and γ-tocopherol standards, these peaks (labelled x and y in Figure 3) eluted immediately after the expected positions of β-and γ-tocopherols but did not overlap with the standard peaks, indicating that they do not correspond to these compounds.

Some LC-MS/MS measurement were performed, but at this stage we could not conclusively identify these unknown peaks.

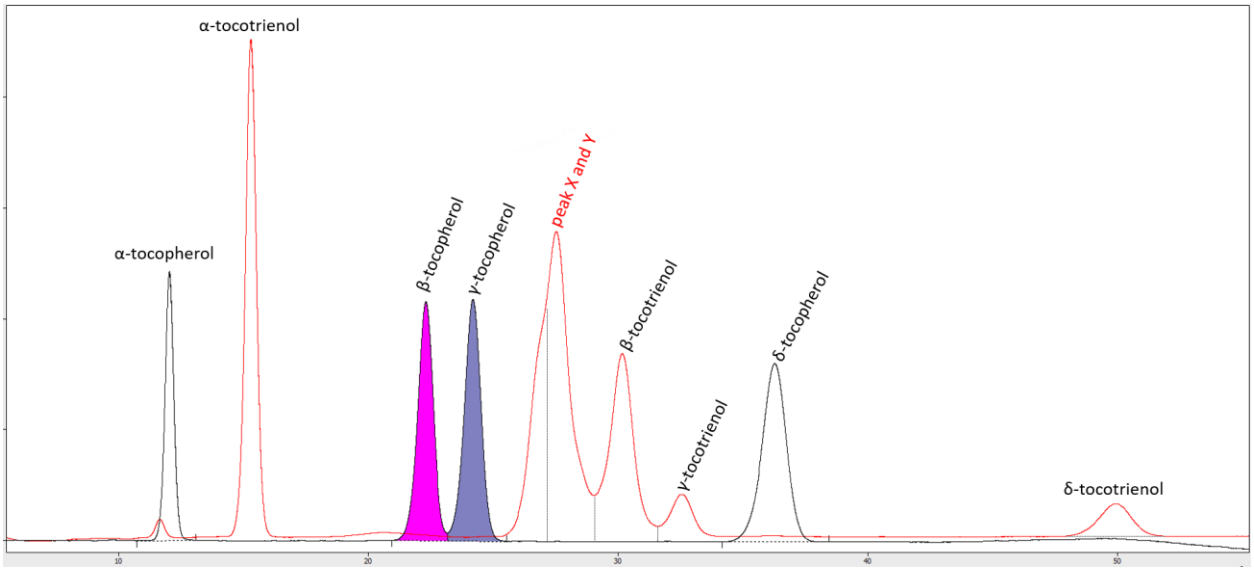


Figure 3 . Overlay of the chromatogram of a date seed oil sample (red trace) and a mixture of authentic tocopherol standards (black trace). The peaks corresponding to β- and γ-tocopherol in the standard are highlighted with colors for comparison.

1.1.3. Phytosterol composition

Phytosterols represent the major components of the unsaponifiable fraction of oils. Among the three oils, CO had the highest phytosterol content (8292–11091 mg/kg), followed by DSO (3422–4827 mg/kg), and then BKO (871–2218 mg/kg). In all three oils, β-sitosterol was the major phytosterol, comprising about 70% of the total phytosterol content. The other key phytosterols were campesterol and Δ5-avenasterol, which are also the major sterols in many vegetable oils such as cottonseed, linseed, and hazelnut oils, according to Codex standards (Codex, 1999). It is worth noting that the phytosterol content in CO is higher than in most edible oils listed in the Codex benchmark. The only exceptions are corn oil, rapeseed oil, rice bran oil, and sesame oil, where either the phytosterol content is higher or CO falls within their reported ranges. Phytosterols have been shown to retard oxidative polymerization in vegetable oils subjected to frying, as evidenced by reduced iodine values and fewer formation of polymer products (Frankel, 2005).

Among the analyzed oils, BKO stood out with a relatively high cholesterol content of about 7%. In comparison, DSO contained around 1%, and although cholesterol was not detected in the analyzed CO samples, previous studies have reported its presence at around 1% of total phytosterols. (Al-Naqeb et al., 2021). It seems to be a common misconception that cholesterol is found only in animal products. In fact, most plants contain relatively small amounts of cholesterol. Many vegetable oils listed in Codex standards also contains cholesterol. For instance, palm oil can reach levels comparable to those observed in BKO, with Codex reporting up to 7% cholesterol content (Codex, 1999). Another well-known example is camelina oil, which also contains relatively high levels of cholesterol (Kirkhus et al., 2013). Particularly noteworthy is the research by Sonawane et al., who made a significant contribution to understanding cholesterol biosynthesis in plants. They elucidated the biosynthetic pathway, which consists of ten

enzymatic steps starting from 2,3-oxidosqualene (Sonawane et al., 2016). These findings confirm that cholesterol is a natural component of plant sterol metabolism.

In summary, BKO, CO, and DSO, as potential oil sources from the Moroccan desert, share several similarities in their chemical composition with other well-known oils reported in Codex standards, while also showing a few particularities. BKO and CO were predominantly unsaturated oils with high Li and Ol content, while DSO exhibited a more balanced FA profile and the broadest diversity of TAGs. Tocochromanol composition also varied significantly, with BKO rich in α-tocopherol, CO in γ-tocopherol, and DSO dominated by tocotrienols. CO showed the highest phytosterol levels, though all three oils shared a profile dominated by β-sitosterol, campesterol, and Δ5-avenasterol.

These compositional traits influence the storability and oxidative stability of the oils. BKO and CO, being rich in unsaturated FAs, are more susceptible to oxidation. Therefore, proper storage practices such as refrigeration, dark storage, and nitrogen flushing are recommended. In contrast, DSO's lower UFA/SFA ratio and high tocotrienol content contribute to excellent stability. This was supported by Rancimat results, where DSO showed the longest induction time (62.5 hours) and highest activation energy (100.46 kJ/mol), compared to BKO (6.5 hours; 84.7 kJ/mol) and CO (6.6 hours; 92.4 kJ/mol) (El Harkaoui et al., 2023). These properties make DSO well-suited for longer shelf life particularly in arid areas with high ambient temperatures. These findings also support the potential application of DSO in culinary and food preservation contexts or as a blend with other oils to improve their oxidative stability

From a nutritional perspective, several key insights emerge. The fatty acid composition of BKO and CO, dominated by Li, an essential omega-6 fatty acid, and Ol, a monounsaturated fatty acid, support the reduction of LDL cholesterol and the improvement of lipid profiles. DSO, while higher in SFAs, still provides substantial Ol content and an UFA/SFA ratio of around 1.2. It also contains LaLaLa, a medium-chain TAG associated with rapid energy metabolism (Lauridsen, 2017). Historically, saturated fats have been associated with cardiovascular disease, and oils like palm oil, comparable to DSO in composition, have sometimes been classified as unhealthy due to their high SFA content. However, saturated fatty acids make up roughly half of the fatty acids in membrane phospholipids and about one-third in triacylglycerols in healthy individuals (Lands, 2008). Thus, the overall saturation level is not the only factor influencing health; the specific type of saturated fatty acid and its position in the TAG molecule also play key roles (Zock et al., 1994; Kritchevsky et al., 2000).

BKO's richness in α-tocopherol is notable, as this is the most bioavailable form of vitamin E and is preferentially retained in the human body (Brigelius-Flohé et al., 2002). CO's high γ-tocopherol content is also beneficial due to its superior antioxidant activity. DSO, with its tocotrienol-rich profile, offers an additional nutritional dimension, as tocotrienols have neuroprotective, anti-cancer, and heart-protective effects not seen in tocopherols (Babura et al., 2017; Nesaretnam et al., 2007). β-sitosterol, campesterol, and Δ5-avenasterol, particularly abundant in CO, are important phytosterols in food applications due to their significant health benefits, particularly in lowering cholesterol levels and reducing cardiovascular risks.

Their incorporation into functional foods enhances the nutritional value and health benefits of these products (Faubel et al., 2024; Normén et al., 2007). In addition, phytosterols have been shown to retard oxidative polymerization in vegetable oils subjected to frying, as evidenced by reduced iodine values and fewer polymer products, which also contributes to health-related benefits (Frankel, 2005).

1.2.Variation of the chemical composition as related to the geographical origin

Origin-related variations were evaluated for BKO, CO, and DSO each of which was investigated in a dedicated study. Due to natural constraints related to plant distribution, accessibility, and availability of regional production, the sampling structure differed across the three manuscripts (I–III). To be sure about the location always seeds were collected and the oil extraction was done in the lab. For BKO, samples could only be collected from the Tata region over three years, thus to assess geographical variation, these samples were compared with others from Sudan and Mauritania (manuscript I). In the case of CO, six single-location samples from different regions of Morocco were analyzed. This approach was necessary due to limited access with verified geographic origins from cooperatives, as cactus plantations had been severely affected by the pest *Dactylopius opuntiae*, which damaged many of the cactus-growing areas. DSO was examined through a more extensive Moroccan dataset (26 samples collected from 3 palm groves) supplemented by a few samples from out of Morocco.

To assess the effect of geographical origin, we combined chemical composition data with PCA to examine clustering trends and correlations. For CO, where only one sample per site was available, comparisons were limited to univariate correlation and variance testing. This allowed us to detect both direct trends in compound concentrations and possible underlying patterns driven by environmental factors such as temperature, rainfall, or altitude.

1.2.1. Variation in fatty acid and triacylglycerol composition

Variation in FA composition related to geographical origin was observed across the three studied oils, with the most pronounced differences seen in BKO. To understand this variation, it is necessary to consider the biosynthetic pathway of FA. The proposed FA pathway highlights a reciprocal quantitative relationship among different FAs, regulated by various enzymes such as ketoacyl synthase and fatty acid desaturase, which mediate the interconversion between FAs (Porokhovinova et al., 2022). The geographical location (or the environmental conditions associated with it mainly water deficit and/or temperature) can influence the activity of these enzymes, thereby shifting FA profiles as a physiological response to environmental stress. Numerous studies have discussed the impact of these factors on the FA composition of various oils. For instance, Li synthesis in cottonseed, sunflower, and camelina oil was shown to increase under water deficit or lower temperatures and it was assumed that this is due to enhanced activity of oleate desaturase, the enzyme that converts Ol to Li (Porokhovinova et al., 2022). Similarly, in sunflower oil, drought stress led to increases in both Li and Pl acids, accompanied by decreases in St and Ol acids. While one explanation for lower Ol levels is its enzymatic desaturation to Li, another hypothesis links the involvement of St in signaling and defense mechanisms (Kachroo et al., 2008), which could divert precursors away from Ol synthesis under stress (Ghaffari et al., 2023). A similar FA pattern, rising Li and Pl with declining Ol and St, was also observed in rapeseed oil under water deficit conditions (Hatzig et al., 2018). These explanations and hypotheses are also relevant to the case of BKO, which shows comparable trends. Particularly when

comparing Moroccan samples to those from Mauritania and Sudan we can see in biplot (Figure 6, manuscript I) that the Moroccan samples correlate positively with Pa and Li and negatively by St and Ol. However, whether this represents a generalizable trend is questionable. Other studies conducted under controlled field conditions have reported different FA variation patterns in response to water deficit and/or temperature (Bettaieb Rebey et al., 2012; Laribi et al., 2009). These discrepancies suggest that stress factors like water deficit and/or temperature influence desaturase enzyme activity in ways that are species-specific. For DSO, variation in FA composition was observed between Moroccan and non-Moroccan samples, particularly in St, Ol, and Pa contents (Figure 3 and Figure S1, manuscript III). While the same biosynthetic explanations may apply here, climatic data from non-Moroccan sites were not available, limiting interpretation. Within Morocco, no significant FA differences were found between the three palm groves, likely due to similar environmental conditions, especially in terms of temperature and rainfall. In the case of CO, although some regional variation was observed, it was minimal and had no influence on the unsaturation index across the six samples. The contents of SFA and UFA were consistent across all samples.

Interestingly, while fatty composition did not show strong variation across the three Moroccan palm groves, TAG composition did. In fact, TAGs emerged as key variables contributing to the clustering observed in the PCA biplot (Figure S1, manuscript III). This observation is noteworthy, as it suggests that the climatic conditions may exert a more pronounced effect on TAG assembly pathways than on the overall fatty acid biosynthesis itself. This hypothesis is supported by the possibility that enzymes involved in TAG synthesis, such as lysophosphatidate acyltransferase and diacylglycerol acyltransferase, may exhibit isoform-specific expression or activity shifts depending on environmental conditions. These shifts could alter the incorporation pattern of individual fatty acids into TAG molecules, resulting in region-specific TAG profiles even when the underlying FA pool remains relatively constant. A similar explanation has been proposed in studies on olive oil, where TAG variability was observed between regions, although direct enzymatic evidence for this mechanism remains limited (Stefanoudaki et al., 2009).

1.2.2. Variation in tocochromanol composition

Environmental stress also appears to influence tocochromanol composition. For example, water deficit has been shown to increase total tocopherol content in oils such as almond, olive, and soybean, indicating a positive correlation between water deficit and tocopherol accumulation (Carrera & Seguin, 2016; Kodad et al., 2018; Romero et al., 2003). Cold temperatures have also been linked to increased tocochromanol synthesis, through enhanced expression of p-hydroxyphenylpyruvate dioxygenase (HPPD), a key enzyme in the biosynthetic pathway (Chaudhary & Khurana, 2009). In addition to changes in total content, stress conditions may alter the ratios between tocopherol homologues, particularly the α-tocopherol/γ-tocopherol and δ-tocopherol/β-tocopherol ratios. These changes have been associated in the literature with the potential effects of water deficit and/or temperature on γ-tocopherol methyltransferase (γ-TMT), an important enzyme responsible for regulating the α-tocopherol/γ-tocopherol and δ-tocopherol/β-tocopherols (Carrera & Seguin, 2016; Kodad et al., 2018; Lushchak & Semchuk, 2012).

In line with this literature, notable differences in tocochromanol content were observed among the studied oils, particularly in relation to geographical origin. For BKO, Moroccan samples from Tata, an arid region with strong seasonal temperature variation (ranging from 6.1 to 41.5 °C annually) and very low precipitation (64 mm/year), exhibited significantly higher tocochromanol levels than samples from Mauritania and Sudan. These samples also showed elevated α-tocopherol/γ-tocopherol ratios, possibly reflecting increased γ-TMT activity under water deficit and/or temperature stress. These trends are further supported by the biplot in manuscript I (Figure 6), where Moroccan samples aligned with higher tocochromanol content. These findings may support the hypothesis that cold and water deficit promote tocopherol biosynthesis. In CO, higher tocopherol contents were found in samples from *Tiznit* and *Hoceima*, two areas relatively close to the coast and at a lower altitude. Some studies suggest a negative correlation between tocopherol content and altitude or distance from the coast (e.g., in olive and argan oil), although these remain speculative and not supported by biosynthetic evidence (Elgadi et al., 2021; Mohamed Mousa et al., 1996). For DSO, this hypothesis did not hold. Despite *Allougoum* being the lowest-altitude site, it showed lower tocochromanol content than *Errachidia* and *Alnif*. This indicates that factors beyond altitude, such as plant variety or other microclimate effects, may influence tocopherol accumulation in DSO.

1.2.3. Variation in phytosterol composition

Several studies have reported that phytosterol content varies in response to water deficit and/or temperature stress, likely due to their role in regulating membrane fluidity and permeability (Schaller, 2003). In sunflower oil, controlled field trials confirmed an increase in total phytosterol content under water deficit and/or temperature effect (Roche et al., 2006). A study on rice under water deficit stress found a proportional increase in phytosterols with increasing water stress duration. This was attributed to higher expression of HMG-CoA reductase, the rate-limiting enzyme in phytosterol biosynthesis (Kumar et al., 2018). Indeed, water deficit and/or low temperatures (cold stress) were also found to increase phytosterol levels (Rogowska & Szakiel, 2020). This reasoning may apply to the BKO samples, where Moroccan oils had higher phytosterol levels compared to those from Mauritania and Sudan. Similarly, in CO, the *Tiznit* sample showed higher phytosterol levels than others, possibly reflecting local climate stress. For DSO, however, no clear trends were found. Rainfall levels were relatively similar across the three Moroccan groves, and while *Errachidia* has a slightly higher altitude and cooler climate, phytosterol content did not correlate with these factors. Although some authors reported a link between altitude and sterol composition, such a pattern was not observed in our DSO dataset.

In summary, geographical origin was found to influence the chemical composition of all three oils, though the extent of variation and the compounds contributing to regional clustering varied depending on the sample set of each oil. Moroccan BKO samples were clearly distinguishable from those collected in Sudan and Mauritania, with clustering mainly driven by major components across all four chemical classes, FA, TAG, tocochromanols, and phytosterols. In contrast, for DSO, where the comparison was focused on a

smaller geographical scale within Morocco, origin-related differences were primarily linked to TAG composition, especially minor TAGs. To interpret the observed variations, the impact of environmental conditions such as water deficit and/or temperature on biosynthetic pathways was considered. These included the influence of stress factors on enzymes like ketoacyl synthase and fatty acid desaturase (for FA), p-hydroxyphenylpyruvate dioxygenase (for tocopherols), and HMG-CoA reductase (for phytosterols), as reported in the literature.

1.3. Effect of geographical origin on the metabolomic profile

Following the analysis of major lipid components in DSO as reported in manuscript III, the investigation was extended toward a more comprehensive understanding of the oil's chemical complexity using an untargeted metabolomics approach. The polar fraction of DSO samples was extracted using MeOH/H_2O and analyzed via UHPLC-ESI-QTOF-MS in both positive and negative electrospray ionization modes, thereby maximizing metabolite detection. A higher number of features were detected in the positive mode, a trend commonly attributed to higher ionization efficiency for many metabolite classes and possibly to instrumental settings favoring positive ion formation. This observation aligns with previous oil metabolomics studies (Dou et al., 2025; Gil-Solsona et al., 2016; Hu et al., 2022; Willenberg et al., 2021). The resulting data underwent a comprehensive and carefully structured processing workflow, including quality control assessment, data filtering, LOWESS normalization, transformation and scaling, multivariate analysis, and finally, feature annotation (Figure 2, manuscript IV). This workflow was considered upon reviewing different recommendations for processing metabolomics data (Broadhurst et al., 2018; Di Guida et al., 2016; Mattoli et al., 2022; Schymanski et al., 2014; Sumner et al., 2007; van den Berg et al., 2006).

PCA was initially used to explore differences among DSO samples from the three Moroccan palm groves: *Allougoum*, *Alnif*, and *Errachidia* (Figure 4, manuscript IV). The PCA score plots revealed no clear separation between *Alnif* and *Errachidia* samples, which is consistent with their geographical proximity and shared environmental conditions. However, samples from *Allougoum*, located further south at a lower altitude and closer to the Atlantic coast, showed a modest but consistent clustering trend in both ionization modes. This observation aligns with previous findings based on chemical composition, where *Allougoum* samples also demonstrated a slightly distinct profile. Although the overall variation between regions was moderate, the trend observed for *Allougoum* samples was consistent, which made it interesting to further investigate the features responsible for this difference. For this reason, an OPLS-DA was performed by grouping the samples from *Alnif* and *Errachidia* into one group (called "AlEr") and comparing them with *Allougoum* samples. Based on this model, VIP scores (Figure 5, manuscript IV) and S-plots (Figure 6, manuscript IV) were generated for both ionization modes to help select the most important features that contribute to the separation. Features were ranked based on their VIP scores, and the top 50 were selected for further investigation and annotation. The identified features are reported in Tables 1 and 2, manuscript IV, which include information about their VIP score, reflecting how important a feature is in distinguishing the samples; retention time, which provides insights into chromatographic behavior and supports compound identification; and the annotation level, based on confidence level from 5 to 1.

In total, 25 metabolites from different chemical classes were tentatively annotated at various confidence levels. All of these compounds are reported for the first time in DSO, except for vanillin, which was already known. Among the 25, three compounds were confirmed with authentic standards: n-(3-oxohexanoyl) homoserine lactone, indole-3-carboxaldehyde, and vanillin. The most represented class among the

identified markers was hydroxy fatty acids, with four compounds (hydroxy tetradecenoic acid, dihydroxy tetradecanoic acid, 12-hydroxydodecanoic acid, and dihydroxy tetracosadienoic acid) tentatively identified. It is currently difficult to fully explain why some of these compounds appeared as top features, or what their exact metabolic origin or biosynthetic pathways might be. However, some assumptions and possible explanations were made. For example, the presence of hydroxy fatty acids might be linked to the activity of desaturase enzymes, which in addition to desaturation, can also catalyze hydroxylation reactions under certain conditions, especially when mutations or changes in enzyme–substrate interactions occur (Cahoon & Li-Beisson, 2020; Schmid, 2021). Another compound of interest is octacosyl ferulate, which was the top-ranked feature in the negative mode. This compound may originate from the esterification of ferulic acid with octacosanol, which is supported by the presence of ferulic acid in DSO and by the fact that octacosanol has been previously reported in other vegetable oils (Jung et al., 2011).

The distribution of the tentatively identified features is visualized in box plots (Figure 7, manuscript IV), which show their content in each DSO sample. The box plots confirm that all identified metabolites are present in all samples, meaning that the geographical origin affects content levels rather than the presence or absence of certain metabolites. This is expected since the samples belong to the same oil type. Some of the compounds were found in higher content in *Allougoum*, while others were more concentrated in the AlEr group. In addition, the box plots revealed wide dispersion of the samples from the same location within the cluster. Since the samples were all collected at the same maturity stage (*Tamr*) and processed under identical extraction and storage conditions, this variation was mainly explained by the different date varieties included in the study. A similar dispersion was also observed in the PCA plot and discussed in manuscript III.

In summary, this section presented the full processing workflow starting from UHPLC-ESI-QTOF-MS analysis to the annotation of features. DSO was used as a case study, but this approach can be applied to other oils like BKO and CO if more samples are available. While each oil may have a different metabolomic profile, the data processing workflow remains applicable. A total of 50 features were investigated, 25 metabolites were tentatively annotated, and 3 were confirmed by authentic standards. The identified markers belonged to various chemical classes and were detected in all samples, showing that although the effect of geographical origin was moderate, it was clear and occurred mainly at the level of content differences, rather than in the presence or absence of specific compounds.

1.4.Adulteration detection using analytical data and machine learning

Adulteration detection is another important part of authenticity that was included in this work, especially because economically motivated adulteration affects both consumer trust and market integrity. The aim here was to develop a machine learning model that can detect adulteration without relying on large sets of physical mixtures. Instead, the approach combines analytical data with simulation algorithms to generate enough artificial data to build robust classification models. CO was chosen as the case study, and refined SO was used as the adulterant. CO was selected because of its high price (around 600 €/L in bulk), and its more established production and commercialization compared to BKO and DSO.

First, sample sets of pure CO and SO were analyzed for their FA, TAG, and tocochromanol composition using GC and HPLC. This multi-aspect analysis approach (using multiple chemical parameters) was followed to strengthen detection, as it makes it harder for counterfeiters to manipulate just one parameter to escape detection. The resulting compositional database formed the basis for subsequent simulations, which were then used to train classification models to assign samples to specific adulteration levels.

1.4.1. Simulation of the data

Generating large datasets for training classification models in adulteration detection typically requires preparing and analyzing numerous pure oil samples and mixtures at varying adulteration levels. This conventional approach is labor- and resource-intensive, involving physical sample preparation and extensive measurements. To reduce workload while maintaining model robustness, data augmentation can be applied, creating new data by modifying existing measurements or simulating datasets based on artificial intelligence and statistical modeling (Georgouli et al., 2018; Gracia Moisés et al., 2023; Shorten & Khoshgoftaar, 2019).

In this study, two simulation models, Monte Carlo (MC) and Conditional Tabular Generative Adversarial Network (CTGAN), were employed to simulate data. The goal was to compare how well they could simulate oil data that reflected the statistical properties of the original samples. Both methods were used to simulate different sample sizes for pure CO and SO, ranging from 1,000 to 1,000,000 samples. Simulation performance is often evaluated using box plots to visually compare the distributions of raw and simulated data (Cui et al., 2024; Zhang et al., 2022). However, for multivariate datasets with many variables, this approach is impractical, as each variable would require a separate plot. Instead, we used the Kullback-Leibler Divergence (KLD; Formula 1, manuscript V), which provides a single quantitative measure of similarity between multivariate distributions. Unlike univariate comparisons, KLD accounts for the dependencies between variables, capturing the overall structure of the data (Kullback & Leibler, 1951). KLD values reported in Table 1 of manuscript V are given in units of 10^3; for clarity, discussion values are presented only in absolute terms (e.g., 1.2 corresponds to 1,200; 89.5 corresponds to 89,500).

The MC method showed consistent performance across all dataset sizes, maintaining low KLD values, especially from 10,000 samples onward. For CO, values ranged between 2.3 and 3.1, and for SO, between

1.2 and 1.4. A noticeable improvement was observed when increasing the number of simulations from 1,000 to 10,000, with KLD dropping from 3.1 to 2.5 for CO and from 1.4 to 1.2 for SO. In contrast, CTGAN's KLD values were much higher and strongly influenced by the number of training epochs. At lower epochs (100–1,000), the simulations were significantly less accurate. Increasing dataset size had only a minor effect on CTGAN performance, whereas increasing training epochs provided more improvement but at a high computational cost. At 50,000 epochs, CTGAN achieved its best performance with KLD values of 89.5 for CO and 68.2 for SO at 10,000 simulations which is still several times higher than those of MC. The superior performance of MC can be attributed to its statistical approach, which samples from the original data distribution using the mean and covariance matrix. This method does not require large datasets to perform effectively, unlike CTGAN, a deep learning model that needs large and diverse training data to learn complex patterns. With only 27 CO and 10 SO original samples, CTGAN likely lacked sufficient data for optimal training.

In the literature, MC has been applied as a simulation model for oil adulteration detection in several studies (Zhang et al., 2014; Zhang et al., 2017a; Zhang et al., 2017b), where small datasets of oil composition were first analyzed and then used to generate simulated data (300, 1,000, or 10,000 samples) for model training and testing. Although these studies did not report explicit performance metrics or compare MC with alternative simulation methods, its repeated use across different oils suggests a robust and generalizable approach. Our findings further support that MC performance is largely independent of the specific chemical composition of the oil, provided that variables follow a normal distribution.

In contrast, to the best of our knowledge, GAN-based models have not yet been applied to the chemical composition of oils using conventional tabular datasets, making direct comparisons with similar studies difficult. However, other GAN variants have been successfully applied in related contexts using hyperspectral images to generate sufficient simulated data for model training. For example, deep convolutional GANs have been used to predict oil content in maize (Zhang et al., 2022), and autoencoder-assisted GANs to predict polyunsaturated fatty acid content in meat (Cui et al., 2024). After multiple training iterations, these models were able to generate synthetic data closely resembling experimental data. While the use of GAN in food authenticity research remains in its early stages (Deng et al., 2024), these examples highlight that the performance of GAN-based approaches is highly dependent on the specific application context, including the data type, dataset size, and study objectives.

Based on all these facts, MC was selected for data simulation. The simulated pure oils data were then used to generate artificially adulterated samples using a weighted sum formula (Formula 2, manuscript V). The number of samples in each class is shown in Table 2 (manuscript V) which served as the input for the classification step.

1.4.2. Classification models

For the classification task, Random Forest (RF) and Neural Network (NN) models were employed. Both were trained on 80% of the MC-simulated data from each class (see Table 2, manuscript V), with the remaining 20% used for testing. For example, from 1,000 samples, 800 were used for training; from 10,000, 8,000; and so on. Hyperparameter tuning was applied to both models to improve performance and avoid overfitting (A Ilemobayo et al., 2024). The optimized parameters are listed in Tables S1 and S2, manuscript V. The models were then evaluated using accuracy, precision, recall, and F1-score, first on simulated test data (Section 3.4.2.1) and then on real-world data (Section 3.4.2.2).

1.4.2.1. Evaluation of RF and NN using simulated data

The RF model showed an improvement in classification accuracy as the number of simulated samples used to train the model increases. Starting with 1,000 samples per class, RF achieved an accuracy of 83.7%. This improved to 90.5% with 10,000 samples per class, and continued to rise, reaching 94% with 1,000,000 samples per class. The trend suggests that RF benefits from larger training datasets, which likely allows the model to capture more complex patterns and variability in the data. However, the accuracy increases from 100,000 samples per class (92.5%) to 1,000,000 samples per class (94%) were relatively modest, suggesting that RF may be approaching a performance plateau. The NN model, on the other hand, behaved differently. It started with 92% accuracy at 1,000 samples, already outperforming RF at this stage. But after that, the accuracy remained stable, around 92–93%, even with 1,000,000 samples.

Previous studies have shown that classification models, including those applied in non-food sectors, generally achieve improved accuracy when trained on larger datasets (Qiu et al., 2018; Liu et al., 2018; Rocha, Werickson Fortunato de Carvalho et al., 2020). NNs, in particular, tend to benefit more than RF from large datasets due to their capacity to learn complex, non-linear representations. Therefore, it was expected that the NN would outperform the RF, especially when trained on the largest sample set (1,000,000 samples). However, this improvement was not observed in our study. A likely explanation was that NN models require extensive hyperparameter tuning to fully exploit large datasets, which was not feasible in our case due to hardware limitations. Consequently, full tuning could not be performed for the 100,000 and 1,000,000 sample scenarios, potentially limiting the NN's performance. Another possible reason was that the dataset may not contain a significant degree of complex relationships beyond those already captured at smaller dataset sizes, resulting in an early performance plateau. These factors together may explain why RF continued to benefit from additional data, while NN performance remained essentially unchanged. Further exploration of expanded hyperparameter ranges and alternative network architectures could help assess whether higher accuracies are achievable.

When looking at confusion matrices, both models could clearly separate CO from SO, regardless of dataset size. The most common errors occurred in distinguishing very similar classes, such as 99:1 from pure CO, or 95:5 from 97:3. In other words, adulteration level differences of ±2% were sometimes misclassified.

However, adulteration levels ≥3% were reliably detected. Interestingly, at 1,000 samples, NN performed slightly better than RF. But as the dataset grew, RF caught up and eventually outperformed NN. Even for 1% adulteration, RF misclassified only 13–14% of samples, still an acceptable result.

1.4.2.2. Evaluation of RF and NN using real word samples

To test if the models trained on simulated data could also work in real-life situations, we tested them using a small real-world dataset of 33 adulterated CO samples, representing about 2.4% of all possible combinations (27 CO × 10 SO × 5 levels). The real-world evaluation (Table 4, manuscript V) showed that accuracy dropped compared to simulated data, but still remained above 70% in all cases, which was quite good given the seven classes, natural variation, and the small sample size. Considering the performances of RF, the differences between simulated and real-world data became smaller when the size of the simulated datasets becomes larger. The smallest difference of 4% points was reached for the RF trained on one million samples of simulated data. This suggests that with larger datasets not only the performances of RF became better, but also the capacity of the model to generalize was enhanced, so that the risk of overfitting was reduced. Considering the performances of NN, the accuracies were in most cases lower than the RF so we can see that RF clearly outperformed NN. Therefore, the NN seemed to be less well generalizable to real world conditions as the RF. This behavior is often observed when a model focuses too much on certain given data points rather than learning the overall pattern.

From an application perspective, model interpretability is also important. RF is more transparent, as its decision trees can be followed and explained, which is helpful for regulatory use. NN behaves more like a black box, which makes it harder to use when interpretability is needed (Roßbach, 2018). Furthermore, RF models are computationally less demanding, requiring less time and fewer hardware resources for training and optimization compared to NN, which can be demanding when working with large datasets which seem to be needed to capture the observed data sufficiently. Although NN shows stable performance and has the ability to capture complex patterns, the comparable and often better accuracy of RF in our case, combined with its interpretability and computational efficiency, suggests that RF is the more suitable model for this particular scenario.

In summary, this study showed that MC is a reliable and efficient simulation tool for oil composition data. It can generate realistic data even when the original dataset is small. CTGAN, while powerful in theory, needs a lot of data and long training times to perform well. RF outperformed NN, reaching 94% accuracy on simulated data and 90% on real-world samples, and could detect adulteration levels as low as 1%. RF was also easier to interpret and less demanding computationally, making it a better model for oil authenticity testing. All original datasets are shared open access in the accompanying data publications (El Harkaoui et al., 2024a, 2024b, 2024c), and the full methodology, implemented in Python, has been uploaded to GitHub for transparency and reproducibility (see manuscript V). While CO was used as a case study due to its more advanced production and commercialization, the same methodology can be applied to BKO and DSO with appropriate retraining and model tuning.

4- Concluding Remarks and Outlook

An important dimension of this PhD project, beyond the analytical methods and statistical models, was the transfer of knowledge to the local cooperatives involved in this project. These cooperatives play a central role in the extraction and commercialization of unconventional oils in Morocco, and their involvement is crucial to ensuring that the benefits of scientific research are embedded within local communities. To this end, results and conclusions were shared through workshops and web meetings organized with cooperative members. These sessions aimed to raise awareness about the importance of oil authenticity, in parallel with the quality aspect as a decisive factor for long-term market acceptance. Discussions included practices that could undermine authenticity, such as mixing seeds from different regions without traceability or processing them in locations different from their origin without proper labeling. While such practices may go unnoticed today due to the niche status of these oils, they pose risks to consumer trust and the establishment of credible origin-based claims in the future. Communication with the cooperatives was horizontal and culturally grounded. Being of Moroccan origin and fluent in both Amazigh and Arabic languages, allowed for a more open, respectful, and effective exchange. Many of the samples used in this research were provided free of charge, but this was not viewed as a one-way transaction; the cooperatives expressed strong interest in the scientific findings and saw the research as a form of promotion for their products and, more broadly, for Moroccan cooperatives. The feedback received underscored a clear desire for continued dialogue and knowledge exchange, suggesting that the scientific contribution supports community engagement and capacity-building.

To provide a scientific basis for assessing the authenticity of the investigated oils, this dissertation introduced a multi-approach framework developed through the study of BKO, CO, and DSO as potential vegetable oil sources in the Moroccan desert. It is considered a multi-approach because it integrates complementary levels of investigation: targeted and untargeted analyses, and the characterization of different compound classes including fatty acids, triacylglycerols, tocochromanols, and phytosterols. It is described as a framework because the workflow is structured yet adaptable, allowing extension to other oils or adulteration scenarios beyond the case studies presented here.

Compositional analysis of fatty acids (FA), triacylglycerols (TAG), tocochromanols, and phytosterols revealed that each oil possessed a distinct lipid fingerprint. BKO and CO were characterized by a high proportion of unsaturated FAs, particularly oleic and linoleic acids, while DSO showed a more balanced composition, including a relatively higher proportion of saturated fatty acids, contributing to its enhanced oxidative stability. TAG profiling confirmed these findings, with BKO and CO exhibiting TAG distributions dominated by unsaturated molecular species, whereas DSO displayed a more complex TAG profile with higher molecular diversity. In terms of tocochromanols, BKO was rich in α-tocopherol, CO in γ-tocopherol, and DSO was distinguished by a high level of tocotrienols. All three oils shared a similar phytosterol composition, dominated by β sitosterol, campesterol, and Δ5-avenasterol, with CO showing the

highest total content. These compositional traits not only reflect oil quality but also contribute to oil identity. The profiles established here provide a basis for a reference database that, with more data, could contribute to regional standards, similar to Morocco's SNIMA standard for argan oil.

Despite limitations in sample size and regional coverage, multivariate analyses revealed clear origin-related trends. While some assumptions (e.g. normality) had to be made, care was taken in interpreting the results. Nevertheless, meaningful patterns emerged. Moroccan BKO samples were distinctly different from Sudanese and Mauritanian ones across all lipid classes, while minor TAG profiles differentiated DSO from nearby Moroccan palm groves. These differences reflect environmental impacts on biosynthesis and enzymatic activities. However, the "effect of geographical origin" is complex, arising from overlapping factors such as climate, soil, sunlight, salinity, and their timing during plant growth, making it difficult to isolate specific causes, especially under wild conditions. Nevertheless, the resulting chemical patterns served as valuable markers for authenticity. For example, the distinction of Moroccan BKO was linked not only to fatty acid variations but also to higher tocochromanol and phytosterol content, compounds known for their nutraceutical value and contribution to oil stability. This advantage motivates the potential development of region-specific quality labels.

The untargeted metabolomic profiling of DSO using UHPLC-QTOF-MS further expanded the understanding of regional variation. Focusing on the polar fraction, this analysis confirmed the same geographical clustering observed in the analysis of the major lipids, with *Allougoum* samples again separating in PCA. Among the top 50 discriminating features in the OPLS-DA model, 25 metabolites were tentatively identified and reported for the first time in DSO. Importantly, the identified markers were present in all samples, indicating that origin-based differentiation was based on content differences rather than the presence or absence of unique compounds. While the strongest compositional shift was observed in *Allougoum* samples, the overall differences between Moroccan regions were moderate, suggesting a relatively stable baseline composition of Moroccan DSO. Most metabolites were identified at tentative confidence levels, and further validation using authentic standards is recommended where possible. Additionally, confirming these results with a larger sample set would help assessing the robustness of the observed patterns. In the future, once confirmed, these markers could be monitored using simpler analytical techniques, making routine analysis more accessible for laboratories with limited resources. The data processing workflow was based entirely on open-source tools, including MS-Dial and MetaboAnalyst, and could be readily applied to other oils such as BKO and CO.

Whether geographical differentiation is necessary for BKO, CO, and DSO remains an open question. Current conclusions are constrained by sample size, and further studies with larger, more representative sets are essential to evaluate the impact of origin on chemical and sensory properties. Nevertheless, the broader context offers guidance. For large-scale commodity oils such as sunflower, rapeseed, or palm oil, where production is large-scale and quality requirements are relatively uniform, regional variation provides

little additional value which is marginal. In contrast, for olive oil or argan oil, geographical origin is highly valued, protected through designations of origin, has an impact of the chemical composition and the sensorial profile and plays a decisive role in consumer preference. By analogy, regional identity is likely to become an important differentiator for BKO, CO, and DSO as market demand develops. Establishing clear links between composition, sensory attributes, and origin would strengthen the case for geographical labeling. While larger datasets are needed, the present findings already highlighted practical potential for origin-based authenticity assessment.

A second major aspect of authenticity addressed in this dissertation was adulteration detection. The study aimed to explore the application of machine learning and data simulation to detect adulteration in CO, using refined sunflower oil as the adulterant. A novel pipeline was developed combining real measurements with Monte Carlo simulation and CTGAN for dataset generation. Among these, Monte Carlo produced more realistic datasets, leading to improved model performance. Between Random Forest and Neural Network classifiers, Random Forest was preferred due to its high accuracy (90%) and interpretability. This method detected adulteration levels as low as 1%, without the need to prepare large numbers of real mixtures. As a proof-of-concept, it highlighted a scalable, cost-effective approach for routine quality control. Given the limited number of collected pure oil samples, the same CO and SO real-world samples used for estimating the underlying distributions for MC simulation were also used as the basis for generating real-world adulterated samples. It is not expected that this has a large impact on the validation since the real-world samples were considered as random realizations of the underlying distribution. However, a positive effect on the model performance cannot be entirely excluded. Therefore, further independent evaluations are recommended to confirm the promising results reported here, providing a more robust validation. The simulation-based pipeline, implemented in Python and available open-source (https://github.com/kida4bmel/oil-adulteration), can be extended to other oils provided suitable input data are available. The methodology supports ongoing model refinement as new data are introduced, making it adaptable to real-world authenticity assessment. Future research should focus on broader adulteration scenarios, robustness across oil types, and improvements in data imputation to handle missing values.

Reference

A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., E Odezuligbo, I., & E Edu, O. (2024). Hyperparameter Tuning in Machine Learning: A Comprehensive Review. *Journal of Engineering Research and Reports*, *26*, 388–395.

Abouri, M., Mousadik, A. E., Msanda, F., Boubaker, H., Saadi, B., & Cherifi, K. (2012). An ethnobotanical survey of medicinal plants used in the Tata Province, Morocco. *International Journal of Medicinal Plant Research*, *1*, 99–123.

Ahmed, A. A. O., Kita, A., Nemś, A., Miedzianka, J., Foligni, R., Abdalla, A. M. A., & Mozzon, M. (2020). Tree-to-tree variability in fruits and kernels of a Balanites aegyptiaca (L.) Del. population grown in Sudan. *Trees*, *34*, 111–119.

Akroud, H., Sbaghi, M., Bouharroud, R., Koussa, T., Boujghagh, M., & El Bouhssini, M. (2021). Antibioisis and antixenosis resistance to Dactylopius opuntiae (Hemiptera: Dactylopiidae) in Moroccan cactus germplasm. *Phytoparasitica*, *49*, 623–631.

Al Juhaimi, F., Özcan, M. M., Adiamo, O. Q., Alsawmahi, O. N., Ghafoor, K., & Babiker, E. E. (2018). Effect of date varieties on physico-chemical properties, fatty acid composition, tocopherol contents, and phenolic compounds of some date seed and oils. *Journal of Food Processing and Preservation*, *42*, e13584.

Alkhoori, M. A., Kong, A. S.-Y., Aljaafari, M. N., Abushelaibi, A., Erin Lim, S.-H., Cheng, W.-H., Chong, C.-M., & Lai, K.-S. (2022). Biochemical Composition and Biological Activities of Date Palm (Phoenix dactylifera L.) Seeds: A Review. *Biomolecules*, *12*, 1626.

Al-Naqeb, G., Fiori, L., Ciolli, M., & Aprea, E. (2021). Prickly Pear Seed Oil Extraction, Chemical Characterization and Potential Health Benefits. *Molecules*, *26*, 5018.

Amaral, J. S. (2021). Target and Non-Target Approaches for Food Authenticity and Traceability. *Foods (Basel, Switzerland)*, *10*, 172.

Avato, P., & Tava, A. (2022). Rare fatty acids and lipids in plant oilseeds: occurrence and bioactivity. *Phytochemistry Reviews*, *21*, 401–428.

Babura, S. R., Abdullah, S. N. A., & Khaza Ai, H. (2017). Advances in Genetic Improvement for Tocotrienol Production: A Review. *Journal of nutritional science and vitaminology*, *63*, 215–221.

Ballin, N. Z., & Laursen, K. H. (2019). To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends in Food Science & Technology*, *86*, 537–543.

Barba, F. J., Putnik, P., Bursać Kovačević, D., Poojary, M. M., Roohinejad, S., Lorenzo, J. M., & Koubaa, M. (2017). Impact of conventional and non-conventional processing on prickly pear (Opuntia spp.) and their derived products: From preservation of beverages to valorization of by-products. *Trends in Food Science & Technology*, *67*, 260–270.

Bellakhdar, J. (1997). *La pharmacopée marocaine traditionnelle: Médecine arabe ancienne et savoirs populaires*.

Besbes, S., Blecker, C., Deroanne, C., Bahloul, N., Lognay, G., Drira, N.-E., & Attia, H. (2004). Date seed oil: phenolic, tocopherol and sterol profiles. *Journal of Food Lipids*, *11*, 251–265.

Bettaieb Rebey, I., Jabri-Karoui, I., Hamrouni-Sellami, I., Bourgou, S., Limam, F., & Marzouk, B. (2012). Effect of drought on the biochemical composition and antioxidant activities of cumin (Cuminum cyminum L.) seeds. *Industrial Crops and Products*, *36*, 238–245.

Bimbo, F., Roselli, L., Carlucci, D., & Gennaro, B. C. de (2020). Consumer Misuse of Country-of-Origin Label: Insights from the Italian Extra-Virgin Olive Oil Market. *Nutrients*, *12*, 2150.

Bora, J., Tongbram, T., Mahnot, N., Mahanta, C. L., & Badwaik, L. S. (2022). Chapter 14 - Tocopherol. In J. Kour, & G. A. Nayik (Eds.), *Nutraceuticals and Health Care* (pp. 259–278): Academic Press.

Bouhlali, E. d. T., Ramchoun, M., Alem, C., Ghafoor, K., Ennassir, J., & Zegzouti, Y. F. (2017). Functional composition and antioxidant activities of eight Moroccan date fruit varieties (Phoenix dactylifera L.). *Journal of the Saudi Society of Agricultural Sciences*, *16*, 257–264.

Brigelius-Flohé, R., Kelly, F. J., Salonen, J. T., Neuzil, J., Zingg, J.-M., & Azzi, A. (2002). The European perspective on vitamin E: current knowledge and future research. *The American Journal of Clinical Nutrition*, *76*, 703–716.

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics*, *14*, 72.

Cahoon, E. B., & Li-Beisson, Y. (2020). Plant unusual fatty acids: learning from the less common. *Current Opinion in Plant Biology*, *55*, 66–73.

Carrera, C. S., & Seguin, P. (2016). Factors Affecting Tocopherol Concentrations in Soybean Seeds. *Journal of agricultural and food chemistry*, *64*, 9465–9474.

Casadei, E., Valli, E., Panni, F., Donarski, J., Farrús Gubern, J., Lucci, P., Conte, L., Lacoste, F., Maquet, A., Brereton, P., Bendini, A., & Gallina Toschi, T. (2021). Emerging trends in olive oil fraud and possible countermeasures. *Food Control*, *124*, 107902.

Chapagain, B. P., Yehoshua, Y., & Wiesman, Z. (2009). Desert date (Balanites aegyptiaca) as an arid lands sustainable bioresource for biodiesel. *Bioresource Technology*, *100*, 1221–1226.

Chaudhary, N., & Khurana, P. (2009). Vitamin E biosynthesis genes in rice: Molecular characterization, expression profiling and comparative phylogenetic analysis. *Plant Science*, *177*, 479–491.

Chbani, M., El Harkaoui, S., Willenberg, I., & Matthäus, B. (2023). Review: Analytical Extraction Methods, Physicochemical Properties and Chemical Composition of Cactus (Opuntia ficus-indica) Seed Oil and Its Biological Activity. *Food Reviews International*, *39*, 4496–4512.

Chbani, M., Matthäus, B., Charrouf, Z., El Monfalouti, H., Kartah, B., Gharby, S., & Willenberg, I. (2020). Characterization of Phenolic Compounds Extracted from Cold Pressed Cactus (Opuntia ficus-indica L.) Seed Oil and the Effect of Roasting on Their Composition. *Foods*, *9*, 1098.

Cheng, W.-Y., Haque Akanda, J. M., & Nyam, K.-L. (2016). Kenaf seed oil: A potential new source of edible oil. *Trends in Food Science & Technology*, *52*, 57–65.

Codex (1999). Codex Alimentarius: Standard for named vegetable oils Codex; CXS 210-1999. Adopted in 1999, Amended in 2023. https://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/. Accessed 08.11.2024.

Cui, J., Li, K., Lv, Y., Liu, S., Cai, Z., Luo, R., Zhang, Z., & Wang, S. (2024). Development of a new hyperspectral imaging technology with autoencoder-assisted generative adversarial network for predicting the content of polyunsaturated fatty acids in red meat. *Computers and Electronics in Agriculture*, *220*, 108842.

Deng, M., Chen, H., Zhang, W., Cahoon, E. B., Zhou, Y., & Zhang, C. (2023). Genetic improvement of tocotrienol content enhances the oxidative stability of canola oil. *Frontiers in Plant Science*, *14*, 1247781.

Deng, Z., Wang, T., Zheng, Y., Zhang, W., & Yun, Y.-H. (2024). Deep learning in food authenticity: Recent advances and future trends. *Trends in Food Science & Technology*, *144*, 104344.

Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., Viant, M. R., & Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, *12*, 93.

Dou, X., N´Diaye, K., Harkaoui, S. E., Willenberg, I., Ma, F., Zhang, L., Li, P., & Matthäus, B. (2025). Authentication of Virgin Olive Oil Based on Untargeted Metabolomics and Chemical Markers. *European Journal of Lipid Science and Technology*, *127*, e202400126.

Echegaray, N., Gullón, B., Pateiro, M., Amarowicz, R., Misihairabgwi, J. M., & Lorenzo, J. M. (2023). Date Fruit and Its By-products as Promising Source of Bioactive Components: A Review. *Food Reviews International*, *39*, 1411–1432.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr, Eddine, Gharby, S., Charrouf, Z., & Matthäus, B. (2024a). Dataset: Fatty acid composition of 27 cold-pressed cactus seed oils from the Moroccan market.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr, Eddine, Gharby, S., Charrouf, Z., & Matthäus, B. (2024b). Dataset: Tocochromanol composition of 27 cold-pressed cactus seed oils from the Moroccan market.

El Harkaoui, S., El Monfalouti, H., Kartah, Badr, Eddine, Gharby, S., Charrouf, Z., & Matthäus, B. (2024c). Dataset: Triacylglycerol composition of 27 cold-pressed cactus seed oils from the Moroccan market.

El Harkaoui, S., Rohn, S., & Matthäus, B. (2023). Comparative Chemical Composition and Kinetic Parameter Determination of Four Cold-Pressed Edible Oils from the Moroccan Desert. *Lebensmittelchemie*, *77*, S3-266-S3-266.

El Kabous, K., & Ouhssine, M. (2024). Examining the methods used to extract various grades of Argan oil sold on the Moroccan market: Implications for quality and safety. *E3S Web of Conferences*, *477*, 32.

El Mekkaoui, A., Moussadek, R., Mrabet, R., Chakiri, S., Douaik, A., Ghanimi, A., & Zouahri, A. (2021). The conservation agriculture in northwest of Morocco (Merchouch area): The impact of no-till systems on physical properties of soils in semi-arid climate. *E3S Web of Conferences*, *234*, 37.

Elgadi, S., Ouhammou, A., Zine, H., Maata, N., Aitlhaj, A., El Allali, H., & El Antari, A. (2021). Discrimination of Geographical Origin of Unroasted Kernels Argan Oil (Argania spinosa (L.) Skeels) Using Tocopherols and Chemometrics. *Journal of Food Quality*, *2021*, 1–9.

Elhoumaizi, M. A., Jdaini, K., Alla, F., & Parmar, A. (2023). Variations in physicochemical and microbiological characteristics of 'Mejhoul' dates (Phoenix dactylifera L.) from Morocco and new countries of its expansion. *Journal of the Saudi Society of Agricultural Sciences*, *22*, 318–326.

Espejel, J., Fandos, C., & Flavián, C. (2008). Consumer satisfaction. *British Food Journal*, *110*, 865–881.

Esteki, M., Simal-Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E., & Vander Heyden, Y. (2018). A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, *93*, 165–182.

European Commission (2023). Annual Report Alert and Cooperation Network (2022). https://food.ec.europa.eu/safety/acn/reports-and-publications_en#agri-foodfraud.

FAO 2023. Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL. Accessed 25.03.2023.

Faouzi, H. (2012). Numéro 14. *Confins. Revue franco-brésilienne de géographie / Revista franco-brasilera de geografia*, *14*.

Farag, M. A., Otify, A., & Baky, M. H. (2023). Phoenix Dactylifera L. Date Fruit By-products Outgoing and Potential Novel Trends of Phytochemical, Nutritive and Medicinal Merits. *Food Reviews International*, *39*, 488–510.

Faubel, N., Barberá, R., & Garcia-Llatas, G. (2024). Human Oral Phase Coupled with In Vitro Dynamic Gastrointestinal Digestion for Assessment of Plant Sterol Bioaccessibility from Wholemeal Rye Bread. *Journal of agricultural and food chemistry*, *72*, 15672–15679.

Frankel, E. N. (2005). *Lipid oxidation.* (Second edition). Woodhead Publishing in Food Science, Technology and Nutrition, Volume 10. Cambridge, England: Woodhead Publishing Limited.

Gao, B., Holroyd, S. E., Moore, J. C., Laurvick, K., Gendel, S. M., & Xie, Z. (2019). Opportunities and Challenges Using Non-targeted Methods for Food Fraud Detection. *Journal of agricultural and food chemistry*, *67*, 8425–8430.

Georgouli, K., Osorio, M. T., Del Martinez Rincon, J., & Koidis, A. (2018). Data augmentation in food science: Synthesising spectroscopic data of vegetable oils for performance enhancement. *Journal of Chemometrics*, *32*, e3004.

Ghaffari, M., Gholizadeh, A., Rauf, S., & Shariati, F. (2023). Drought-stress induced changes of fatty acid composition affecting sunflower grain yield and oil quality. *Food Science & Nutrition*, *11*, 7718–7731.

Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, R., Ibáñez, M., Beltran, J., Sancho, J. V., & Hernández, F. J. (2016). Metabolomic approach for Extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-of-flight mass spectrometry. *Food Control*, *70*, 350–359.

Gracia Moisés, A., Vitoria Pascual, I., Imas González, J. J., & Ruiz Zamarreño, C. (2023). Data Augmentation Techniques for Machine Learning Applied to Optical Spectroscopy Datasets in Agrifood Applications: A Comprehensive Review. *Sensors*, *23*, 8562.

Gumus, V., El Moçayd, N., Seker, M., & Seaid, M. (2024). Future projection of droughts in Morocco and potential impact on agriculture. *Journal of Environmental Management*, *367*, 122019.

Harkat, H., Bousba, R., Benincasa, C., Atrouz, K., Gültekin-Özgüven, M., Altuntaş, Ü., Demircan, E., Zahran, H. A., & Özçelik, B. (2022). Assessment of Biochemical Composition and Antioxidant Properties of Algerian Date Palm (Phoenix dactylifera L.) Seed Oil. *Plants*, *11*, 381.

Hatzig, S. V., Nuppenau, J.-N., Snowdon, R. J., & Schießl, S. V. (2018). Drought stress has transgenerational effects on seeds and seedlings in winter oilseed rape (Brassica napus L.). *BMC Plant Biology*, *18*, 297.

Hu, Q., Zhang, J., Xing, R., Yu, N., & Chen, Y. (2022). Integration of lipidomics and metabolomics for the authentication of camellia oil by ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry coupled with chemometrics. *Food Chemistry*, *373*, 131534.

Ilić, M., Pastor, K., Romanić, R., Vujić, Đ., & Ačanski, M. (2022). A New Challenge in Food Authenticity: Application of a Novel Mathematical Model for Rapid Quantification of Vegetable Oil Blends by Gas Chromatography – Mass Spectrometry (GC-MS). *Analytical Letters*, *55*, 2752–2763.

Indelicato, S., Bongiorno, D., Pitonzo, R., Di Stefano, V., Calabrese, V., Indelicato, S., & Avellone, G. (2017). Triacylglycerols in edible oils: Determination, characterization, quantitation, chemometric approach and evaluation of adulterations. *Journal of chromatography. A*, *1515*, 1–16.

International Olive Council (IOC) (2024). *Trade Standard Applying To Olive Oils And Olive Pomace Oils: COI/T.15/NC No 3/ Rev.20/2024*. https://www.internationaloliveoil.org/. Accessed 05.05.2025.

Jung, D. M., Lee, M. J., Yoon, S. H., & Jung, M. Y. (2011). A gas chromatography-tandem quadrupole mass spectrometric analysis of policosanols in commercial vegetable oils. *Journal of food science*, *76*, C891-9.

Kachroo, A., Fu, D.-Q., Havens, W., Navarre, D., Kachroo, P., & Ghabrial, S. A. (2008). An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. *Molecular plant-microbe interactions MPMI*, *21*, 564–575.

Kamal, M., & Karoui, R. (2015). Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. *Trends in Food Science & Technology*, *46*, 27–48.

Kettani, R., Nabloussi, A., Bahri, N., Ferrahi, M., Moussadek, R., Bentaibi, A., & Tirazi, R. (Eds.) (2018). *Agriculture résiliente dans un contexte de changement climatique*.

Kirkhus, B., Lundon, A. R., Haugen, J.-E., Vogt, G., Borge, G. I. A., & Henriksen, B. I. F. (2013). Effects of environmental factors on edible oil quality of organically grown Camelina sativa. *Journal of agricultural and food chemistry*, *61*, 3179–3185.

Kodad, O., Socias i Company, R., & Alonso, J. M. (2018). Genotypic and Environmental Effects on Tocopherol Content in Almond. *Antioxidants*, *7*, 6.

Kritchevsky, D., Tepper, S. A., Chen, S. C., Meijer, G. W., & Krauss, R. M. (2000). Cholesterol vehicle in experimental atherosclerosis. 23. Effects of specific synthetic triglycerides. *Lipids*, *35*, 621–625.

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*, 79–86.

Kumar, M. S. S., Mawlong, I., Ali, K., & Tyagi, A. (2018). Regulation of phytosterol biosynthetic pathway during drought stress in rice. *Plant physiology and biochemistry PPB*, *129*, 11–20.

Lago-Olveira, S., Ouhemi, H., Idrissi, O., Moreira, M. T., & González-García, S. (2024). Promoting more sustainable agriculture in the Moroccan drylands by shifting from conventional wheat monoculture to a rotation with chickpea and lentils. *Cleaner Environmental Systems*, *12*, 100169.

Lands, B. (2008). A critique of paradoxes in current advice on dietary lipids. *Progress in Lipid Research*, *47*, 77–106.

Laribi, B., Bettaieb, I., Kouki, K., Sahli, A., Mougou, A., & Marzouk, B. (2009). Water deficit effects on caraway (Carum carvi L.) growth, essential oil and fatty acid composition. *Industrial Crops and Products*, *30*, 372–379.

Lauridsen, C. (2017). Chapter 22 - Digestion, absorption and metabolism of lipids. In C. C. Akoh (Ed.), *Food Lipids* (pp. 591–602): CRC Press.

Liu, T., Abd-Elrahman, A., Morton, J., & Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience & Remote Sensing*, *55*, 243–264.

Lucini, L., Rocchetti, G., & Trevisan, M. (2020). Extending the concept of terroir from grapes to other agricultural commodities: an overview. *Current Opinion in Food Science*, *31*, 88–95.

Lushchak, V. I., & Semchuk, N. M. (2012). Tocopherol biosynthesis: chemistry, regulation and effects of environmental factors. *Acta Physiologiae Plantarum*, *34*, 1607–1628.

Lykke, A. M., Gregersen, S. B., Padonou, E. A., Bassolé, I. H. N., & Dalsgaard, T. K. (2021). Potential of Unconventional Seed Oils and Fats from West African Trees: A Review of Fatty Acid Composition and Perspectives. *Lipids*, *56*, 357–390.

Maqsood, S., Adiamo, O., Ahmad, M., & Mudgil, P. (2020). Bioactive compounds from date fruit and seed as potential nutraceutical and functional food ingredients. *Food Chemistry*, *308*, 125522.

Mariod, A. A. (Ed.) (2019). *Wild fruits: Composition, nutritional value and products.* (1st ed.): Springer Cham.

Mattoli, L., Gianni, M., & Burico, M. (2022). Mass spectrometry-based metabolomic analysis as a tool for quality control of natural complex products. *Mass spectrometry reviews*, e21773.

Medina, S., Pereira, J. A., Silva, P., Perestrelo, R., & Câmara, J. S. (2019). Food fingerprints - A valuable tool to monitor food authenticity and safety. *Food Chemistry*, *278*, 144–162.

Mengoub, F. E., Uri, D., Abdelaaziz, A. A., & Isabelle, T. (2022). *The Russia-Ukraine war and food security in morocco*: Policy center for the new south.

Ministère de l'agriculture (2024). Filière arganier. https://www.agriculture.gov.ma/fr/filiere/arganier. Accessed 29.07.2024.

Mohamed, A. M., Wolf, W., & Spiess, W. E. L. (2002). Physical, morphological and chemical characteristics, oil recovery and fatty acid composition of Balanites aegyptiaca Del. kernels. *Plant Foods for Human Nutrition*, *57*, 179–189.

Mohamed Mousa, Y., Gerasopoulos, D., Metzidakis, I., & Kiritsakis, A. (1996). Effect of Altitude on Fruit and Oil Quality Characteristics of 'Mastoides' Olives. *Journal of the Science of Food and Agriculture*, *71*, 345–350.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of food science*, *77*, R118-26.

Moussadek, R., Ouabbou, H., El Gharras, O., Dahan, R., & El Mourid, M. (2024). *Research For Promoting Sustainable Farming Systems In Arid And Semi-Arid Areas Of Morocco, Challenges, Achievements And Future Prospects*.

Mrabet, A., Jiménez-Araujo, A., Guillén-Bejarano, R., Rodríguez-Arcos, R., & Sindic, M. (2020). Date Seeds: A Promising Source of Oil with Functional Properties. *Foods*, *9*, 787.

Murthy, H. N., Yadav, G. G., Dewir, Y. H., & Ibrahim, A. (2020). Phytochemicals and Biological Activity of Desert Date (Balanites aegyptiaca (L.) Delile). *Plants (Basel, Switzerland)*, *10*.

Nehdi, I., Omri, S., Khalil, M. I., & Al-Resayes, S. I. (2010). Characteristics and chemical composition of date palm (Phoenix canariensis) seeds and seed oil. *Industrial Crops and Products*, *32*, 360–365.

Nehdi, I. A., Sbihi, H. M., Tan, C. P., Rashid, U., & Al-Resayes, S. I. (2018). Chemical Composition of Date Palm (Phoenix dactylifera L.) Seed Oil from Six Saudi Arabian Cultivars. *Journal of food science*, *83*, 624–630.

Nesaretnam, K., Yew, W. W., & Wahid, M. B. (2007). Tocotrienols and cancer: Beyond antioxidant activity. *European Journal of Lipid Science and Technology*, *109*, 445–452.

Ngo-Duy, C.-C., Destaillats, F., Keskitalo, M., Arul, J., & Angers, P. (2009). Triacylglycerols of Apiaceae seed oils: Composition and regiodistribution of fatty acids. *European Journal of Lipid Science and Technology*, *111*, 164–169.

Normén, L., Ellegård, L., Brants, H., Dutta, P., & Andersson, H. (2007). A phytosterol database: Fatty foods consumed in Sweden and the Netherlands. *Journal of Food Composition and Analysis*, *20*, 193–201.

ODCO (2024). Statistics: Breakdown of cooperatives by sector. https://www.odco.gov.ma/statistics/?lang=en. Accessed 29.07.2024.

Osorio, M. T., Haughey, S. A., Elliott, C. T., & Koidis, A. (2014). Evaluation of methodologies to determine vegetable oil species present in oil mixtures: Proposition of an approach to meet the EU legislation demands for correct vegetable oils labelling. *Food Research International*, *60*, 66–75.

Othman, S., Mavani, N. R., Hussain, M. A., Rahman, N. A., & Mohd Ali, J. (2023). Artificial intelligence-based techniques for adulteration and defect detections in food and agricultural industry: A review. *Journal of Agriculture and Food Research*, *12*, 100590.

Ouédraogo, S., Ouédraogo, O., Thiombiano, A., & Boussim, J. I. (2023). The role of Balanites aegyptiaca (L) Delile in the livelihood and local economy in Sahelian and Sudano-Sahelian zones of Burkina Faso: basis for its conservation. *Environment, Development and Sustainability*, *25*, 1420–1440.

Ozenda, P. (1991). *Flore et vegetation du Sahara.* (3rd ed.). CNRS: Paris, France.

Porokhovinova, E. A., Matveeva, T. V., Khafizova, G. V., v. d. Bemova, Doubovskaya, A. G., Kishlyan, N. V., Podolnaya, L. P., & Gavrilova, V. A. (2022). Fatty acid composition of oil crops: genetics and genetic engineering. *Genetic Resources and Crop Evolution*, *69*, 2029–2045.

Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., & Zhang, C. (2018). Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network. *Applied Sciences*, *8*, 212.

Ramadan, M. F., Tamer, E. M. A., & Sascha Rohn (2021). *Opuntia spp.: chemistry, bioactivity and industrial applications*. Cham: Springer International Publishing.

Rocha, Werickson Fortunato de Carvalho, Prado, C. B. d., & Blonder, N. (2020). Comparison of Chemometric Problems in Food Analysis Using Non-Linear Methods. *Molecules*, *25*, 3025.

Roche, J., Bouniols, A., Mouloungui, Z., Barranco, T., & Cerny, M. (2006). Management of environmental crop conditions to produce useful sunflower oil components. *European Journal of Lipid Science and Technology*, *108*, 287–297.

Rogowska, A., & Szakiel, A. (2020). The role of sterols in plant response to abiotic stress. *Phytochemistry Reviews*, *19*, 1525–1538.

Romero, M. P., Tovar, M. J., Ramo, T., & Motilva, M. J. (2003). Effect of crop season on the composition of virgin olive oil with protected designation of origin "Les garrigues". *Journal of the American Oil Chemists' Society*, *80*, 423–430.

Roßbach, P. (2018). Neural networks vs. random forests–does it always have to be deep learning. *Frankfurt School of Finance and Management*.

Sabbahi, R., & Hock, V. (2022). Control of the prickly pear cochineal, Dactylopius opuntiae (Cockerell), in Morocco: an overview. *Journal of Plant Diseases and Protection*, *129*, 1323–1330.

Saini, M. K., Prasad, J., Raju, P. V. S., Kothari, S. L., Harish, Shukla, J. K., & Gour, V. S. (2021). Morphological Descriptors and Heritability as Markers for Oil Yield in Balanites aegyptiaca (L.) Del.: A Potential Biodiesel Xerophyte. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, *91*, 695–706.

Schaller, H. (2003). The role of sterols in plant growth and development. *Progress in Lipid Research*, *42*, 163–175.

Schmid, K. M. (2021). Lipid metabolism in plants. In N. D. Ridgway, & R. S. McLeod (Eds.), *Biochemistry of Lipids, Lipoproteins and Membranes (Seventh Edition)* (pp. 121–159): Elsevier.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environmental Science & Technology*, *48*, 2097–2098.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*, 1–48.

Silva, M. A., Albuquerque, T. G., Pereira, P., Ramalho, R., Vicente, F., Oliveira, M. B. P. P., & Costa, H. S. (2021). Opuntia ficus-indica (L.) Mill.: A Multi-Benefit Potential to Be Exploited. *Molecules (Basel, Switzerland)*, *26*, 951.

Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., Unger, T., Malitsky, S., Arendt, P., Pauwels, L., Almekias-Siegl, E., Rogachev, I., Meir, S., Cárdenas, P. D., Masri, A., Petrikov, M., Schaller, H., Schaffer, A. A., Kamble, A., Giri, A. P., Goossens, A., & Aharoni, A. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nature Plants*, *3*, 16205.

Stefanoudaki, E., Williams, M., Chartzoulakis, K., & Harwood, J. (2009). Effect of irrigation on quality attributes of olive oil. *Journal of agricultural and food chemistry*, *57*, 7048–7055.

Sudhakar, A., Chakraborty, S. K., Mahanti, N. K., & Varghese, C. (2023). Advanced techniques in edible oil authentication: A systematic review and critical analysis. *Critical reviews in food science and nutrition*, *63*, 873–901.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden, J. J., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, *3*, 211–221.

Tahir, H. E., Mariod, A. A., & Xiaobo, Z. (2022). Chapter 1 - Unconventional oils production, utilization worldwide. In A. A. Mariod (Ed.), *Multiple Biological Activities of Unconventional Seed Oils* (pp. 1–15): Academic Press.

Trela, A., & Szymańska, R. (2019). Less widespread plant oils as a good source of vitamin E. *Food Chemistry*, *296*, 160–166.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*, *7*, 142.

Wang, Y., Gu, H.-W., Yin, X.-L., Geng, T., Long, W., Fu, H., & She, Y. (2024). Deep leaning in food safety and authenticity detection: An integrative review and future prospects. *Trends in Food Science & Technology*, *146*, 104396.

Willenberg, I., & Matthäus, B. (2019). Authenticity of Edible Oils—Heading for New Methods. *European Journal of Lipid Science and Technology*, *121*, 1900021.

Willenberg, I., Parma, A., Bonte, A., & Matthäus, B. (2021). Development of Chemometric Models Based on a LC-qToF-MS Approach to Verify the Geographic Origin of Virgin Olive Oil. *Foods*, *10*, 479.

Yao, Y., Liu, W., Zhou, H., Di Zhang, Li, R., Li, C., & Wang, S. (2019). The Relations between Minor Components and Antioxidant Capacity of Five Fruits and Vegetables Seed Oils in China. *Journal of Oleo Science*, *68*, 625–635.

Yuan, Z., Zhang, L., Du Wang, Jiang, J., Harrington, P. d. B., Mao, J., Zhang, Q., & Li, P. (2020). Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis. *LWT*, *125*, 109247.

Zhang, L., Huang, X., Li, P., Na, W., Jiang, J., Mao, J., Ding, X., & Zhang, Q. (2017a). Multivariate adulteration detection for sesame oil. *Chemometrics and Intelligent Laboratory Systems*, *161*, 147–150.

Zhang, L., Li, P., Sun, X., Wang, X., Xu, B., Wang, X., Ma, F., Zhang, Q., & Ding, X. (2014). Classification and adulteration detection of vegetable oils based on fatty acid profiles. *Journal of agricultural and food chemistry*, *62*, 8745–8751.

Zhang, L., Wang, Y., Wei, Y., & An, D. (2022). Near-infrared hyperspectral imaging technology combined with deep convolutional generative adversarial network to predict oil content of single maize kernel. *Food Chemistry*, *370*, 131047.

Zhang, L., Yuan, Z., Li, P., Wang, X., Mao, J., Zhang, Q., & Hu, C. (2017b). Targeted multivariate adulteration detection based on fatty acid profiles and Monte Carlo one-class partial least squares. *Chemometrics and Intelligent Laboratory Systems*, *169*, 94–99.

Zock, P. L., Vries, J. H. de, & Katan, M. B. (1994). Impact of myristic acid versus palmitic acid on serum lipid and lipoprotein levels in healthy women and men. *Arteriosclerosis and thrombosis a journal of vascular biology*, *14*, 567–575.

Appendix

List of conference contributions

Oral Presentations

Roggenland, A., **El Harkaoui, S**., Cruz, C.O., Schneider, M., Rohn, S., Drusch, S. and Matthäus, B. Adulteration detection in cactus seed oil: integrating analytical chemistry and machine learning approaches. AI for Research in Food, Agriculture and Environment. 2 – 3 December 2024, Leipzig, Germany.

El Harkaoui, S., Chemical composition of Moroccan Toogga oil for potential cosmetic use. 3rd International Congress on Natural Products and Sustainable Development. 25 – 27 Mai 2023, Rabat, Morocco.

Matthäus, B., **El Harkaoui**, **S.,** El Monfalouti, H., Kartah, B., Charrouf, Z., Approaches for the evaluation of quality and authenticity of different seed oils from Morocco. 3rd International Congress on Natural Products and Sustainable Development. 25 – 27 May 2023, Rabat, Morocco.

El Harkaoui, S., Exploring new sources of edible oil in the Moroccan desert area. 7th Conference for Student Research at Humboldt-Universität zu Berlin, 4 and 5 October 2022, Berlin, Germany.

Poster Presentations

El Harkaoui, S., Rohn, S., El Kaourat, A., Kartah, B.E., El Monfalouti, H. and Matthäus, B. *Balanites aegyptiaca* Kernel Oil: Chemical Composition and Geographical Variability. 52. Deutsche Lebensmittelchemietage, 16 – 18 September 2024, Freising, Germany.

El Harkaoui, S., N'Diaye, K., Vosmann, K., Gharby, S., Al-Hilal, M., Charrouf, Z., Rohn, S., Drusch, S., Matthäus, B. 2023. Composition and Geographical Variations of Moroccan Date Seed Oils (*Phoenix dactylifera* L.)- A Targeted and Untargeted Metabolomic approach. 19th Euro Fed Lipid Congress and Expo, 17 – 20 September 2023, Poznan, Poland.

El Harkaoui, S., Rohn, S., El Kaourat, A., Kartah, B.E., El Monfalouti, H. and Matthäus, B. Comparative chemical composition and kinetic parameter determination of four cold-pressed edible oils from the Moroccan desert. 51. Deutsche Lebensmittelchemietage, 21 – 23 August 2023, Bonn, Germany.

El Harkaoui, S., Aissa R., Bousaid Z., Bijla L., Matthäus, B., Gharby S. Chemical composition and kinetic parameter determination of unrefined and refined argan oil under the Rancimat test condition. Congrès International sur l'Arganier, 10 – 13 Mai 2022, Agadir, Morocco.

El Harkaoui, S., Charrouf, Z., El Monfalouti, H., Kartah, B., Gharby, S. and Matthäus, B. Exploring date palm, desert date, and acacia as promising sources of edible oil in Moroccan deserts area. Tropentag, September 15 – 17, 2021, hybrid conference, Germany.

Workshops

El Harkaoui, S., Matthäus, B. Exploring the Sensorial Properties of Cactus seed oil, Faculty of Sciences, University Mohammed V in Rabat, Morocco, 27 Mai 2023.

List of additional publications

The following publications includes additional information

El Harkaoui, S., El Monfalouti, H., Kartah, B.E., Gharby, S., Charrouf, Z., Matthäus, B., 2025. Dataset: Fatty acid composition of 27 cold-pressed cactus seed oils from the Moroccan market. OpenAgrar Repository. https://doi.org/10.25826/Data20240930-105507-0.

El Harkaoui, S., El Monfalouti, H., Kartah, B.E., Gharby, S., Charrouf, Z., Matthäus, B., 2025. Dataset: Triacylglycerol composition of 27 cold-pressed cactus seed oils from the Moroccan market. OpenAgrar Repository. https://doi.org/10.25826/Data20240930-114743-0.

El Harkaoui, S., El Monfalouti, H., Kartah, B.E., Gharby, S., Charrouf, Z., Matthäus, B., 2025. Dataset: Tocochromanol composition of 27 cold-pressed cactus seed oils from the Moroccan market. OpenAgrar Repository. https://doi.org/10.25826/Data20240930-113217-0.

Chbani, M., **El Harkaoui, S**., Willenberg, I. and Matthäus, B., 2022. Analytical extraction methods, physicochemical properties and chemical composition of cactus (*Opuntia ficus-indica*) seed oil and its biological activity. Food Reviews International, 39(7), pp.4496-4512. https://doi.org/10.1080/87559129.2022.2027437.

El Harkaoui, S., Rohn, S. and Matthäus, B., 2023. Comparative Chemical Composition and Kinetic Parameter Determination of Four Cold-Pressed Edible Oils from the Moroccan Desert. Lebensmittelchemie, 77, pp. S3-266. https://doi.org/10.1002/lemi.202359234.

El Harkaoui, S., Rohn, S., El Kaourat, A., Kartah, B.E., El Monfalouti, H. and Matthäus, B., 2024. Balanites aegyptiaca Kernel Oil: Chemical Composition and Geographical variability. Lebensmittelchemie, 78, pp. S3-077. https://doi.org/10.1002/lemi.202459061.

Supplementary data manuscript I

Table S1. Additional information on sample set

Sample Code	Harvest location	Description
Mo1	*Tata* - Morocco	*Balanites* fruits were collected From *Tata* (Morocco) in 2020 and the extraction was done as described in the method part
Mo2	*Tata* - Morocco	*Balanites* fruit were collected From *Tata* (Morocco) in 2021 and the extraction was done as described in the method part
Mo3	*Tata* – Morocco	*Balanites* fruit were collected From *Tata* (Morocco) in 2022 and the extraction was done as described in the method part
Mau1	*Guidimakha* – Mauritania	Cosmetic *Balanites* kernel oil (extracted from non-roosted kernels) was purchased from a local cooperative in *Guidimakha*, in Mauritania
Mau2	*Guidimakha* – Mauritania	Alimentary *Balanites* kernel oil (extracted from kernels roasted for 10 min at 100 °C in the oven) was purchased from a local cooperative in *Guidimakha*, in Mauritania
Su1	*El Fulah* – Sudan	*Balanites* fruits were collected From *El Fulah* (Sudan) in 2021 and the extraction was done as described in the method part
Su2	*Al Fashir* – Sudan	*Balanites* fruits were collected From *Al Fashir* (Sudan) in 2021 and the extraction was done as described in the method part
Su3	*Al ʼAbbasiyah* – Sudan	*Balanites* fruits were collected From *Al ʼAbbasiyah* (Sudan) in 2021 and the extraction was done as described in the method part

Table S2. Fatty acid composition of *Balanites* kernel oil (%)

Sample	Mo1	Mo2	Mo3	Mau1	Mau2	Su1	Su2	Su3
14:0	0.1±0.0	0.1±0.0	nd	nd	nd	nd	0.1±0.0	0.1±0.0
16:0	14.5±0.5	12.8±0.1	13.9±0.4	11.5±0.3	11.3±0.2	11.1±0.1	12.8±0.1	12.5±0.1
16:1 Δ9	0.1±0.0	0.1±0.0	nd	nd	nd	nd	0.1±0.0	0.1±0.0
17:0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0
18:0	11.1±0.2	9.7±0.0	11.0±0.1	13.0±0.1	12.8±0.1	11.3±0.0	12.7±0.0	12.6±0.0
18:1Δ9	23.4±0.1	28.3±0.1	27.3±0.0	33.3±0.0	33.8±0.0	35.6±0.0	34.4±0.1	31.7±0.1
18:1Δ11	0.8±0.0	1.2±0.0	1.0±0.0	0.8±0.0	0.8±0.0	1.2±0.0	1.0±0.0	0.9±0.0
18:2Δ9t,12 t	nd	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	nd	nd	nd
18:2Δ9,12	48.5±0.3	44.4±0.1	44.6±0.1	39.9±0.2	39.8±0.1	38.6±0.0	38.0±0.1	41.0±0.0
18:3Δ9,12,15	0.7±0.0	1.7±0.0	0.8±0.0	nd	nd	0.9±0.0	0.1±0.0	0.1±0.0
20:0	0.4±0.0	0.4±0.0	0.4±0.0	0.5±0.0	0.5±0.0	0.4±0.0	0.4±0.0	0.4±0.0
20:1Δ11	0.2±0.0	0.3±0.0	0.2±0.0	0.1±0.0	0.1±0.0	0.2±0.0	0.1±0.0	0.1±0.0
22:0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0
24:0	0.1±0.0	0.5±0.0	0.2±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0	0.1±0.0

Results are expressed as mean ± SD (n = 3), nd: not detected

Table S3. Triacylglycerol composition of *Balanites* kernel oil (%)

Sample	Mo1	Mo2	Mo3	Mau1	Mau2	Su1	Su2	Su3
PaOlPa	2.4±0.0	2.3±0.0	2.8±0.0	2.4±0.0	2.6±0.0	2.6±0.0	3.2±0.0	2.8±0.0
PaLiPa	7.0±0.0	6.2±0.0	6.2±0.0	3.1±0.0	3.1±0.0	3.6±0.0	4.3±0.0	4.4±0.0
PaOlSt	2.7±0.0	2.5±0.0	3.1±0.0	3.3±0.0	4.1±0.0	4.0±0.0	4.5±0.1	4.4±0.0
PaOlOl	3.8±0.0	4.6±0.0	5.3±0.0	8.2±0.0	8.3±0.0	7.6±0.0	8.7±0.1	7.2±0.0
PaLiSt	8.8±0.0	7.7±0.0	7.9±0.0	4.8±0.0	5.3±0.0	5.7±0.0	6.8±0.1	6.9±0.0
PaLiOl	14.6±0.0	13.9±0.0	15.4±0.0	14.1±0.0	13.7±0.0	13.7±0.1	14.7±0.1	14.8±0.1
PaLiLi	16.5±0.0	14.7±0.0	13.9±0.0	9.5±0.0	9.1±0.0	9.0±0.1	8.6±0.1	9.9±0.1
StOlSt	0.8±0.0	0.7±0.0	0.9±0.0	1.3±0.0	1.8±0.0	1.8±0.1	1.8±0.1	1.8±0.1
StOlOl	1.9±0.0	2.1±0.0	2.6±0.0	6.3±0.0	6.2±0.0	5.2±0.1	5.4±0.1	4.9±0.1
OlOlOl	3.3±0.0	6.8±0.0	4.5±0.0	6.3±0.0	6.5±0.0	7.5±0.0	6.1±0.1	4.8±0.1
StLiOl	8.4±0.0	8.0±0.0	9±0.0	11.1±0.0	10.7±0.0	9.4±0.1	9.9±0.1	9.9±0.1
OlLiOl	7.2±0.0	9.1±0.0	8.8±0.0	11.4±0.0	11.2±0.0	11.9±0.0	10.8±0.0	10.3±0.0
LiLiOl	13.2±0.0	13.1±0.0	12.8±0.0	12.2±0.0	11.7±0.0	12.5±0.0	11.0±0.0	12.2±0.0
LiLiLi	9.2±0.0	8.2±0.0	6.9±0.0	6.0±0.0	5.8±0.0	5.7±0.0	4.2±0.0	5.7±0.0

Results are expressed as mean ± SD (n = 3)

Pa: palmitic acid, St: searic acid, Ol: oleic acid, Li: linoleic acid

Table S4. Tocochromanol composition of *Balanites* kernel oil (mg/kg of oil)

	Mo1	Mo2	Mo3	Mau1	Mau2	Su1	Su2	Su3
α-tocopherol	607 ± 20	574 ± 2	551 ± 7	445 ± 8	426 ± 5	365 ± 9	324 ± 7	404 ± 17
β-tocopherol	17±1	13±0	10±0	7±0	7±0	1±0 <LOQ	1±0 <LOQ	2±0
γ-tocopherol	175±4	212±1	226±2	120±0	124±3	183±1	215±1	197±4
β-tocotrienol	nd	3±1	4±1	nd	nd	nd	nd	nd
plastochromanol-8	5 ± 0	12 ± 0	7 ± 0	2 ± 0	7 ± 1	6 ± 0	3 ± 1	4 ± 0
γ-tocotrienol	nd	3 ± 0	2 ± 0	1±0 <LOQ	4 ± 0	1±0 <LOQ	3 ± 0	2 ± 0
δ-tocopherol	14±1	12±1	13±0	11±1	12±0	3±0	5±0	5±1
Sum	819 ± 26	828 ± 5	812 ± 9	585 ± 8	580 ± 3	559 ± 10	552 ± 6	614 ± 22

Results are expressed as mean ± SD (n = 3), nd: not detected, LOQ: limit of quantification

Table S5. Phytosterol composition of *Balanites* kernel oil (mg/kg of oil)

Serols	Mo1	Mo2	Mo3	Mau1	Mau2	Su1	Su2	Su3
Cholesterol	82.6 ± 2.5	112.4 ± 22.1	80.3 ± 1.8	70.2 ± 0.2	77.9 ± 0.9	101.8 ± 1.6	110.9 ± 0.2	102.0 ± 1.0
Brassicasterol	51.9 ± 5.6	131.1 ± 0.6	61.9 ± 1.3	nd	nd	72.0 ± 4.1	nd	nd
24-Methylenecholesterol	13.0 ± 0.2	30.8 ± 0.8	14.0 ± 0.6	1.9 ± 0.1	nd	19.2 ± 0.2	2.8 ± 0.2	2.8 ± 0.0
Campesterol	194.9 ± 2.7	452 ± 0.9	204.6 ± 2.6	21.0 ± 0.2	28.1 ± 2.2	260.8 ± 1.9	28.6 ± 0.6	28.5 ± 0.3
Stigmasterol	57.3 ± 1.5	53.6 ± 4.7	53.6 ± 1.9	32.5 ± 2.2	49.4 ± 0.7	50.9 ± 1	58.0 ± 0.8	55.9 ± 0.6
Δ7-Campesterol	7.4 ± 0.6	5.0 ± 0.1	5.4 ± 0.3	4.8 ± 0.1	7.3 ± 0.6	4.2 ± 0.9	4.2 ± 0.3	3.0 ± 0.1
Δ5,23-Stigmastadienol	13.6 ± 1.6	12.1 ± 0.8	9.3 ± 0.3	7.9 ± 0.5	9.9 ± 1	9.1 ± 1.2	5.0 ± 0.2	5.7 ± 0.7
β-Sitosterol	1024.6 ± 16.6	1295.3 ± 12.9	938.1 ± 11.7	723.1 ± 5.4	819.3 ± 21.7	888.5 ± 5.2	569.9 ± 6.8	603.3 ± 3.2
Sitostanol	9.3 ± 0.7	7.6 ± 0.4	5.9 ± 0.5	4.9 ± 0.6	7.5 ± 1.2	6.4 ± 0.5	6.2 ± 0.2	6.2 ± 0.9
Δ5-Avenasterol	64.0 ± 0.9	100.2 ± 2.8	54.6 ± 8.1	116.7 ± 2.9	85.4 ± 7	105.4 ± 7.2	80.0 ± 2.1	76.5 ± 1.7
Δ5,24-Stigmastadienol	6.5 ± 0.7	7.3 ± 0.6	3.9 ± 0.9	4.8 ± 0.4	4.1 ± 1	4.7 ± 0.2	3.4 ± 0.5	3.4 ± 0.3
Δ7-Stigmastenol	5.1 ± 0.1	3.6 ± 0.7	2.4 ± 0.2	2.8 ± 0.4	20.1 ± 1.1	1.8 ± 0.4	1.3 ± 0.3	nd
Δ7-Avenasterol	5.2 ± 0.7	6.5 ± 0.1	1.6 ± 0.4	7.3 ± 0.3	2.0 ± 0.4	1.1 ± 0.2	1.0 ± 0.2	nd
Sum	1535.5 ± 27.4	2217.7 ± 6.2	1435.5 ± 19.9	997.8 ± 10.3	1111.0 ± 32.6	1526.0 ± 19.8	871.3 ± 10.4	887.3 ± 7.2

Results are expressed as mean ± SD (n = 3), nd: not detected

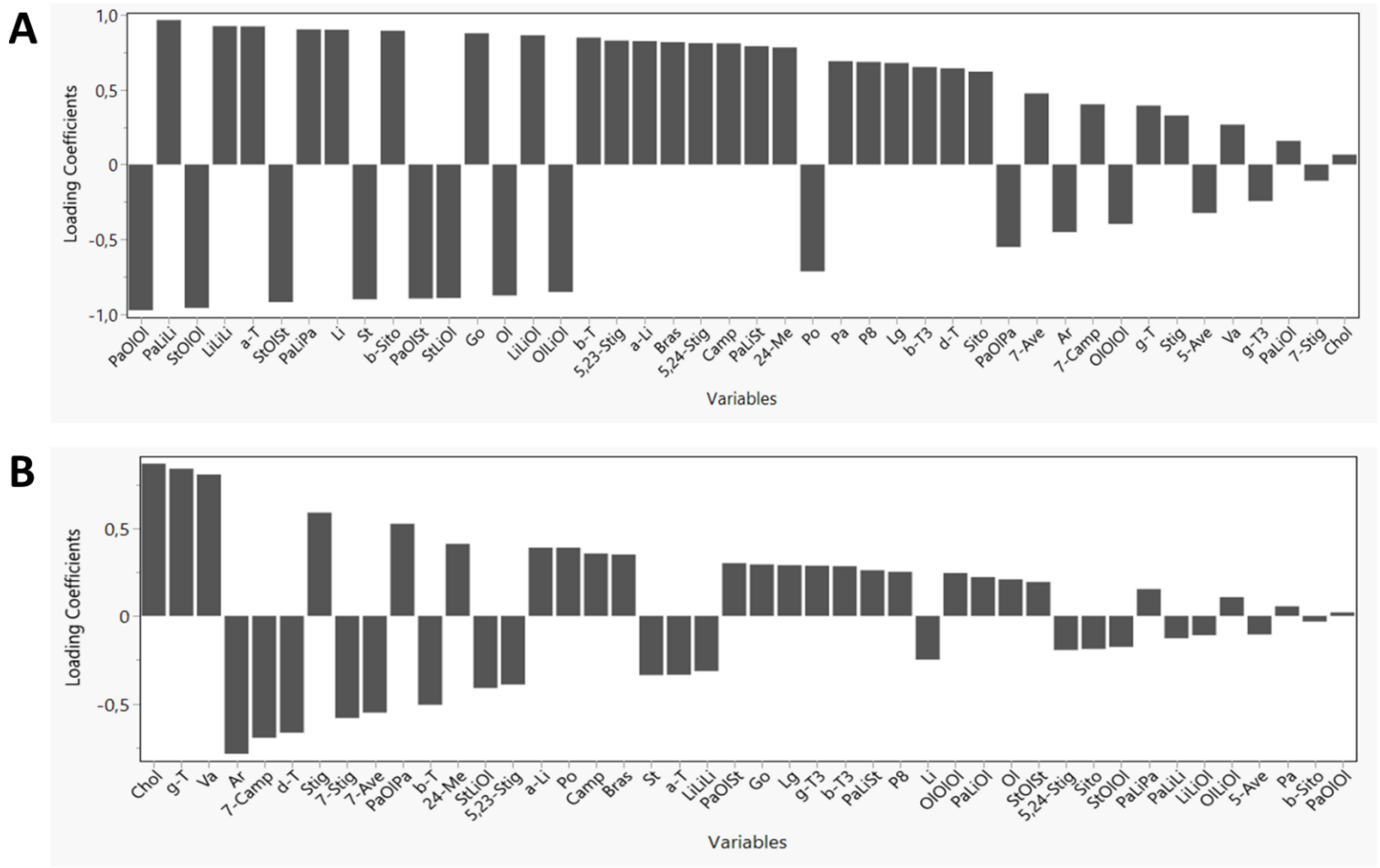


Figure S1. Loadings of the two principal component analysis, A: for PC1 and B: for PC2. Ol: oleic acid, Va: Vaccenic acid, Li: Linoleic acid, a-Li: α-linolenic acid, Ar: arachidic acid, Go: gondoic acid, Lg: lignoceric acid, a-T: α-tocopherol, b-T: β-tocopherol, g-T: γ tocopherol, b-T3: β-tocotrienol, P8: Plastochromanol-8, g-T3: γ-tocotrienol, d-T: δ-tocopherol, Chol: Cholesterol, Bras: brassicasterol, 24-Me: 24-methylenecholesterol, Camp: campesterol, Stig: stigmasterol, 7-Camp: Δ7-campesterol, 5,23-Stig: Δ5,23-stigmastadienol, b-Sito: β-sitosterol, Sito: sitostanol, 5-Ave: Δ5-avenasterol, 5,24-Stig: Δ5,24-stigmastadienol, 7-Stig: Δ7-stigmastenol, 7-Ave: Δ7-avenasterol.

Supplementary data manuscript II

Table S1: Effect of seeds geographical origin on acidity (FFA), peroxide value (PV) and specific extinctions (232 and 270 nm).

	Hoceima	Bejaâd	Rhamna	Ait Baha	Tiznit	Sidi Ifni
FFA (as oleic acid g/100 g)	0.17± 0.01^{a}	0.31± 0.05^{b}	0.4± 0.03^{c}	0.61± 0.09^{d}	0.33± 0.01^{bc}	1.23 ± 0.03^{e}
PV (mEq.O_2/kg)	4.94± 0.06^{a}	3.51± 0.03^{b}	8.65± 0.21^{c}	8.75± 0.34^{c}	9.97± 0.05^{d}	4.63± 0.03^{e}
K_{232}	2.88 ± 0.07^{a}	3.21± 0.07^{bc}	3.32± 0.12^{c}	4.59± 0.05^{d}	3.13± 0.06^{b}	3.31 ± 0.1^{c}
K_{270}	0.42± 0.01^{a}	0.49± 0.02^{bc}	0.51± 0.01^{c}	0.68± 0.02^{d}	0.48± 0.01^{b}	0.55 ± 0.01^{e}

Mean values ± SD of determination for triplicate samples. Means followed by similar letters superscript in the same line are not significantly different according to the Tukey's test ($p<0.05$).

Table S2: Pearson's correlation matrix coefficient between the variables: FFA, PV, K_{232}, K_{270}, Total sterol, Total tocopherol, SFA, UFA, Polyphenol, and DPPH ($1/IC_{50}$ value) of the different samples of cactus oils.

Variables	FFA	PV	K232	K270	Total sterols	Total tocopherols	SFA	UFA	Polyphenols	DPPH ($1/IC_{50}$ value)
FFA	**1**	-0.162	0.288	0.499	-0.239	-0.708	-0.109	0.276	0.712	0.509
PV		**1**	0.386	0.345	-0.201	0.270	**-0.952**	**0.816**	0.505	0.298
K_{232}			**1**	**0.969**	**-0.936**	-0.676	-0.419	0.676	0.537	0.084
K_{270}				**1**	**-0.875**	-0.779	-0.430	0.714	0.696	0.213
Total sterols					**1**	0.625	0.260	-0.414	-0.294	0.177
Total tocopherols						**1**	-0.147	-0.289	-0.527	-0.215
SFA							**1**	**-0.834**	-0.649	-0.390
UFA								**1**	**0.851**	0.604
Polyphenols									**1**	0.722
DPPH ($1/IC_{50}$ value)										**1**

The values in bold are different from 0 at a significance level alpha = 0.05.

Table S3: *p*-values of the correlation matrix coefficient between all variables.

Variables	FFA	PV	K_{232}	K_{270}	Total sterols	Total tocopherols	SFA	UFA	Polyphenols	DPPH ($1/IC_{50}$ value)
FFA	**0**	0.60	0.580	0.314	0.649	0.115	0.838	0.597	0.112	0.303
PV		**0**	0.450	0.503	0.702	0.605	**0.003**	**0.048**	0.307	0.566
K_{232}			**0**	**0.001**	**0.006**	0.141	0.408	0.140	0.271	0.875
K_{270}				**0**	**0.022**	0.068	0.394	0.111	0.125	0.686
Total sterols					**0**	0.184	0.619	0.415	0.571	0.737
tocopherols						**0**	0.781	0.579	0.282	0.682
SFA							**0**	**0.039**	0.163	0.445
UFA								**0**	**0.032**	0.205
Polyphenols									**0**	0.105
DPPH($1/IC_{50}$ values)										**0**

Supplementary data manuscript III

Table S1. Origin, name, name abbreviation and oil yield of the date seeds studied.

Origin of the seeds	Sample name	Name abbreviation	Oil yield (g/100g)
Allougoum - Morocco	Sayer 1	Sy1Ag	10.2 ± 0.1
	Sayer 2	Sy2Ag	10.1 ± 0.1
	Bousthammi (black)	BtbAg	10.3 ± 0.0
	Bousthammi (withe)	BtwAg	10.3 ± 0.1
	Iklan	IkAg	9.3 ± 0.0
	Lmtrwah	LtAg	8.0 ± 0.1
	Boufeggous	BfAg	8.1 ± 0.1
	Amchaw	AmAg	9.5 ± 0.1
	Mejhoul	MeAg	9.8 ± 0.1
Errachidia - Morocco	Mejhoul	MeE	8.9 ± 0.1
	Tarzawa	TzEr	8.9 ± 0.1
	Racetmar	RcEr	8.6 ± 0.1
	Boufeggous 1	Bf1Er	8.7 ± 0.1
	Boufeggous 2	Bf2Er	9.2 ± 0.0
	Bouhmroune	BhEr	9.0 ± 0.0
	Khalt 1	Kh1Er	9.1 ± 0.1
	Khalt 2	Kh2Er	10.6 ± 0.5
Alnif - Morocco	Mejhoul 1	Me1Al	8.8 ± 0.2
	Boufeggous 1	Bf1Al	8.8 ± 0.0
	Boutshammi	BtAl	9.6 ± 0.0
	Mejhoul 2	Me2Al	9.3 ± 0.1
	Boufeggous 2	Bf2Al	9.3 ± 0.3
	Tarzawa (Long)	TzLAl	10.0 ± 0.1
	Tarzawa (Small)	TzSAl	8.7 ± 0.1
	Tahmout	ThAl	10.0 ± 0.0
	Jihel	JhAl	9.2 ± 0.0
Algeria	Deglet Nour	DgnM*	8.0 ± 0.0
Kuwait	Berhi	BenM*	10.2 ± 0.1
Saudi Aribia	Ikhlas	IkhnM*	7.9 ± 0.0
Palestine	Mejhoul	MenM*	9.0 ± 0.0

*Samples from Algeria, Kuwait, Saudi Arabia, and Palestine were combined in one grouped named non-Morocco (nM). The oil yield results are expressed as mean ± SD (n = 3)

Table S2. Calibration curves used for the quantification of the tocochromanols

Compound	Calibration curve	R2
α-tocotrienol	y = 159.79x-2.87	0.9999
β-tocotrienol	y = 258.96x-14.43	1.0000
γ-tocotrienol	y = 268.88x-5.94	1.0000
δ-tocotrienol	y = 347.03x-15.46	1.0000
α-tocopherol	y = 159.79x-2.87	0.9999

Table S3. fatty acid composition of DSO samples (%)

	8:0 %	10:0 %	12:0 %	14:0 %	16:0%	16:1Δ7 %	16:1 Δ9%	17:0 %	18:0 %	18:1Δ9 %
Sy1Ag	0.3 ± 0	0.4 ± 0	21.6 ± 0.4	12.6 ± 0.2	10.2 ± 0	nd	0.1 ± 0	0.1 ± 0	4.4 ± 0.1	41.5 ± 0.4
Sy2Ag	0.2 ± 0	0.3 ± 0	14.8 ± 0.1	9.6 ± 0	10.5 ± 0	nd	0.1 ± 0	0.1 ± 0	4.6 ± 0	49.9 ± 0.1
BtbAg	0.3 ± 0	0.3 ± 0	14.9 ± 0.5	9.2 ± 0.2	10.6 ± 0	0.1 ± 0	0.1 ± 0	0.1 ± 0	4.1 ± 0	50.2 ± 0.5
BtwAg	0.3 ± 0	0.3 ± 0	16.2 ± 0.8	9.8 ± 0.1	10.3 ± 0.1	nd	0.1 ± 0	0.1 ± 0	4.2 ± 0.1	48.4 ± 0.6
IkAg	0.3 ± 0	0.4 ± 0	18.1 ± 0.2	10.1 ± 0	9.6 ± 0.1	0.1 ± 0	nd	0.1 ± 0	3.4 ± 0	47.6 ± 0.1
LtAg	0.3 ± 0	0.4 ± 0	17.3 ± 1.1	9.6 ± 0.4	9.6 ± 0	nd	0.1 ± 0	0.1 ± 0	3.0 ± 0.1	48.4 ± 1.1
BfAg	0.2 ± 0	0.2 ± 0	14.0 ± 0.3	10.2 ± 0.2	11.2 ± 0	nd	0.1 ± 0	0.1 ± 0	4.1 ± 0	48.0 ± 0.4
AmAg	0.3 ± 0	0.3 ± 0	14.8 ± 1.0	9.6 ± 0.2	11.2 ± 0.2	nd	0.1 ± 0	0.1 ± 0	3.9 ± 0.1	49.1 ± 0.8
MeAg	0.3 ± 0	0.4 ± 0	19.0 ± 0.4	10.5 ± 0	10.0 ± 0.1	nd	0.1 ± 0	0.1 ± 0	4.0 ± 0	46.2 ± 0.2
MeE	0.3 ± 0	0.3 ± 0	15.4 ± 0.2	9.2 ± 0.1	9.8 ± 0	nd	0.1 ± 0	0.1 ± 0	4.1 ± 0	49.0 ± 0.2
TzEr	0.3 ± 0	0.3 ± 0	15.6 ± 1.1	9.0 ± 0.4	10.1 ± 0	0.1 ± 0	0.1 ± 0	0.1 ± 0	4.8 ± 0.1	49.5 ± 1.2
RcEr	0.3 ± 0	0.4 ± 0	16.8 ± 0.6	9.5 ± 0.2	10.4 ± 0.1	nd	0.1 ± 0	0.1 ± 0	3.2 ± 0.1	48.6 ± 0.6
Bf1Er	0.3 ± 0	0.3 ± 0	15.6 ± 0.2	9.6 ± 0	10.3 ± 0	nd	0.1 ± 0	0.1 ± 0	4.0 ± 0	48.6 ± 0.2
Bf2Er	0.3 ± 0	0.3 ± 0	17.0 ± 0.3	10.7 ± 0.1	10.7 ± 0	0.1 ± 0	nd	0.1 ± 0	4.4 ± 0	46.1 ± 0.4
BhEr	0.3 ± 0	0.3 ± 0	17.1 ± 0.3	9.5 ± 0.1	10.4 ± 0	0.1 ± 0	nd	0.1 ± 0	3.4 ± 0	47.2 ± 0.3
Kh1Er	0.3 ± 0	0.3 ± 0	16.6 ± 0.3	10.6 ± 0.1	11.3 ± 0.1	0.1 ± 0	nd	0.1 ± 0	4.3 ± 0	46.0 ± 0.2
Kh2Er	0.3 ± 0	0.4 ± 0	20.9 ± 1.2	11.0 ± 0.1	10.9 ± 0.1	0.1 ± 0	nd	0.1 ± 0	3.6 ± 0.1	41.9 ± 0.7
Me1Al	0.3 ± 0	0.3 ± 0	15.7 ± 1.3	9.7 ± 0.2	10.8 ± 0.2	nd	0.1 ± 0	0.1 ± 0	3.9 ± 0.1	49.7 ± 1.1
Bf1Al	0.4 ± 0	0.4 ± 0	20.9 ± 0.9	10.9 ± 0.6	9.5 ± 0.2	nd	0.1 ± 0	0.1 ± 0	3.3 ± 0.1	44.3 ± 1.1
BtAl	0.3 ± 0	0.3 ± 0	17.9 ± 0.7	11.1 ± 0.5	11.6 ± 0.3	nd	0.1 ± 0	0.1 ± 0	3.9 ± 0.1	45.0 ± 1.2
Me2Al	0.3 ± 0	0.3 ± 0	16.0 ± 0.4	9.6 ± 0.1	10.2 ± 0.1	0.1 ± 0	nd	0.1 ± 0	4.1 ± 0	47.4 ± 0.4
Bf2Al	0.3 ± 0	0.3 ± 0	16.4 ± 0.5	10.4 ± 0.2	11.6 ± 0.1	0.1 ± 0	nd	0.1 ± 0	4.4 ± 0.1	46.0 ± 0.5
TzLAl	0.3 ± 0	0.3 ± 0	17.9 ± 0.6	10.7 ± 0.2	10.5 ± 0	0.1 ± 0	nd	0.1 ± 0	4.1 ± 0.1	45.5 ± 0.6
TzSAl	0.2 ± 0	0.3 ± 0	15.0 ± 0.1	9.6 ± 0.1	11.2 ± 0.1	0.1 ± 0	nd	0.1 ± 0	4.4 ± 0	48.4 ± 0.2
ThAl	0.2 ± 0	0.3 ± 0	14.7 ± 0.3	9.7 ± 0.2	11.2 ± 0.1	0.1 ± 0	nd	0.1 ± 0	4.4 ± 0	49.5 ± 0.3
JhAl	0.3 ± 0	0.3 ± 0	15.5 ± 0.4	9.9 ± 0.1	11.3 ± 0	0.1 ± 0	nd	0.1 ± 0	4.0 ± 0	47.2 ± 0.4
DgnM	0.4 ± 0	0.5 ± 0	22.5 ± 0.2	9.8 ± 0.1	9.0 ± 0	0.1 ± 0	nd	0.1 ± 0	3.1 ± 0	44.6 ± 0.2
BenM	0.4 ± 0	0.4 ± 0	20.6 ± 2.0	10.7 ± 0.8	9.7 ± 0.2	nd	0.1 ± 0	0.1 ± 0	2.4 ± 0.2	45.5 ± 2.4
IkhnM	0.5 ± 0.1	0.5 ± 0.1	24.2 ± 2.2	12.3 ± 0.5	9.6 ± 0.1	nd	0.1 ± 0	0.1 ± 0	2.9 ± 0.2	40.8 ± 2.1
MenM	0.3 ± 0	0.3 ± 0	16.7 ± 0.6	9.9 ± 0.2	9.4 ± 0	0.1 ± 0	0.1 ± 0	0.1 ± 0	4.0 ± 0.1	46.6 ± 0.5

Continue Table S3. fatty acid composition of DSO samples (%)

	18:1Δ11 %	18:2Δ9,12 %	18:3Δ9,12,15 %	20:0 %	20:1Δ11 %	22:0 %	22:2 Δ13, 16 %	23:0 %	24:0 %
Sy1Ag	0.2 ± 0	7.2 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.2 ± 0	nd	nd	0.1 ± 0
Sy2Ag	0.3 ± 0	7.9 ± 0	nd	0.6 ± 0	0.4 ± 0	0.3 ± 0	0.1 ± 0	nd	0.2 ± 0
BtbAg	0.3 ± 0	8.2 ± 0.1	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.4 ± 0	0.1 ± 0	nd	0.2 ± 0
BtwAg	0.3 ± 0	8.4 ± 0.1	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.4 ± 0	0.1 ± 0	nd	0.2 ± 0
IkAg	0.2 ± 0	8.6 ± 0	nd	0.5 ± 0	0.5 ± 0	0.3 ± 0	nd	nd	0.2 ± 0
LtAg	0.3 ± 0	9.8 ± 0.2	0.1 ± 0	0.3 ± 0	0.4 ± 0	0.2 ± 0	nd	nd	0.1 ± 0
BfAg	0.3 ± 0	9.9 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.3 ± 0	0.1 ± 0	nd	0.2 ± 0
AmAg	0.3 ± 0	8.6 ± 0.1	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.4 ± 0	0.1 ± 0	nd	0.2 ± 0
MeAg	0.2 ± 0	7.7 ± 0	0.1 ± 0	0.5 ± 0	0.4 ± 0	0.3 ± 0	nd	nd	0.2 ± 0
MeE	0.3 ± 0	9.9 ± 0.1	0.1 ± 0	0.5 ± 0	0.4 ± 0	0.3 ± 0	0.1 ± 0	nd	0.2 ± 0
TzEr	0.6 ± 0	7.7 ± 0.1	0.1 ± 0	0.7 ± 0	0.4 ± 0	0.4 ± 0	0.1 ± 0	nd	0.2 ± 0
RcEr	0.5 ± 0	8.1 ± 0.1	0.1 ± 0	0.5 ± 0	0.5 ± 0	0.4 ± 0	0.1 ± 0	nd	0.2 ± 0
Bf1Er	0.3 ± 0	9.2 ± 0	0.1 ± 0	0.5 ± 0	0.4 ± 0	0.3 ± 0	0.1 ± 0	nd	0.2 ± 0
Bf2Er	0.6 ± 0	8.2 ± 0.1	0.1 ± 0	0.6 ± 0	0.3 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
BhEr	0.7 ± 0	9.4 ± 0	0.1 ± 0	0.5 ± 0	0.4 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
Kh1Er	0.6 ± 0	8.1 ± 0.1	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
Kh2Er	0.5 ± 0	8.8 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.3 ± 0	nd	nd	0.1 ± 0
Me1Al	0.4 ± 0	7.1 ± 0.2	nd	0.6 ± 0	0.4 ± 0	0.4 ± 0	nd	nd	0.2 ± 0
Bf1Al	0.4 ± 0	8.4 ± 0.2	nd	0.4 ± 0	0.3 ± 0	0.2 ± 0	nd	nd	0.1 ± 0
BtAl	0.5 ± 0	7.5 ± 0.2	0.1 ± 0	0.6 ± 0	0.3 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
Me2Al	0.6 ± 0	9.8 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.3 ± 0	nd	nd	0.2 ± 0
Bf2Al	0.5 ± 0	8.2 ± 0.1	0.1 ± 0	0.6 ± 0	0.3 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
TzLAl	0.5 ± 0	8.5 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.3 ± 0	nd	nd	0.2 ± 0
TzSAl	0.6 ± 0	8.1 ± 0	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.4 ± 0	nd	0.1 ± 0	0.2 ± 0
ThAl	0.5 ± 0	7.5 ± 0.1	nd	0.6 ± 0	0.4 ± 0	0.4 ± 0	nd	0.1 ± 0	0.2 ± 0
JhAl	0.6 ± 0	8.9 ± 0.1	0.1 ± 0	0.6 ± 0	0.4 ± 0	0.4 ± 0	nd	0.1 ± 0	0.2 ± 0
DgnM	0.5 ± 0	7.6 ± 0	nd	0.4 ± 0	0.5 ± 0	0.3 ± 0	nd	0.1 ± 0	0.2 ± 0
BenM	0.6 ± 0.1	7.7 ± 0.4	nd	0.5 ± 0.1	0.6 ± 0.1	0.4 ± 0.1	nd	nd	0.2 ± 0
IkhnM	0.4 ± 0.2	7.1 ± 0.3	nd	0.4 ± 0.1	0.3 ± 0	0.3 ± 0.1	nd	nd	0.2 ± 0
MenM	0.5 ± 0	10.7 ± 0.1	0.1 ± 0	0.5 ± 0	0.3 ± 0	0.3 ± 0	nd	nd	0.2 ± 0

Results are expressed as mean ± SD (n = 3), nd: not detected

Table S4. triacylglycerol composition of DSO samples (%)

	CaCaLa	CaLaLa/CyLaMy	LaLaLa	LaLaMy	CyOlLa	LaMyMy	LaMyPa	LaOlLa	LaLiLa	LaPaPa	LaOlMy	LaLiMy
Sy1Ag	0.2 ± 0	0.3 ± 0	4.8 ± 0	4.2 ± 0	0.8 ± 0	3.1 ± 0	1.7 ± 0	14.5 ± 0.1	2.6 ± 0	0.8 ± 0	15.2 ± 0	2.9 ± 0
Sy2Ag	nd	0.1 ± 0	2.0 ± 0	2.6 ± 0	0.5 ± 0	2.5 ± 0	1.4 ± 0	9.5 ± 0	1.9 ± 0	0.7 ± 0	10.9 ± 0.1	2.4 ± 0
BtbAg	0.1 ± 0	0.1 ± 0	1.9 ± 0	2.1 ± 0	0.5 ± 0	2.0 ± 0	1.0 ± 0	10.4 ± 0	2.0 ± 0	0.5 ± 0	11.6 ± 0	2.4 ± 0
BtwAg	0.1 ± 0	0.1 ± 0	2.0 ± 0	2.1 ± 0	0.6 ± 0	2.0 ± 0	1.0 ± 0	11.9 ± 0.1	2.3 ± 0	0.5 ± 0	12.6 ± 0.1	2.7 ± 0
IkAg	0.2 ± 0	0.2 ± 0	3.4 ± 0	2.9 ± 0	0.5 ± 0	2.4 ± 0	1.1 ± 0	11.0 ± 0	2.3 ± 0	0.5 ± 0	12.1 ± 0	2.8 ± 0
LtAg	0.1 ± 0	0.2 ± 0	2.3 ± 0	2.3 ± 0	0.6 ± 0	2.0 ± 0	0.9 ± 0	12.9 ± 0	3.0 ± 0	0.4 ± 0	13.2 ± 0	3.3 ± 0
BfAg	nd	nd	1.4 ± 0	1.6 ± 0	0.4 ± 0	1.8 ± 0.1	0.8 ± 0	9.1 ± 0.1	2.1 ± 0	0.5 ± 0	11.9 ± 0	2.9 ± 0
AmAg	0.1 ± 0	0.1 ± 0	2.1 ± 0	2.5 ± 0	0.5 ± 0	2.6 ± 0	1.3 ± 0	10.1 ± 0.1	2.2 ± 0	0.7 ± 0	11.2 ± 0.1	2.6 ± 0
MeAg	0.2 ± 0	0.2 ± 0	3.5 ± 0	2.9 ± 0	0.7 ± 0	2.2 ± 0	1.1 ± 0	12.9 ± 0.1	2.3 ± 0	0.5 ± 0	13.5 ± 0	2.7 ± 0
MeE	nd	0.1 ± 0	1.4 ± 0	1.5 ± 0	0.6 ± 0	1.7 ± 0	0.7 ± 0	11.5 ± 0.1	2.6 ± 0	0.4 ± 0	12.4 ± 0.1	3.0 ± 0.1
TzEr	0.1 ± 0	0.1 ± 0	2.1 ± 0	2.2 ± 0	0.7 ± 0	1.9 ± 0.1	1.0 ± 0	11.9 ± 0.1	2.2 ± 0	0.5 ± 0	12.1 ± 0.1	2.5 ± 0.1
RcEr	0.1 ± 0	0.1 ± 0	1.8 ± 0	2.0 ± 0	0.8 ± 0	1.9 ± 0	0.9 ± 0	12.8 ± 0.1	2.4 ± 0	0.5 ± 0	12.7 ± 0.1	2.8 ± 0
Bf1Er	nd	0.1 ± 0	1.5 ± 0	1.6 ± 0	0.6 ± 0	1.8 ± 0.1	0.8 ± 0	11.3 ± 0	2.4 ± 0	0.4 ± 0	12.7 ± 0	2.9 ± 0
Bf2Er	0.1 ± 0	0.1 ± 0	1.7 ± 0	1.9 ± 0	0.5 ± 0	1.8 ± 0.1	0.9 ± 0	10.3 ± 0.1	2.0 ± 0	0.5 ± 0	12.4 ± 0.1	2.5 ± 0
BhEr	0.1 ± 0	0.1 ± 0	2.4 ± 0	1.9 ± 0	0.5 ± 0	1.5 ± 0.1	0.7 ± 0	9.8 ± 0	2.2 ± 0	0.4 ± 0	10.7 ± 0	2.9 ± 0
Kh1Er	0.1 ± 0	0.1 ± 0	1.5 ± 0	1.8 ± 0	0.5 ± 0	1.8 ± 0	0.9 ± 0.1	10.0 ± 0.1	2.0 ± 0.1	0.5 ± 0	11.9 ± 0.1	2.5 ± 0
Kh2Er	0.1 ± 0	0.1 ± 0	2.1 ± 0	2.1 ± 0	0.7 ± 0	1.6 ± 0.1	0.8 ± 0	13.1 ± 0	2.9 ± 0	0.5 ± 0.1	13.3 ± 0.1	3.1 ± 0.1
Me1Al	nd	0.1 ± 0	1.2 ± 0	1.4 ± 0	0.4 ± 0	1.4 ± 0	0.8 ± 0	9.2 ± 0	1.5 ± 0.1	0.5 ± 0	11.1 ± 0	2.1 ± 0
Bf1Al	nd	0.1 ± 0	1.2 ± 0	1.2 ± 0	0.5 ± 0	1.1 ± 0	0.6 ± 0	11.0 ± 0.1	2.4 ± 0	0.3 ± 0	12.2 ± 0	2.9 ± 0.1
BtAl	nd	nd	0.8 ± 0	1.1 ± 0	0.4 ± 0	1.2 ± 0	0.6 ± 0	9.6 ± 0	1.7 ± 0	0.4 ± 0.1	12.3 ± 0	2.5 ± 0
Me2Al	0.1 ± 0	0.1 ± 0	1.0 ± 0	1.0 ± 0	0.5 ± 0	1.1 ± 0	0.5 ± 0	10.3 ± 0.1	2.3 ± 0	0.3 ± 0	11.5 ± 0.1	2.8 ± 0
Bf2Al	0.1 ± 0	0.1 ± 0	1.2 ± 0	1.5 ± 0	0.5 ± 0	1.5 ± 0	0.8 ± 0	9.9 ± 0.1	1.9 ± 0	0.5 ± 0	12.0 ± 0.1	2.5 ± 0
TzLAl	0.1 ± 0	0.1 ± 0	1.9 ± 0	1.7 ± 0	0.5 ± 0	1.5 ± 0	0.8 ± 0	11.2 ± 0	2.3 ± 0	0.4 ± 0	12.9 ± 0	2.8 ± 0
TzSAl	0.1 ± 0	0.1 ± 0	1.2 ± 0	1.5 ± 0	0.5 ± 0	1.6 ± 0	0.8 ± 0	9.5 ± 0	1.8 ± 0	0.5 ± 0	11.2 ± 0	2.4 ± 0
ThAl	0.1 ± 0	0.1 ± 0	1.5 ± 0	1.9 ± 0	0.4 ± 0	1.7 ± 0.1	0.9 ± 0	9.0 ± 0	1.6 ± 0	0.5 ± 0	10.7 ± 0.1	2.2 ± 0
JhAl	0.1 ± 0	0.1 ± 0	2.1 ± 0	2.0 ± 0	0.4 ± 0	1.8 ± 0	0.9 ± 0	8.2 ± 0.1	1.8 ± 0	0.5 ± 0	10.7 ± 0.1	2.5 ± 0
DgnM	0.1 ± 0	0.2 ± 0	2.2 ± 0	2.0 ± 0	1.0 ± 0	1.6 ± 0	0.8 ± 0	16.9 ± 0	3.1 ± 0	0.4 ± 0	13.5 ± 0	2.8 ± 0
BenM	nd	0.1 ± 0	1.8 ± 0	2.4 ± 0	0.4 ± 0	2.0 ± 0	1.0 ± 0	10.2 ± 0	2.1 ± 0	0.5 ± 0	11.5 ± 0	2.6 ± 0
IkhnM	nd	0.1 ± 0	1.8 ± 0	2.1 ± 0	0.5 ± 0	1.7 ± 0	0.9 ± 0	11.1 ± 0	2.3 ± 0	0.4 ± 0	12.4 ± 0	2.7 ± 0
MenM	nd	0.1 ± 0	1.5 ± 0	1.7 ± 0	0.6 ± 0	1.7 ± 0	0.8 ± 0	12.2 ± 0.1	3.0 ± 0	0.4 ± 0	13.0 ± 0.1	3.4 ± 0

Continue Table S4. triacylglycerol composition of DSO samples (%)

	LaOlPa	LaLiPa	PaPaPa	MyOlPa	LaOlOl	LaLiOl	PaOlPa	MyOlOl	PaLiPa	MyLiOl	PaOlSt
Sy1Ag	13.8 ± 0	2.4 ± 0	0.2 ± 0	7.3 ± 0	8.9 ± 0	2.3 ± 0	2.8 ± 0	3.8 ± 0	0.5 ± 0	0.9 ± 0	0.9 ± 0
Sy2Ag	12.3 ± 0	2.1 ± 0	0.2 ± 0	6.9 ± 0.1	13.8 ± 0.1	3.4 ± 0	3.2 ± 0	5.9 ± 0	0.6 ± 0	1.4 ± 0	1.3 ± 0
BtbAg	12.8 ± 0	2.2 ± 0.1	0.2 ± 0	6.9 ± 0	13.6 ± 0	3.5 ± 0	3.2 ± 0	5.7 ± 0	0.6 ± 0	1.4 ± 0	1.2 ± 0
BtwAg	13.1 ± 0	2.5 ± 0	0.1 ± 0	7.0 ± 0	13 ± 0	3.5 ± 0	3.0 ± 0	5.5 ± 0	0.6 ± 0	1.4 ± 0	1.1 ± 0
IkAg	12.3 ± 0	2.5 ± 0	0.1 ± 0	6.1 ± 0	14 ± 0	4.0 ± 0	2.5 ± 0	5.7 ± 0	0.5 ± 0	1.5 ± 0	0.8 ± 0
LtAg	13.4 ± 0	2.8 ± 0.1	nd	6.0 ± 0	13.4 ± 0	4.3 ± 0	2.1 ± 0	4.7 ± 0	0.5 ± 0	1.3 ± 0	0.6 ± 0
BfAg	13.4 ± 0.1	3.1 ± 0	nd	7.6 ± 0	12.0 ± 0	3.9 ± 0	3.4 ± 0	6.4 ± 0.1	0.9 ± 0	2.0 ± 0	1.2 ± 0
AmAg	13.0 ± 0.1	2.4 ± 0	0.2 ± 0	6.8 ± 0	13.2 ± 0	3.6 ± 0	3.1 ± 0	5.6 ± 0	0.7 ± 0	1.4 ± 0	1.1 ± 0
MeAg	13.2 ± 0.1	2.4 ± 0.1	0.1 ± 0	6.9 ± 0	12.1 ± 0	3.2 ± 0	2.8 ± 0	5.0 ± 0	0.6 ± 0	1.2 ± 0	1.0 ± 0
MeE	12.8 ± 0	2.7 ± 0	0.1 ± 0	6.6 ± 0.1	13.2 ± 0.1	4.2 ± 0	2.7 ± 0.1	5.5 ± 0.1	0.7 ± 0	1.7 ± 0	1.1 ± 0
TzEr	13.2 ± 0.1	2.2 ± 0	0.1 ± 0	7.1 ± 0	13.6 ± 0	3.4 ± 0	3.0 ± 0	5.4 ± 0	0.6 ± 0	1.3 ± 0	1.3 ± 0
RcEr	13.7 ± 0.1	2.4 ± 0	0.1 ± 0	6.5 ± 0.1	13.6 ± 0.6	3.5 ± 0	2.7 ± 0	5.3 ± 0	0.6 ± 0	1.4 ± 0	0.8 ± 0
Bf1Er	13.4 ± 0.1	2.7 ± 0	0.1 ± 0	7.0 ± 0.1	12.6 ± 0.1	3.8 ± 0	3.0 ± 0	5.6 ± 0	0.7 ± 0	1.6 ± 0.1	1.1 ± 0
Bf2Er	13.5 ± 0	2.4 ± 0	0.2 ± 0	7.7 ± 0.1	11.7 ± 0.1	3.3 ± 0	3.4 ± 0	5.6 ± 0	0.7 ± 0	1.5 ± 0	1.5 ± 0
BhEr	12.3 ± 0.1	2.8 ± 0.1	0.2 ± 0	6.2 ± 0.1	14.0 ± 0	4.3 ± 0	2.8 ± 0	5.6 ± 0	0.8 ± 0	1.7 ± 0	1.1 ± 0
Kh1Er	13.8 ± 0.1	2.5 ± 0	0.2 ± 0	7.8 ± 0.1	11.6 ± 0.1	3.3 ± 0	3.4 ± 0.5	5.6 ± 0.1	0.8 ± 0.1	1.5 ± 0	1.6 ± 0.1
Kh2Er	14.0 ± 0.1	3.2 ± 0.1	0.2 ± 0.1	7.1 ± 0.1	10.7 ± 0.1	3.4 ± 0.1	3.5 ± 0.5	4.6 ± 0	0.8 ± 0	1.4 ± 0	1.2 ± 0
Me1Al	13.0 ± 0	2.1 ± 0.1	0.1 ± 0	7.3 ± 0	13.8 ± 0.1	3.3 ± 0	3.7 ± 0	6.7 ± 0.1	0.7 ± 0	1.7 ± 0.1	1.5 ± 0
Bf1Al	12.6 ± 0	2.6 ± 0	0.1 ± 0	6.6 ± 0	14.0 ± 0	4.3 ± 0	2.8 ± 0	6.0 ± 0	0.7 ± 0	1.8 ± 0	1.1 ± 0
BtAl	14.7 ± 0	2.6 ± 0.1	0.1 ± 0	8.4 ± 0	11.6 ± 0.1	3.2 ± 0.1	4.1 ± 0	6.3 ± 0	0.9 ± 0.1	1.7 ± 0	1.6 ± 0
Me2Al	12.6 ± 0.1	2.7 ± 0.1	0.2 ± 0	7.0 ± 0.1	13.6 ± 0.9	4.3 ± 0	3.1 ± 0	5.7 ± 0	0.8 ± 0	1.9 ± 0	1.4 ± 0
Bf2Al	14.0 ± 0.2	2.6 ± 0.1	0.2 ± 0	8.0 ± 0.1	11.9 ± 0.7	3.2 ± 0	3.9 ± 0	5.6 ± 0.1	0.9 ± 0	1.5 ± 0	1.6 ± 0
TzLAl	13.7 ± 0	2.8 ± 0.1	0.2 ± 0	7.5 ± 0	12.3 ± 0.1	3.5 ± 0	3.3 ± 0	5.6 ± 0	0.8 ± 0	1.6 ± 0	1.3 ± 0
TzSAl	13.5 ± 0.1	2.4 ± 0.1	0.2 ± 0	7.7 ± 0.1	12.8 ± 0.1	3.5 ± 0	3.7 ± 0	5.9 ± 0	0.7 ± 0	1.5 ± 0	1.6 ± 0
ThAl	12.6 ± 0	2.1 ± 0.1	0.2 ± 0	7.3 ± 0.1	13.5 ± 0.1	3.3 ± 0	3.7 ± 0.1	6.3 ± 0	0.7 ± 0	1.5 ± 0	1.7 ± 0.1
JhAl	13.0 ± 0	2.6 ± 0.1	0.1 ± 0	7.1 ± 0.1	12.0 ± 0.1	3.7 ± 0.1	3.4 ± 0	5.6 ± 0	0.8 ± 0	1.6 ± 0	1.5 ± 0
DgnM	13.2 ± 0	2.3 ± 0	0.1 ± 0	6.2 ± 0	13.0 ± 0	3.6 ± 0	2.3 ± 0	4.4 ± 0	0.4 ± 0	1.2 ± 0	0.9 ± 0
BenM	12.7 ± 0.1	2.4 ± 0.2	0.1 ± 0	5.6 ± 0.1	14.2 ± 0.1	4.0 ± 0	2.4 ± 0	6.0 ± 0.1	0.6 ± 0	1.7 ± 0.1	0.8 ± 0
IkhnM	13.1 ± 0.1	2.4 ± 0	0.1 ± 0	6.8 ± 0	12.8 ± 0.1	3.7 ± 0.1	3.0 ± 0	6.1 ± 0	0.7 ± 0	1.7 ± 0	1.1 ± 0
MenM	12.6 ± 0	3.1 ± 0	0.1 ± 0	6.6 ± 0	12.8 ± 0.1	4.5 ± 0	2.5 ± 0	5.3 ± 0.1	0.8 ± 0	1.9 ± 0	0.9 ± 0

Continue Table S4. triacylglycerol composition of DSO samples (%)

	PaOlOl	PaLiSt	PaLiOl	PaLiLi	StOlSt	StOlOl	OlOlOl	StLiOl	OlLiOl	LiLiOl	LiLiLi
Sy1Ag	2.1 ± 0	0.2 ± 0	0.6 ± 0	0.1 ± 0	0.2 ± 0	0.7 ± 0	0.7 ± 0	0.2 ± 0	0.3 ± 0	nd	nd
Sy2Ag	5.2 ± 0.1	0.3 ± 0	1.3 ± 0	0.1 ± 0	0.4 ± 0	1.7 ± 0	3.4 ± 0	0.5 ± 0	1.0 ± 0	0.2 ± 0	0.2 ± 0.1
BtbAg	5.1 ± 0	0.3 ± 0	1.4 ± 0	0.2 ± 0	0.3 ± 0	1.5 ± 0	3.3 ± 0	0.4 ± 0	1.1 ± 0	0.2 ± 0	0.1 ± 0
BtwAg	4.1 ± 0	0.3 ± 0	1.2 ± 0	0.2 ± 0	0.3 ± 0	1.2 ± 0	2.3 ± 0	0.4 ± 0	0.8 ± 0	0.2 ± 0	0.1 ± 0
IkAg	3.9 ± 0	0.2 ± 0	1.2 ± 0	0.2 ± 0	0.2 ± 0	1.1 ± 0	2.2 ± 0	0.3 ± 0	0.8 ± 0	0.2 ± 0	0.1 ± 0
LtAg	3.4 ± 0	0.2 ± 0	1.2 ± 0	0.2 ± 0	0.1 ± 0	0.8 ± 0	2.1 ± 0	0.2 ± 0	0.9 ± 0	0.3 ± 0	0.1 ± 0
BfAg	5 ± 0.1	0.4 ± 0	1.7 ± 0	0.2 ± 0	0.3 ± 0	1.3 ± 0	2.6 ± 0	0.5 ± 0	1.1 ± 0	0.3 ± 0	0.2 ± 0
AmAg	4.9 ± 0.1	0.3 ± 0	1.4 ± 0	0.2 ± 0	0.3 ± 0	1.3 ± 0	2.7 ± 0	0.4 ± 0	0.9 ± 0	0.2 ± 0	0.1 ± 0
MeAg	3.5 ± 0	0.2 ± 0	1.0 ± 0	0.1 ± 0	0.2 ± 0	1.0 ± 0	1.6 ± 0	0.3 ± 0	0.5 ± 0	0.1 ± 0	0.1 ± 0
MeE	4.3 ± 0	0.3 ± 0	1.5 ± 0	0.2 ± 0	0.3 ± 0	1.3 ± 0	2.8 ± 0	0.5 ± 0	1.2 ± 0.1	0.3 ± 0	0.1 ± 0
TzEr	4.3 ± 0	0.3 ± 0	1.1 ± 0	0.1 ± 0	0.4 ± 0	1.5 ± 0	2.4 ± 0	0.4 ± 0	0.7 ± 0	0.1 ± 0	0.1 ± 0
RcEr	4.3 ± 0.1	0.2 ± 0	1.2 ± 0.1	0.1 ± 0	0.2 ± 0	1.0 ± 0	2.4 ± 0	0.3 ± 0	0.8 ± 0	0.2 ± 0	0.1 ± 0
Bf1Er	4.4 ± 0	0.3 ± 0	1.4 ± 0	0.2 ± 0	0.3 ± 0	1.2 ± 0	2.7 ± 0	0.4 ± 0	1.1 ± 0	0.2 ± 0	0.1 ± 0
Bf2Er	4.7 ± 0.1	0.4 ± 0	1.4 ± 0.1	0.2 ± 0.1	0.5 ± 0	1.7 ± 0	2.8 ± 0	0.5 ± 0	1.0 ± 0	0.2 ± 0	0.1 ± 0
BhEr	5.1 ± 0	0.3 ± 0	1.8 ± 0	0.3 ± 0.1	0.3 ± 0	1.4 ± 0	3.3 ± 0.1	0.5 ± 0	1.4 ± 0	0.3 ± 0	0.2 ± 0
Kh1Er	5.1 ± 0.1	0.4 ± 0	1.6 ± 0.1	0.3 ± 0.1	0.5 ± 0.1	1.7 ± 0	2.8 ± 0	0.6 ± 0	1.0 ± 0	0.2 ± 0	0.2 ± 0
Kh2Er	3.6 ± 0	0.3 ± 0	1.2 ± 0	0.2 ± 0	0.3 ± 0	1.1 ± 0	1.4 ± 0	0.4 ± 0	0.6 ± 0	0.2 ± 0	0.2 ± 0
Me1Al	6.3 ± 0	0.3 ± 0	1.7 ± 0.1	0.2 ± 0	0.5 ± 0	1.9 ± 0	3.8 ± 0	0.4 ± 0	1.1 ± 0	0.2 ± 0	0.2 ± 0
Bf1Al	4.7 ± 0	0.3 ± 0	1.6 ± 0	0.2 ± 0	0.3 ± 0	1.5 ± 0	3.1 ± 0	0.5 ± 0.1	1.2 ± 0.1	0.3 ± 0	0.1 ± 0
BtAl	5.6 ± 0.1	0.4 ± 0	1.6 ± 0.1	0.2 ± 0	0.4 ± 0	1.6 ± 0	2.6 ± 0	0.5 ± 0	0.9 ± 0	0.2 ± 0	0.2 ± 0
Me2Al	4.8 ± 0.1	0.4 ± 0	1.8 ± 0.1	0.3 ± 0.1	0.4 ± 0.1	1.6 ± 0.1	3.1 ± 0.1	0.6 ± 0	1.4 ± 0.1	0.4 ± 0.1	0.2 ± 0
Bf2Al	5.1 ± 0.1	0.4 ± 0	1.5 ± 0	0.3 ± 0	0.5 ± 0	1.6 ± 0	2.6 ± 0	0.5 ± 0	1.0 ± 0.1	0.2 ± 0	0.2 ± 0
TzLAl	4.1 ± 0	0.3 ± 0	1.2 ± 0	0.2 ± 0	0.4 ± 0	1.4 ± 0	2.0 ± 0	0.4 ± 0	0.8 ± 0	0.2 ± 0	0.2 ± 0
TzSAl	5.5 ± 0	0.3 ± 0	1.5 ± 0	0.2 ± 0	0.5 ± 0	1.8 ± 0	3.2 ± 0	0.5 ± 0	1.1 ± 0	0.2 ± 0	0.2 ± 0
ThAl	6 ± 0	0.3 ± 0	1.6 ± 0	0.2 ± 0	0.5 ± 0.1	2.0 ± 0	3.8 ± 0	0.5 ± 0	1.2 ± 0	0.2 ± 0	0.2 ± 0
JhAl	5.9 ± 0	0.4 ± 0.1	1.9 ± 0	0.3 ± 0	0.4 ± 0.1	1.8 ± 0	3.9 ± 0	0.6 ± 0	1.5 ± 0	0.3 ± 0	0.2 ± 0
DgnM	3.0 ± 0	0.2 ± 0	0.8 ± 0	0.1 ± 0	0.3 ± 0	0.9 ± 0	1.5 ± 0	0.3 ± 0	0.6 ± 0	0.1 ± 0	0.1 ± 0
BenM	5.3 ± 0.1	0.2 ± 0	1.6 ± 0	0.2 ± 0	0.2 ± 0	1.3 ± 0	3.9 ± 0	0.4 ± 0	1.4 ± 0	0.2 ± 0	0.2 ± 0
IkhnM	4.5 ± 0	0.3 ± 0	1.3 ± 0	0.1 ± 0	0.4 ± 0	1.3 ± 0	2.8 ± 0	0.4 ± 0	1.0 ± 0.1	0.2 ± 0	0.1 ± 0
MenM	3.5 ± 0.1	0.3 ± 0	1.4 ± 0	0.2 ± 0	0.2 ± 0	1.1 ± 0	2.1 ± 0	0.4 ± 0	1.0 ± 0	0.3 ± 0	0.1 ± 0

Results are expressed as mean ± SD (n = 3), nd: not detected, Cy: caprylic acid, Ca: capric acid, La: lauric acid, My: myristic acid, Pa: palmitic acid, St: searic acid, Ol: oleic acid, Li: linoleic acid

Table S5. tocochromanol composition of DSO samples (mg/kg of oil)

	α-Tocopherol	α-Tocotrienol	β-Tocotrienol	γ-Tocotrienol	δ-Tocotrienol	Sum
Sy1Ag	14 ± 1	395 ± 6	102 ± 2	30 ± 1	24 ± 0	565 ± 5
Sy2Ag	8 ± 0	351 ± 2	63 ± 0	38 ± 0	13 ± 1	473 ± 2
BtbAg	5 ± 0	387 ± 4	70 ± 0	39 ± 0	16 ± 0	517 ± 4
BtwAg	12 ± 0	350 ± 2	80 ± 1	48 ± 0	22 ± 0	512 ± 3
IkAg	2 ± 0	440 ± 2	72 ± 2	98 ± 0	30 ± 2	642 ± 5
LtAg	19 ± 1	360 ± 4	64 ± 2	20 ± 1	10 ± 0	473 ± 5
BfAg	nd	382 ± 5	79 ± 4	38 ± 0	20 ± 1	519 ± 8
AmAg	9 ± 1	334 ± 24	117 ± 11	28 ± 1	21 ± 1	508 ± 39
MeAg	18 ± 1	324 ± 2	94 ± 3	33 ± 1	18 ± 0	486 ± 7
MeE	nd	362 ± 3	51 ± 3	43 ± 0	11 ± 1	467 ± 10
TzEr	nd	352 ± 3	69 ± 7	50 ± 0	19 ± 2	491 ± 9
RcEr	8 ± 0	382 ± 11	86 ± 2	52 ± 0	28 ± 1	557 ± 16
Bf1Er	nd	343 ± 3	57 ± 2	45 ± 0	15 ± 1	459 ± 7
Bf2Er	36 ± 2	441 ± 15	86 ± 3	26 ± 1	12 ± 1	601 ± 20
BhEr	21 ± 1	508 ± 4	129 ± 1	33 ± 1	13 ± 1	704 ± 6
Kh1Er	25 ± 0	480 ± 10	80 ± 2	24 ± 1	11 ± 1	593 ± 57
Kh2Er	18 ± 2	405 ± 1	65 ± 2	67 ± 1	18 ± 0	573 ± 4
Me1Al	nd	422 ± 6	49 ± 3	35 ± 3	10 ± 0	516 ± 6
Bf1Al	nd	433 ± 12	84 ± 4	33 ± 1	18 ± 2	568 ± 12
BtAl	3 ± 0	342 ± 10	49 ± 0	24 ± 1	7 ± 0	424 ± 9
Me2Al	26 ± 1	563 ± 7	61 ± 3	32 ± 2	10 ± 0	692 ± 12
Bf2Al	16 ± 0	504 ± 5	113 ± 2	28 ± 1	19 ± 0	679 ± 8
TzLAl	31 ± 3	491 ± 2	74 ± 1	56 ± 1	18 ± 0	671 ± 6
TzSAl	48 ± 2	539 ± 7	89 ± 3	41 ± 1	15 ± 0	733 ± 12
ThAl	24 ± 0	466 ± 3	71 ± 5	30 ± 1	11 ± 0	602 ± 3
JhAl	18 ± 0	519 ± 9	169 ± 5	29 ± 1	25 ± 1	760 ± 13
DgnM	17 ± 1	560 ± 1	71 ± 3	28 ± 1	11 ± 1	687 ± 6
BenM	nd	429 ± 14	53 ± 1	14 ± 0	5 ± 0	502 ± 16
IkhnM	nd	382 ± 10	64 ± 2	26 ± 0	13 ± 2	485 ± 13
MenM	20 ± 0	372 ± 3	46 ± 2	24 ± 0	6 ± 0	468 ± 4

Results are expressed as mean ± SD (n = 3), nd: not detected

Table S6. phytosterol composition of DSO samples (mg/kg of oil)

	Cholesterol	24-Methylenecholesterol	Campesterol	Campestanol	Stigmasterol	Δ5,23-Stigmastadienol	β-Sitosterol	Sitostanol	Δ5-Avenasterol	Δ5,24-Stigmastadienol	Δ7-Stigmastenol	Δ7-Avenasterol	Sum
Sy1Ag	29 ± 1	21 ± 1	372 ± 6	11 ± 1	129 ± 3	35 ± 1	2540 ± 57	30 ± 2	517 ± 10	46 ± 2	1 ± 0	8 ± 2	3739 ± 69
Sy2Ag	27 ± 1	24 ± 5	327 ± 6	8 ± 1	115 ± 0	34 ± 2	2329 ± 25	33 ± 3	443 ± 6	51 ± 4	11 ± 0	21 ± 1	3422 ± 24
BtbAg	41 ± 1	50 ± 10	372 ± 12	6 ± 0	114 ± 0	42 ± 2	2830 ± 25	22 ± 3	919 ± 2	62 ± 2	10 ± 3	19 ± 2	4487 ± 17
BtwAg	40 ± 2	32 ± 6	379 ± 5	6 ± 0	128 ± 2	41 ± 0	2917 ± 47	22 ± 2	700 ± 6	57 ± 2	13 ± 0	22 ± 1	4357 ± 45
IkAg	54 ± 2	59 ± 2	506 ± 12	4 ± 1	127 ± 2	43 ± 3	3075 ± 60	17 ± 1	653 ± 5	46 ± 6	2 ± 0	13 ± 2	4600 ± 67
LtAg	38 ± 1	25 ± 3	325 ± 3	9 ± 0	124 ± 1	33 ± 1	2413 ± 14	32 ± 2	554 ± 1	48 ± 2	9 ± 2	23 ± 0	3631 ± 9
BfAg	34 ± 0	37 ± 3	442 ± 1	4 ± 0	163 ± 0	40 ± 4	3079 ± 20	25 ± 4	557 ± 9	51 ± 5	12 ± 3	10 ± 2	4454 ± 25
AmAg	34 ± 1	32 ± 6	307 ± 7	6 ± 0	136 ± 2	41 ± 2	2753 ± 5	27 ± 3	739 ± 6	55 ± 1	7 ± 2	14 ± 1	4151 ± 15
MeAg	34 ± 0	26 ± 4	396 ± 3	6 ± 0	128 ± 3	36 ± 1	2653 ± 44	22 ± 2	537 ± 5	50 ± 2	6 ± 1	10 ± 1	3903 ± 50
MeE	31 ± 4	39 ± 0	390 ± 6	3 ± 1	213 ± 2	41 ± 1	2786 ± 37	22 ± 2	565 ± 10	48 ± 1	4 ± 0	14 ± 4	4156 ± 60
TzEr	35 ± 2	56 ± 1	368 ± 9	5 ± 0	174 ± 3	41 ± 2	2290 ± 58	28 ± 2	737 ± 15	57 ± 3	6 ± 1	20 ± 1	3816 ± 90
RcEr	32 ± 1	76 ± 2	550 ± 19	14 ± 2	156 ± 3	37 ± 1	2168 ± 28	37 ± 4	753 ± 14	60 ± 1	5 ± 1	23 ± 1	3912 ± 59
Bf1Er	39 ± 2	43 ± 8	360 ± 20	4 ± 1	204 ± 4	38 ± 2	2534 ± 40	24 ± 4	626 ± 17	47 ± 4	3 ± 0	9 ± 3	3931 ± 86
Bf2Er	34 ± 1	58 ± 2	445 ± 15	11 ± 1	141 ± 2	40 ± 1	2341 ± 69	33 ± 1	607 ± 38	50 ± 2	3 ± 1	3 ± 1	3765 ± 126
BhEr	50 ± 2	54 ± 0	491 ± 3	10 ± 1	154 ± 0	40 ± 1	2554 ± 9	33 ± 1	601 ± 5	54 ± 1	8 ± 2	18 ± 3	4067 ± 11

Continue Table S6. phytosterol composition of DSO samples (mg/kg of oil)

	Cholesterol	24-Methylenecholesterol	Campesterol	Campestanol	Stigmasterol	Δ5,23-Stigmastadienol	β-Sitosterol	Sitostanol	Δ5-Avenasterol	Δ5,24-Stigmastadienol	Δ7-Stigmasterol	Δ7-Avenasterol	Sum
Kh1Er	35 ± 3	61 ± 3	414 ± 22	7 ± 1	154 ± 10	37 ± 2	2258 ± 110	34 ± 1	692 ± 31	56 ± 2	3 ± 1	8 ± 1	3760 ± 163
Kh2Er	43 ± 3	23 ± 1	386 ± 20	7 ± 0	121 ± 4	35 ± 2	2584 ± 125	30 ± 2	325 ± 11	39 ± 2	8 ± 1	16 ± 2	3616 ± 160
Me1Al	33 ± 4	35 ± 1	376 ± 4	5 ± 1	223 ± 4	53 ± 4	2913 ± 27	28 ± 1	546 ± 13	52 ± 5	6 ± 1	9 ± 1	4278 ± 34
Bf1Al	49 ± 6	55 ± 0	470 ± 8	7 ± 0	181 ± 0	56 ± 4	2809 ± 36	31 ± 0	652 ± 19	51 ± 4	7 ± 1	9 ± 0	4377 ± 32
BtAl	67 ± 9	65 ± 1	412 ± 5	5 ± 0	137 ± 2	58 ± 1	2795 ± 68	23 ± 1	848 ± 5	65 ± 3	6 ± 1	13 ± 1	4494 ± 93
Me2Al	28 ± 2	29 ± 1	379 ± 18	8 ± 1	191 ± 8	39 ± 2	2745 ± 135	30 ± 1	437 ± 16	42 ± 2	4 ± 1	8 ± 1	3939 ± 183
Bf2Al	44 ± 3	45 ± 1	466 ± 3	9 ± 1	165 ± 2	43 ± 1	2832 ± 20	36 ± 0	634 ± 4	58 ± 0	7 ± 1	11 ± 1	4349 ± 26
TzLAl	40 ± 4	31 ± 1	359 ± 5	9 ± 0	173 ± 7	37 ± 2	2676 ± 29	41 ± 0	461 ± 1	45 ± 1	7 ± 2	6 ± 1	3884 ± 48
TzSAl	28 ± 2	88 ± 2	672 ± 5	20 ± 0	152 ± 2	45 ± 1	2585 ± 22	46 ± 1	630 ± 5	57 ± 1	5 ± 1	10 ± 2	4338 ± 28
ThAl	40 ± 0	61 ± 1	352 ± 4	6 ± 1	119 ± 1	41 ± 1	2534 ± 23	29 ± 1	873 ± 8	61 ± 3	3 ± 1	14 ± 1	4133 ± 40
JhAl	51 ± 3	66 ± 1	471 ± 3	10 ± 0	150 ± 4	44 ± 1	2789 ± 48	43 ± 2	710 ± 9	60 ± 2	4 ± 0	10 ± 1	4408 ± 70
DgnM	65 ± 3	39 ± 1	498 ± 0	12 ± 0	180 ± 2	41 ± 0	2585 ± 5	42 ± 1	446 ± 2	54 ± 2	5 ± 1	9 ± 1	3976 ± 6
BenM	52 ± 3	156 ± 3	562 ± 5	14 ± 1	240 ± 6	53 ± 2	2484 ± 30	50 ± 2	1101 ± 9	88 ± 3	4 ± 1	24 ± 1	4827 ± 51
IkhnM	68 ± 4	66 ± 1	383 ± 7	8 ± 1	180 ± 8	45 ± 1	2469 ± 40	32 ± 0	829 ± 16	77 ± 1	3 ± 0	14 ± 1	4175 ± 73
MenM	31 ± 1	17 ± 4	391 ± 5	12 ± 2	213 ± 2	43 ± 2	2968 ± 26	39 ± 2	490 ± 2	49 ± 1	3 ± 0	6 ± 1	4262 ± 32

Results are expressed as mean ± SD (n = 3)

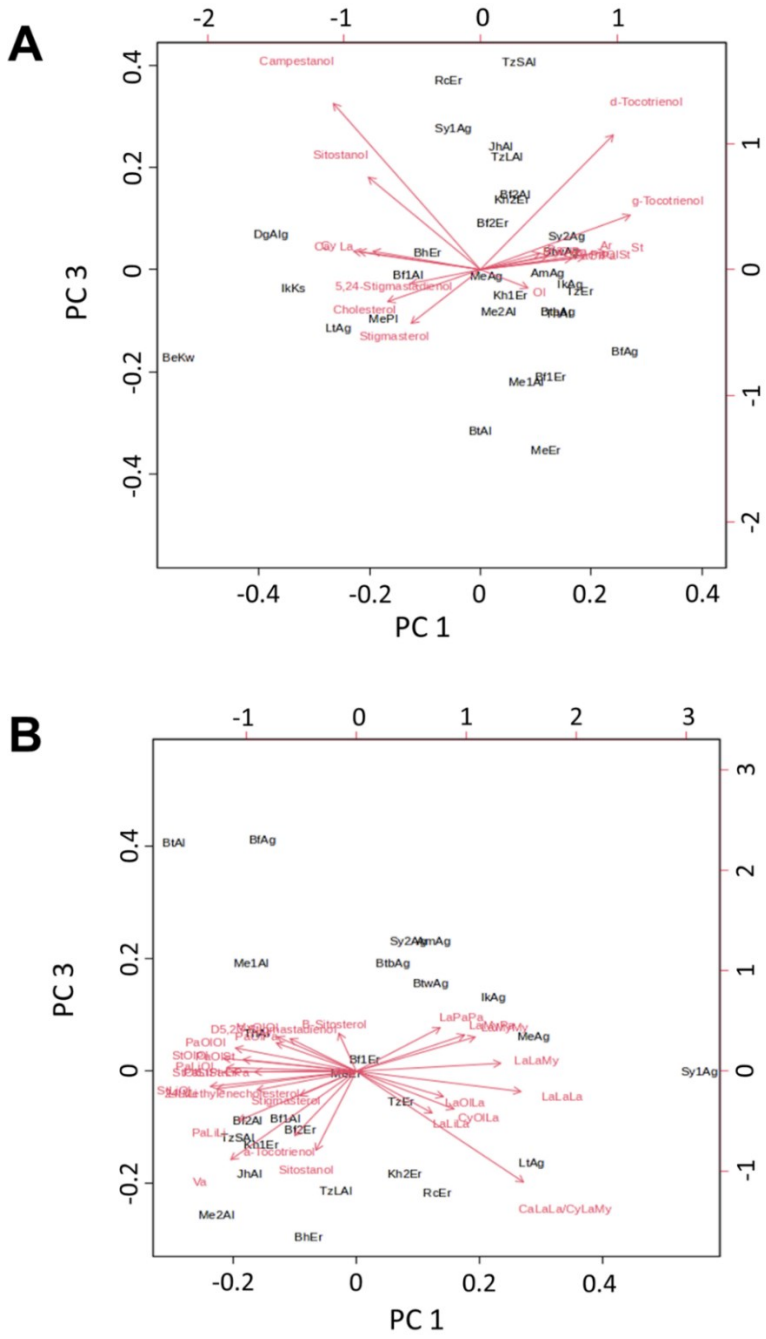


Fig. S1. Principal Component analysis (PCA) biplots showing which variables influence the principal components. A: comparison of Moroccan (M) samples against non-Moroccan (nM) samples B: comparison of Moroccan samples (*Allougoum* (Ag), *Errachidia* (Er), *Alnif* (Al)) against each other. Cy: caprylic acid, Ca: capric acid, La: lauric acid, My: myristic acid, Pa: palmitic acid, St: stearic acid, Ol: oleic acid, Li: linoleic acid, Va: vaccenic acid, Ar: arachidic acid. The full sample names are available in Table S1 for reference.

Supplementary data manuscript IV

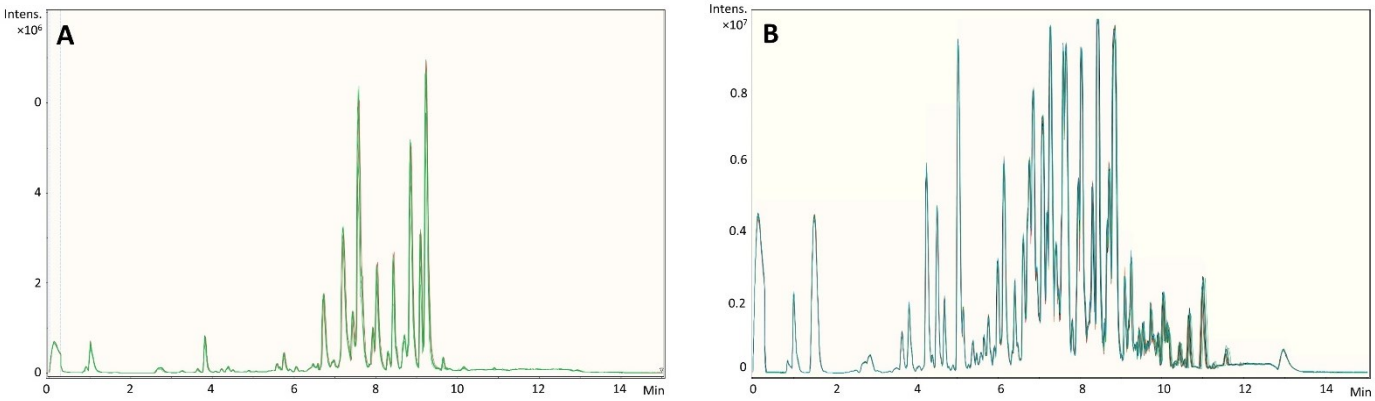


Fig.S1. Data quality assessment by an overlapped base peak chromatogram of the QC samples in the negative mode (A) and positive mode (B)

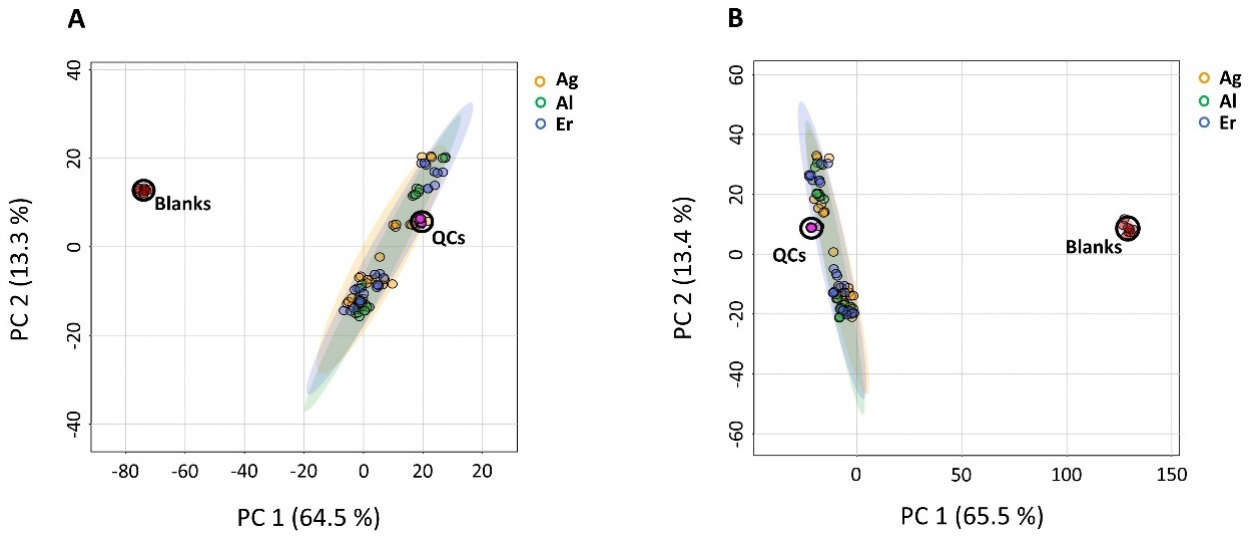


Fig. S2. Visual inspection of the score plots including the samples, QCs, and blanks visualization in negative (A) and positive (B) ionization modes

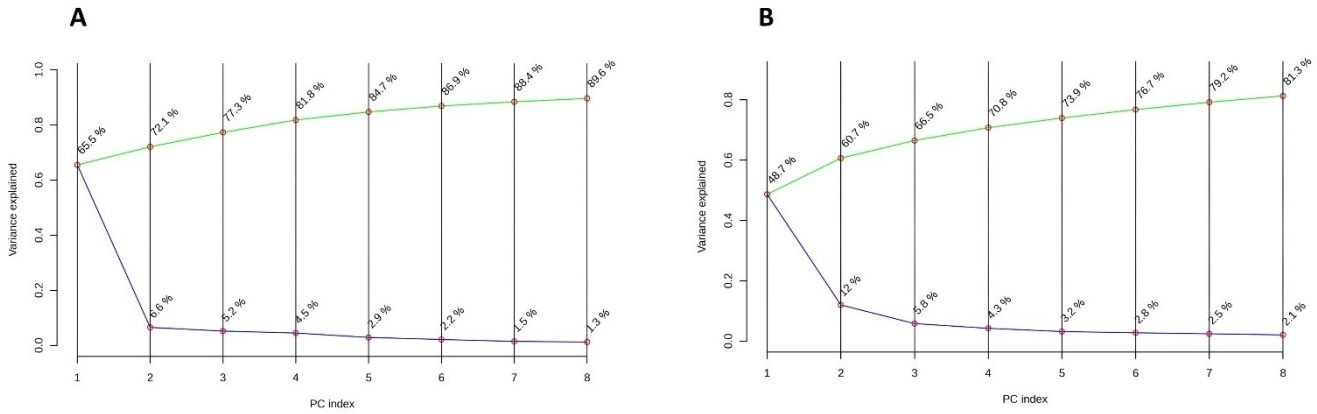


Fig. S3. Scree plot visualization in negative (A) and positive (B) ionization modes

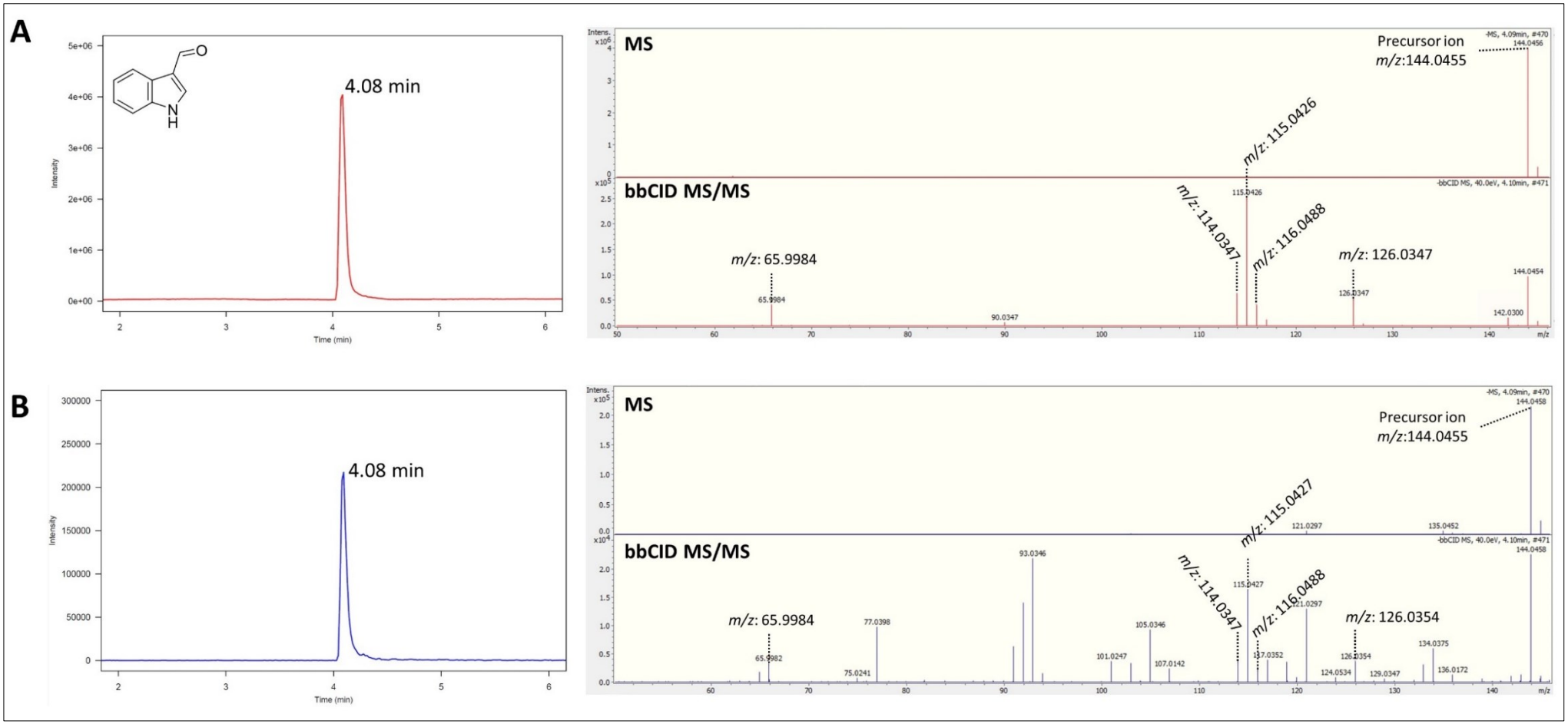


Fig. S4. Extracted ion chromatogram (EIC) and MS/MS spectrum of indole-3-carboxaldehyde, including annotated fragment ions. (A) Authentic standard; (B) Corresponding signal detected in the DSO sample.

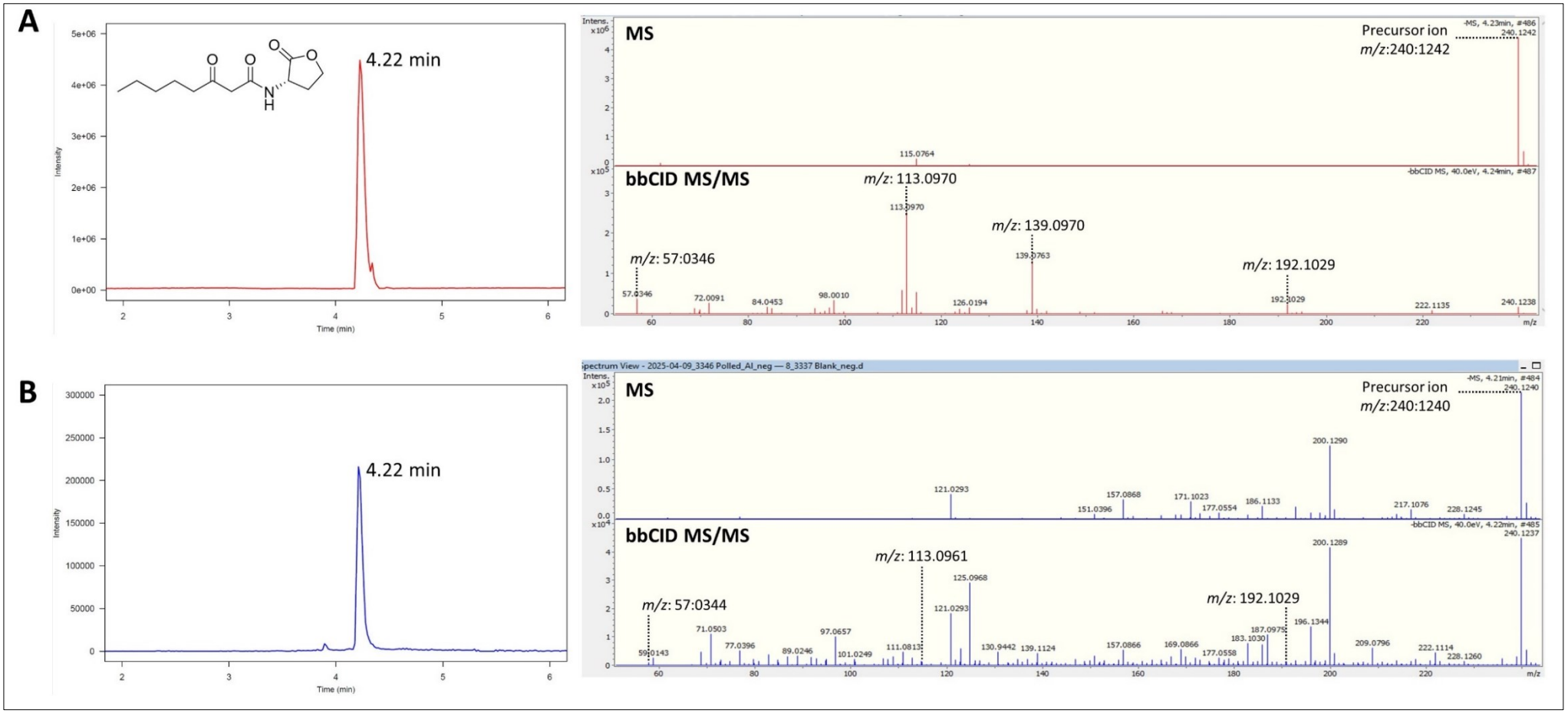


Fig. S5. Extracted ion chromatogram (EIC) and MS/MS spectrum of n-(3-oxohexanoyl) homoserine lactone, including annotated fragment ions. (A) Authentic standard; (B) Corresponding signal detected in the DSO sample.

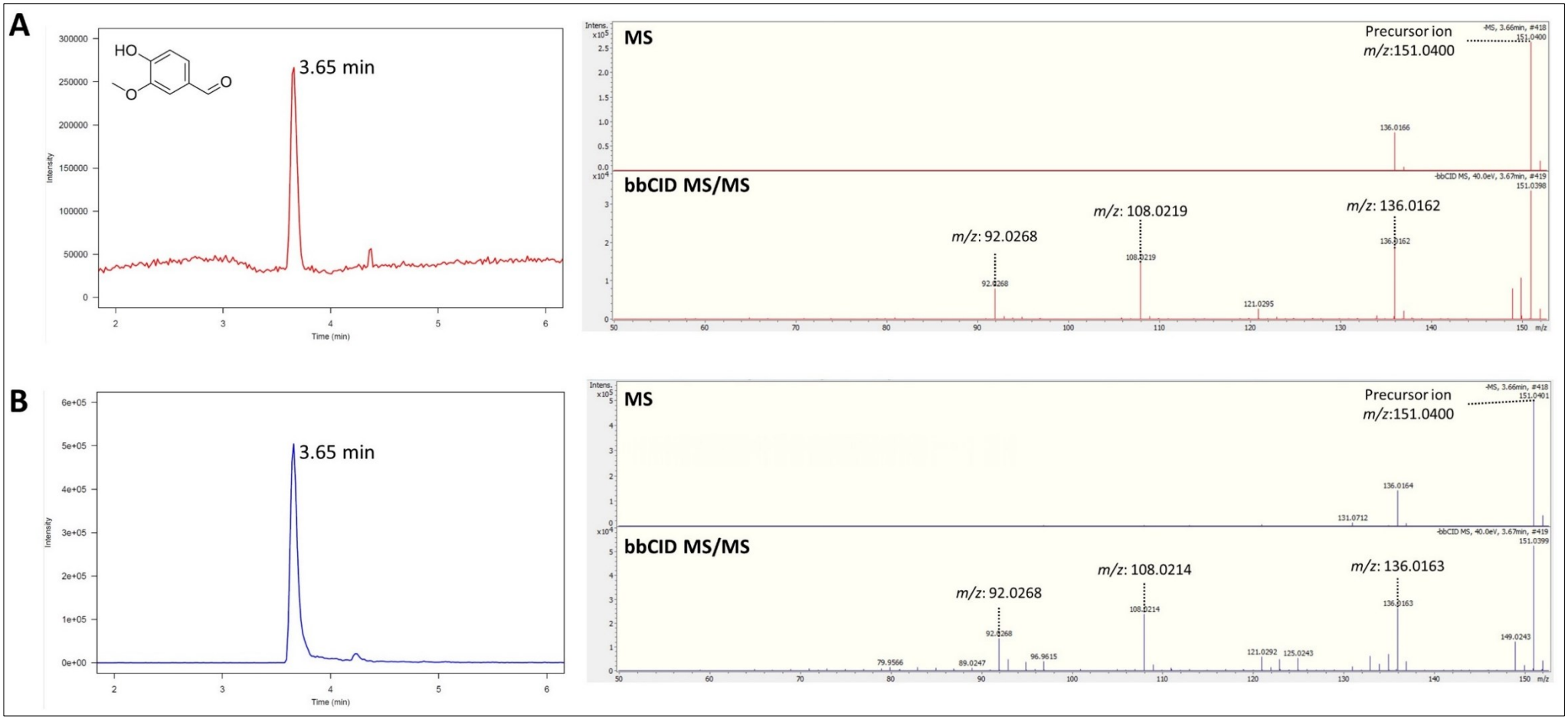


Fig. S6. Extracted ion chromatogram (EIC) and MS/MS spectrum of vanillin, including annotated fragment ions. (A) Authentic standard; (B) Corresponding signal detected in the DSO sample

Table S1. Applied MS-Dial (5.5) pre-processing parameters.

Data collection	
MS1 tolerance	0.01 Da
Retention time begin	1 min
Retention time end	12 min
MS1 mass range begin	50 Da
MS1 mass range end	1,000 Da
MS/MS mass range begin	50 Da
MS/MS mass range end	1,000 Da
Peak detection	
Minimum peak height	5,000 and 10,000 amplitude for negative and positive modes, respectively
Mass slice width	0.1 Da
Smoothing method	Linear weighted moving average
Smoothing level	3 scans
Adduct	
Positive	$[M+H]^+$
Negative	$[M-H]^-$
Alignment	
Retention time tolerance	0.05 min
MS1 tolerance	0.015 Da
Retention time factor	50%
MS1 factor	50%
Remove features based on blank information	Yes

Table S2. Top 50 features in negative modes classified according to their VIP scores

No.	Features (*mz*: $[M-H]^-$/RT)	VIP scores
1	585.487/9.254	2.820
2	390.24988/5.505	2.745
3	377.34232/8.546	2.664
4	309.16895/6.846	2.657
5	240.12329/4.222	2.585
6	475.36255/8.818	2.465
7	533.45569/9.114	2.424
8	439.34274/8.545	2.382
9	349.23672/9.342	2.366
10	151.07635/8.372	2.336
11	349.23599/9.293	2.308
12	441.41028/10.903	2.278
13	485.40012/10.913	2.252
14	327.17935/5.596	2.251
15	295.18887/8.447	2.228
16	241.18056/6.854	2.219
17	259.19217/5.597	2.148
18	541.37299/5.598	2.110
19	873.71466/8.676	2.012
20	425.36426/8.683	1.992
21	513.42828/11.593	1.968
22	439.39279/10.563	1.947
23	381.37244/8.686	1.944
24	379.35889/8.686	1.935
25	473.32813/8.366	1.918
26	483.38245/10.563	1.884
27	473.32755/8.329	1.876
28	421.2926/9.594	1.861
29	481.35443/9.106	1.860
30	144.04556/4.081	1.860
31	477.39307/8.445	1.856
32	457.36789/10.329	1.836
33	545.34723/8.547	1.827
34	313.21722/9.18	1.778
35	193.10242/5.53	1.758
36	265.08749/5.531	1.738
37	358.25916/6.955	1.703
38	225.07692/3.507	1.694
39	215.16576/5.767	1.689
40	413.3053/8.341	1.671
41	373.20163/7.639	1.664
42	169.15991/5.767	1.662
43	309.20755/5.865	1.611
44	255.12161/4.905	1.598
45	187.13432/4.905	1.597
46	141.00238/7.595	1.583
47	151.04033/3.654	1.557
48	281.2489/9.244	1.554
49	665.57214/9.356	1.529
50	455.10184/1.109	1.529

Table S3. Top 50 features in positive modes classified according to their VIP scores

No.	Features (*mz*: $[M+H]^+$/RT)	VIP scores
1	206.11487/4.046	2.592
2	485.38245/6.004	2.500
3	207.17369/6.005	2.489
4	189.16277/6.004	2.489
5	179.17937/6.004	2.471
6	243.19565/6.005	2.451
7	265.17712/6.005	2.445
8	507.36346/6.004	2.442
9	637.38867/8.733	2.420
~~10~~	220.13071/4.381	2.420
11	266.17242/4.666	2.416
12	281.1424/6.007	2.365
13	383.26126/6.004	2.342
14	409.2659/6.587	2.339
15	163.14793/6.153	2.329
16	397.33002/6.152	2.324
17	129.09044/6.153	2.322
18	199.16901/6.152	2.322
19	181.15848/6.152	2.316
20	302.99905/4.704	2.314
21	84.95966/4.707	2.311
22	321.00946/4.705	2.301
23	419.3125/6.153	2.291
24	327.14719/6.002	2.282
25	234.10933/5.916	2.277
26	459.06839/4.703	2.277
27	350.25711/8.671	2.271
28	241.02948/4.704	2.270
29	416.30338/6.154	2.263
30	271.09412/7.856	2.259
31	259.03983/4.704	2.257
32	741.51031/9.825	2.256
33	219.04752/4.704	2.254
34	193.11955/5.143	2.253
35	549.41174/9.386	2.251
36	221.1508/6.155	2.250
37	189.01544/7.859	2.248
38	477.43839/7.73	2.248
39	713.47766/9.395	2.242
40	248.12498/5.763	2.236
41	237.0578/4.702	2.223
42	222.14633/7.765	2.206
43	353.17145/7.857	2.197
44	463.33813/8.554	2.185
45	681.50262/7.858	2.174
46	214.11987/4.95	2.163
47	198.10974/3.55	2.162
48	153.12688/5.143	2.161
49	171.13736/5.143	2.158
50	363.24948/5.143	2.157

Table S4. Main MS/MS fragment ions for compounds annotated at Level 2, compared with reference databases

Compound	Formula	MS	MS/MS	Database / Reference Spectrum
12-hydroxydodecanoic	$C_{12}H_{24}O_3$	216.17303	169.1596 213.1494 197.1543	MassBanl of North America Spectrum MoNA034669 (last accessed 22/11/2024)
δ-dodecalactone	$C_{12}H_{22}O_2$	198.16224	136.1481 97.1010 121.1010	MassBanl of North America Spectrum FIO00053 (last accessed 22/11/2024)
2,4-dodecadienal	$C_{12}H_{20}O$	180.1513	81.0701 79.0544 93.0698	Foodb *Predicted Spectrum FDB019215 (last accessed 22/11/2024)

* Note: “Predicted spectrum” refers to in-silico spectra provided in the FooDB entry

Supplementary data manuscript V

Supplementary data A

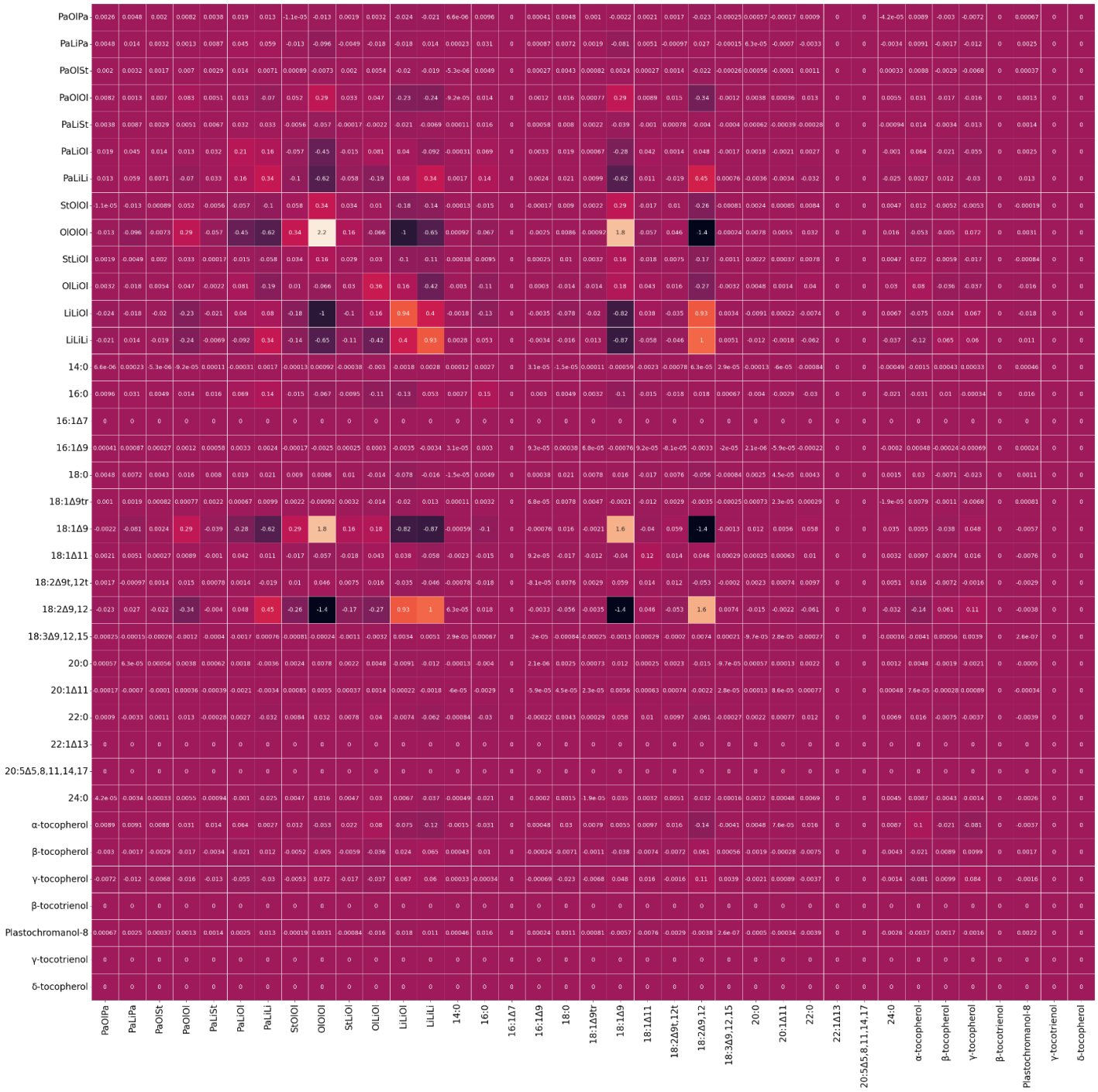


Fig. S1. Covariance matrix for the chemical composition of refined sunflower oil

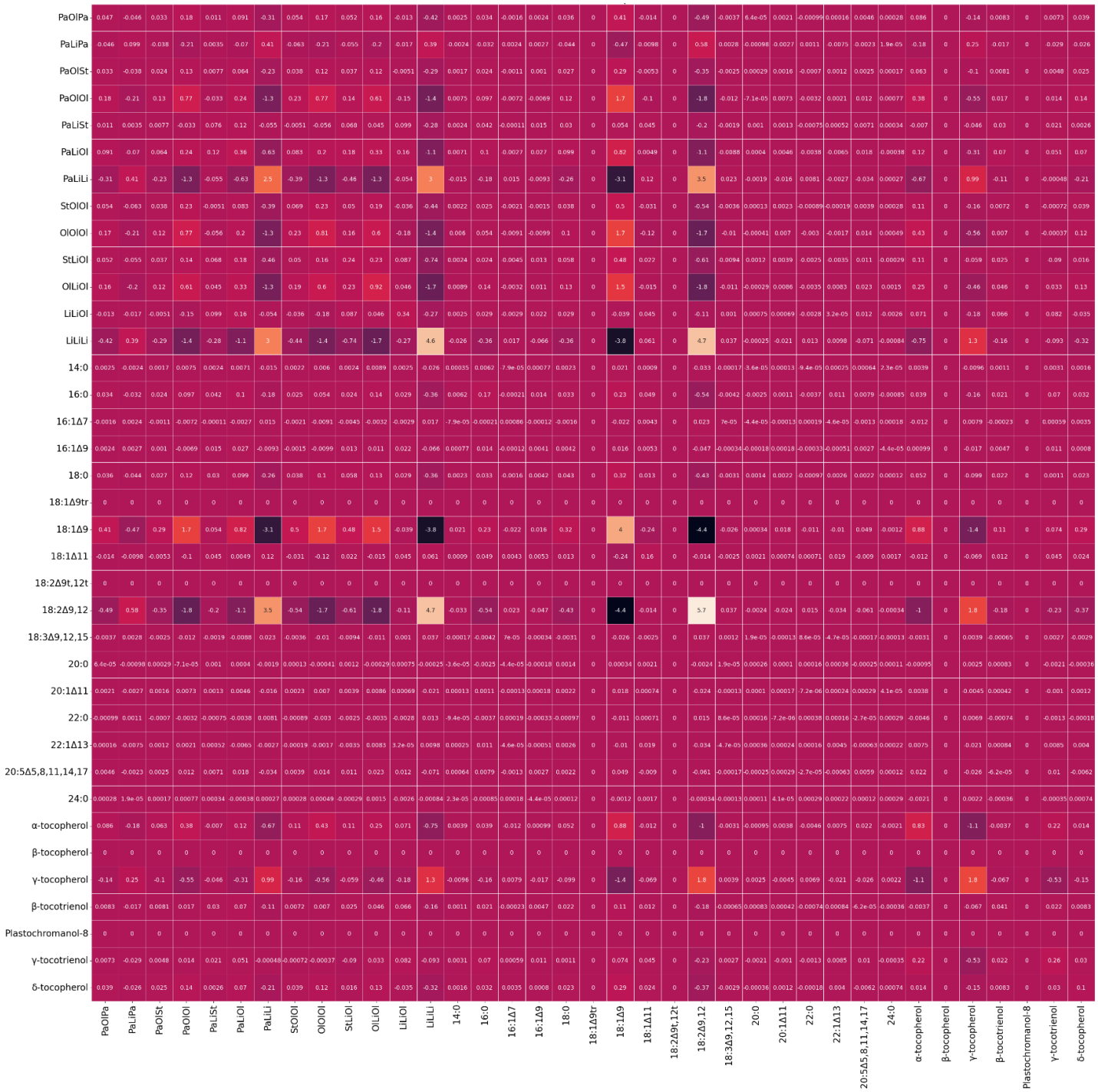


Fig. S2. Covariance matrix for the chemical composition of cactus seed oil

Table S1

Hyperparameter set for Random Forest (RF) models trained on data sets of different sizes.

Model	max_depth	Bootstrap	min_samples_split	max_features	min_samples_leaf	n_estimators
RF trained on 1,000 samples	9	True	6	All features	3	1,500
RF trained on 10,000 samples	24	True	2	All features	9	500
RF trained on 100,000 samples	24	True	2	All features	9	500
RF trained on 1,000,000 samples	24	True	2	All features	9	100

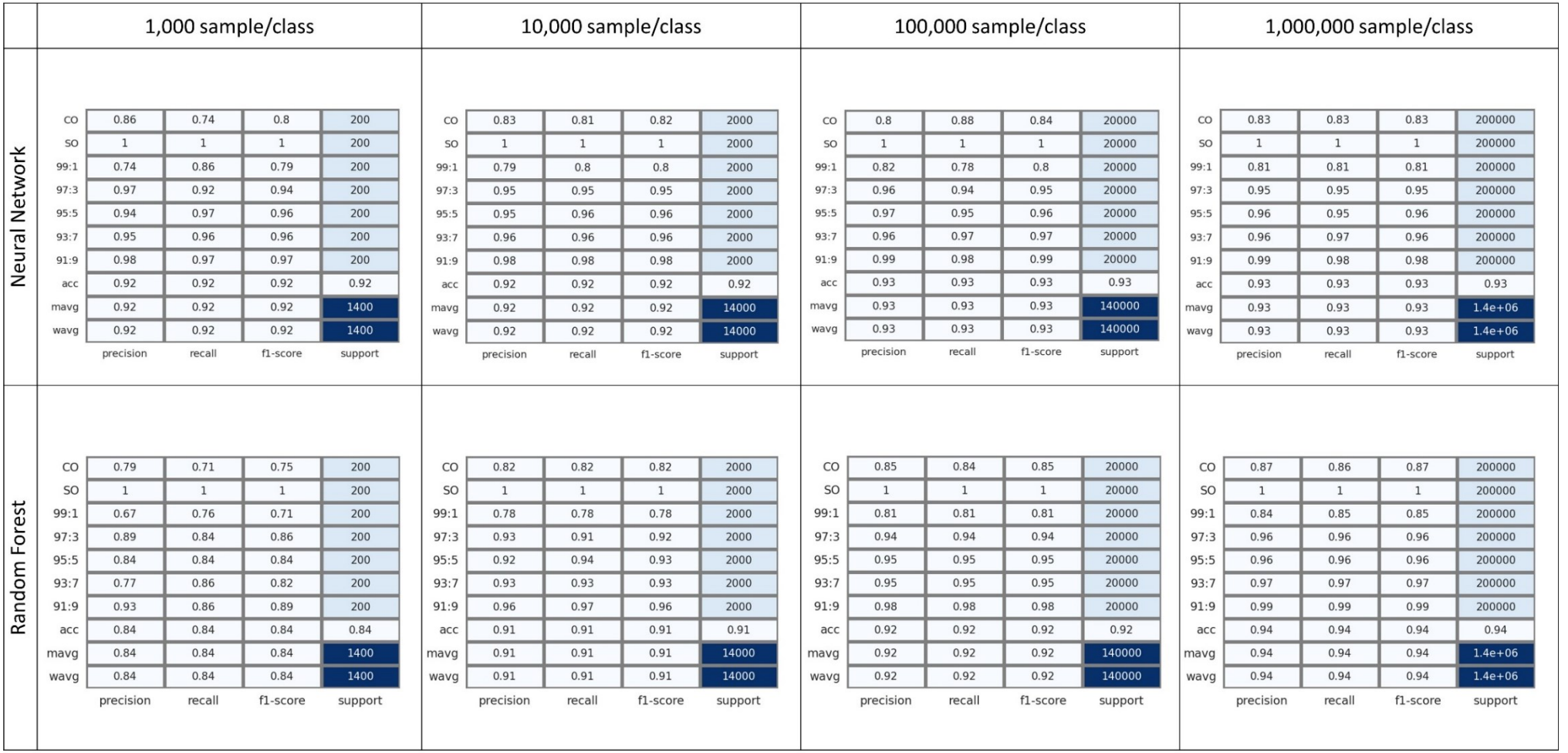
The following values were used in the grid for tuning Random Forest: max_depth (values from 2 to 24), Bootstrap (True or False), min_samples_split (values from 2 to 8), max_features (all features or sqrt (square root of number of features), min_samples_leaf (values from 3 to 9), n_estimators (values from 100 to 1500). Note: For 1,000,000 simulations, the number of n_estimators were fixed at 100 and was not determined by grid search.

Table S2

Hyperparameter set for Neural Network (NN) models trained on data sets of different sizes.

Model	Number of Hidden-Layers	Dimension for each hidden Layer	Batch_size
NN trained on 1,000 samples	1	Layer1: 91	16
NN trained on 10,000 samples	2	Layer1: 51 Layer2: 69	18
NN trained on 100,000 samples	2	Layer1: 51 Layer2: 69	18
NN trained on 1,000,000 samples	2	Layer1: 51 Layer2: 69	18

The following values were used for the tuning: Number of Hidden-Layers (values of 1 to 3), Dimension for each hidden Layer (values between 32 and 128), Batch size (values between 16 and 64). Note: For 100,000 and 1,000,000 simulations, complete tuning was not feasible due to hardware limitations. Thus, the hyperparameters optimized for 10,000 simulations were reused.



Neural Network	precision	recall	f1-score	support
CO	0.86	0.74	0.8	200
SO	1	1	1	200
99:1	0.74	0.86	0.79	200
97:3	0.97	0.92	0.94	200
95:5	0.94	0.97	0.96	200
93:7	0.95	0.96	0.96	200
91:9	0.98	0.97	0.97	200
acc	0.92	0.92	0.92	0.92
mavg	0.92	0.92	0.92	1400
wavg	0.92	0.92	0.92	1400

Neural Network	precision	recall	f1-score	support
CO	0.83	0.81	0.82	2000
SO	1	1	1	2000
99:1	0.79	0.8	0.8	2000
97:3	0.95	0.95	0.95	2000
95:5	0.95	0.96	0.96	2000
93:7	0.96	0.96	0.96	2000
91:9	0.98	0.98	0.98	2000
acc	0.92	0.92	0.92	0.92
mavg	0.92	0.92	0.92	14000
wavg	0.92	0.92	0.92	14000

Neural Network	precision	recall	f1-score	support
CO	0.8	0.88	0.84	20000
SO	1	1	1	20000
99:1	0.82	0.78	0.8	20000
97:3	0.96	0.94	0.95	20000
95:5	0.97	0.95	0.96	20000
93:7	0.96	0.97	0.97	20000
91:9	0.99	0.98	0.99	20000
acc	0.93	0.93	0.93	0.93
mavg	0.93	0.93	0.93	140000
wavg	0.93	0.93	0.93	140000

Neural Network	precision	recall	f1-score	support
CO	0.83	0.83	0.83	200000
SO	1	1	1	200000
99:1	0.81	0.81	0.81	200000
97:3	0.95	0.95	0.95	200000
95:5	0.96	0.95	0.96	200000
93:7	0.96	0.97	0.96	200000
91:9	0.99	0.98	0.98	200000
acc	0.93	0.93	0.93	0.93
mavg	0.93	0.93	0.93	1.4e+06
wavg	0.93	0.93	0.93	1.4e+06

Random Forest	precision	recall	f1-score	support
CO	0.79	0.71	0.75	200
SO	1	1	1	200
99:1	0.67	0.76	0.71	200
97:3	0.89	0.84	0.86	200
95:5	0.84	0.84	0.84	200
93:7	0.77	0.86	0.82	200
91:9	0.93	0.86	0.89	200
acc	0.84	0.84	0.84	0.84
mavg	0.84	0.84	0.84	1400
wavg	0.84	0.84	0.84	1400

Random Forest	precision	recall	f1-score	support
CO	0.82	0.82	0.82	2000
SO	1	1	1	2000
99:1	0.78	0.78	0.78	2000
97:3	0.93	0.91	0.92	2000
95:5	0.92	0.94	0.93	2000
93:7	0.93	0.93	0.93	2000
91:9	0.96	0.97	0.96	2000
acc	0.91	0.91	0.91	0.91
mavg	0.91	0.91	0.91	14000
wavg	0.91	0.91	0.91	14000

Random Forest	precision	recall	f1-score	support
CO	0.85	0.84	0.85	20000
SO	1	1	1	20000
99:1	0.81	0.81	0.81	20000
97:3	0.94	0.94	0.94	20000
95:5	0.95	0.95	0.95	20000
93:7	0.95	0.95	0.95	20000
91:9	0.98	0.98	0.98	20000
acc	0.92	0.92	0.92	0.92
mavg	0.92	0.92	0.92	140000
wavg	0.92	0.92	0.92	140000

Random Forest	precision	recall	f1-score	support
CO	0.87	0.86	0.87	200000
SO	1	1	1	200000
99:1	0.84	0.85	0.85	200000
97:3	0.96	0.96	0.96	200000
95:5	0.96	0.96	0.96	200000
93:7	0.97	0.97	0.97	200000
91:9	0.99	0.99	0.99	200000
acc	0.94	0.94	0.94	0.94
mavg	0.94	0.94	0.94	1.4e+06
wavg	0.94	0.94	0.94	1.4e+06

Fig. S3. Classification metrics on a class level for Random Forest and Neural Network using test sets of different simulation numbers. Accuracy (acc), Macro average (mavg), Weighted average (wavg).

Table S3
Number of real samples for each class.

Classes	Number of real samples
CO (100% Cactus oil)	27
SO (100% Sunflower oil)	10
99:1 (Mixture of 99% CO and 1% SO)	5
97:3 (Mixture of 97% CO and 3% SO)	7
95:5 (Mixture of 95% CO and 5% SO)	7
93:7 (Mixture of 93% CO and 7% SO)	7
91:9 (Mixture of 91% CO and 9% SO)	7

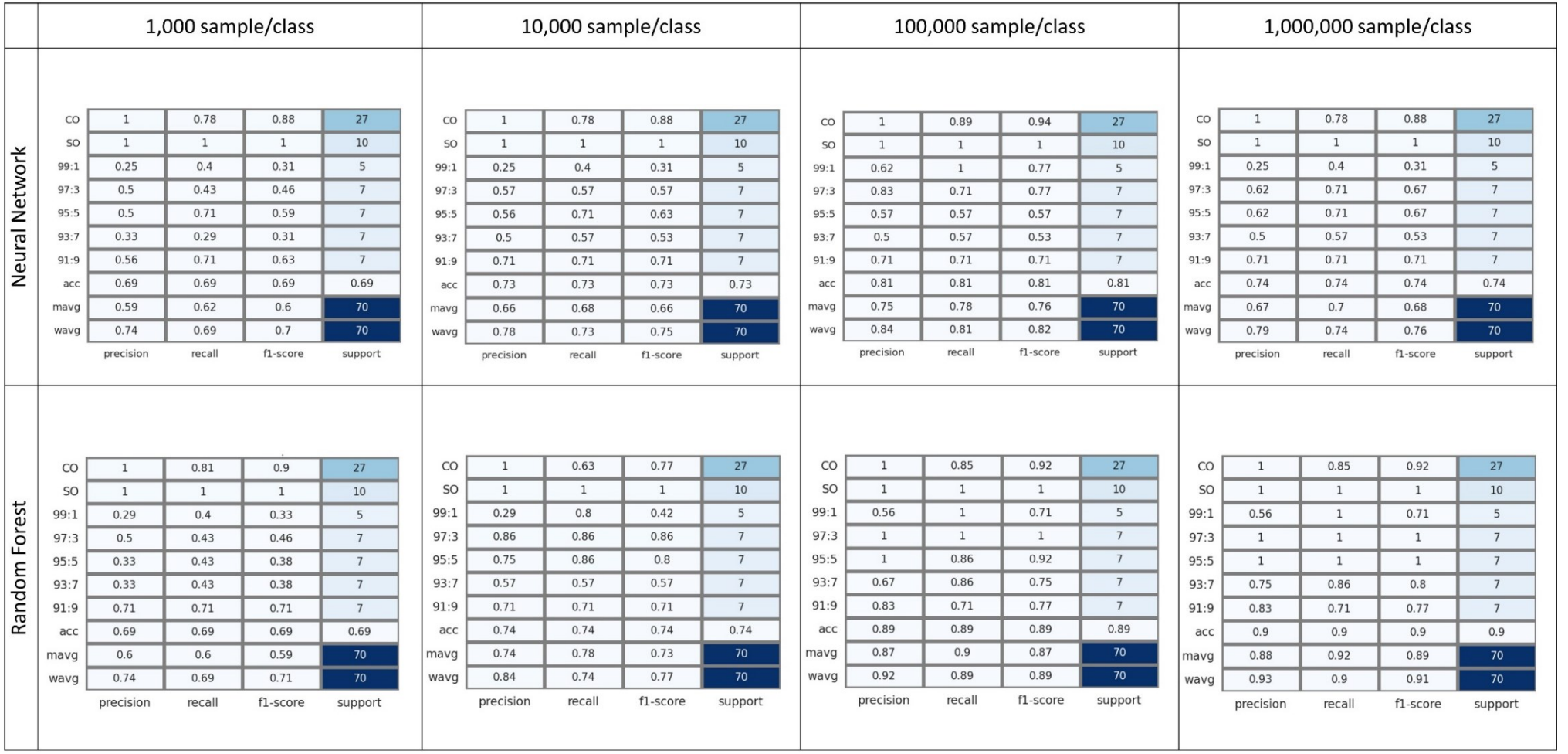


Fig. S4. Classification metrics on a class level for Random Forest and Neural Network tested on real-world samples after being trained on different numbers of simulated samples. Accuracy (acc), Macro average (mavg), Weighted average (wavg)

Supplementary data A: (Dataset, available in the journal website in Excel format)

Fatty acid composition of refined sunflower oil (mean values in %)

	C14:0	C16:0	C16:1Δ9	C17:0	C18:0	C18:1Δ9tr	C18:1Δ9	C18:1Δ11	C18:2Δ9tr,12tr	C18:2Δ9,12	C18:3Δ9,12,15	C20:0	C20:1Δ11	C20:2Δ11,14	C21:0	C20:3Δ11,14,17	C22:0	C22:1Δ13	C20:5Δ5,8,11, 14,17	C23:0	C24:0	C24:1Δ15
SO1	0.07	6.54	0.11	<LOQ	3.60	0.25	28.53	1.01	0.23	58.10	0.08	0.29	0.15	<LOQ	<LOQ	<LOQ	0.72	<LOQ	<LOQ	<LOQ	0.22	<LOQ
SO2	0.07	5.60	0.09	<LOQ	3.16	0.07	28.64	1.20	0.12	59.14	0.09	0.26	0.16	<LOQ	<LOQ	<LOQ	0.79	<LOQ	<LOQ	<LOQ	0.33	<LOQ
SO3	0.07	6.24	0.12	<LOQ	3.45	0.06	30.66	1.24	0.22	55.97	0.09	0.30	0.16	<LOQ	<LOQ	<LOQ	0.88	<LOQ	<LOQ	<LOQ	0.34	<LOQ
SO4	0.06	5.73	0.10	<LOQ	3.39	0.11	32.68	1.13	0.28	54.59	0.07	0.30	0.17	<LOQ	<LOQ	<LOQ	0.89	<LOQ	<LOQ	<LOQ	0.35	<LOQ
SO5	0.08	6.47	0.12	<LOQ	3.42	0.16	30.48	1.33	0.25	55.97	0.07	0.29	0.15	<LOQ	<LOQ	<LOQ	0.78	<LOQ	<LOQ	<LOQ	0.29	<LOQ
SO6	0.09	6.46	0.10	<LOQ	3.38	0.09	29.65	0.99	<LOQ	57.40	0.08	0.25	0.14	<LOQ	<LOQ	<LOQ	0.58	<LOQ	<LOQ	<LOQ	0.17	<LOQ
SO7	0.05	6.13	0.11	<LOQ	3.35	0.05	28.75	1.89	0.21	57.71	0.06	0.28	0.15	<LOQ	<LOQ	<LOQ	0.80	<LOQ	<LOQ	<LOQ	0.28	<LOQ
SO8	0.07	6.49	0.11	<LOQ	3.14	<LOQ	29.61	1.86	0.13	57.27	0.09	0.25	0.15	<LOQ	<LOQ	<LOQ	0.71	<LOQ	<LOQ	<LOQ	0.21	<LOQ
SO9	0.08	6.40	0.12	<LOQ	3.36	0.08	28.71	0.92	<LOQ	56.63	0.06	0.26	0.14	<LOQ	<LOQ	<LOQ	0.71	<LOQ	<LOQ	<LOQ	0.25	<LOQ
SO10	0.08	6.88	0.12	<LOQ	3.18	0.06	30.01	1.03	<LOQ	57.33	0.09	0.23	0.14	<LOQ	<LOQ	<LOQ	0.57	<LOQ	<LOQ	<LOQ	0.17	<LOQ

Fatty acid composition of refined sunflower oil (standard deviation)

	C14:0	C16:0	C16:1Δ9	C17:0	C18:0	C18:1Δ9tr	C18:1Δ9	C18:1Δ11	C18:2Δ9tr,12tr	C18:2Δ9,12	C18:3Δ9,12,15	C20:0	C20:1Δ11	C20:2Δ11,14	C21:0	C20:3Δ11,14,17	C22:0	C22:1Δ13	C20:5Δ5,8,11, 14,17	C23:0	C24:0	C24:1Δ15
SO1	0.07	6.54	0.11	<LOQ	3.60	0.25	28.53	1.01	0.23	58.10	0.08	0.29	0.15	<LOQ	<LOQ	<LOQ	0.72	<LOQ	<LOQ	<LOQ	0.22	<LOQ
SO2	0.07	5.60	0.09	<LOQ	3.16	0.07	28.64	1.20	0.12	59.14	0.09	0.26	0.16	<LOQ	<LOQ	<LOQ	0.79	<LOQ	<LOQ	<LOQ	0.33	<LOQ
SO3	0.07	6.24	0.12	<LOQ	3.45	0.06	30.66	1.24	0.22	55.97	0.09	0.30	0.16	<LOQ	<LOQ	<LOQ	0.88	<LOQ	<LOQ	<LOQ	0.34	<LOQ
SO4	0.06	5.73	0.10	<LOQ	3.39	0.11	32.68	1.13	0.28	54.59	0.07	0.30	0.17	<LOQ	<LOQ	<LOQ	0.89	<LOQ	<LOQ	<LOQ	0.35	<LOQ
SO5	0.08	6.47	0.12	<LOQ	3.42	0.16	30.48	1.33	0.25	55.97	0.07	0.29	0.15	<LOQ	<LOQ	<LOQ	0.78	<LOQ	<LOQ	<LOQ	0.29	<LOQ
SO6	0.09	6.46	0.10	<LOQ	3.38	0.09	29.65	0.99	<LOQ	57.40	0.08	0.25	0.14	<LOQ	<LOQ	<LOQ	0.58	<LOQ	<LOQ	<LOQ	0.17	<LOQ
SO7	0.05	6.13	0.11	<LOQ	3.35	0.05	28.75	1.89	0.21	57.71	0.06	0.28	0.15	<LOQ	<LOQ	<LOQ	0.80	<LOQ	<LOQ	<LOQ	0.28	<LOQ
SO8	0.07	6.49	0.11	<LOQ	3.14	<LOQ	29.61	1.86	0.13	57.27	0.09	0.25	0.15	<LOQ	<LOQ	<LOQ	0.71	<LOQ	<LOQ	<LOQ	0.21	<LOQ
SO9	0.08	6.40	0.12	<LOQ	3.36	0.08	28.71	0.92	<LOQ	56.63	0.06	0.26	0.14	<LOQ	<LOQ	<LOQ	0.71	<LOQ	<LOQ	<LOQ	0.25	<LOQ
SO10	0.08	6.88	0.12	<LOQ	3.18	0.06	30.01	1.03	<LOQ	57.33	0.09	0.23	0.14	<LOQ	<LOQ	<LOQ	0.57	<LOQ	<LOQ	<LOQ	0.17	<LOQ

Notes: LOQ= 0.05 %; The following fatty acids were analyzed but not detectable: C4:0, C6:0 ,C8:0, C10:0, C11:0, C12:0, C13:0, C14:1Δ9, C15:0, C15:1Δ10, C16:1Δ7, C17:1, C18:3Δ6,9,12, C20:3Δ8,11,14, C20:4Δ5,8,11,14, C22:2Δ13,16, C22:6Δ4,7,10,13,16,19

Triacylglycerol composition of refined sunflower oil (mean values in %)

	PaOlPa	PaLiPa	PaOlSt	PaOlOl	PaLiSt	PaLiOl	PaLiLi	StOlOl	OlOlOl	StLiOl	OlLiOl	LiLiOl	LiLiLi
SO1	0.52	1.63	0.43	3.24	1.41	9.08	12.48	1.16	7.49	3.08	12.01	26.65	20.82
SO2	0.39	1.30	0.34	2.86	1.19	8.59	11.17	0.97	6.89	2.89	13.53	29.48	20.39
SO3	0.55	1.54	0.47	3.79	1.39	9.40	11.38	1.34	8.58	3.17	13.63	26.61	18.15
SO4	0.45	1.25	0.39	3.81	1.20	8.12	10.18	1.87	11.89	3.46	13.38	25.82	18.18
SO5	0.54	1.52	0.45	3.74	1.37	9.39	11.36	1.31	8.46	3.14	13.72	26.78	18.20
SO6	0.47	1.45	0.41	3.38	1.31	8.66	11.71	1.31	8.96	2.99	12.48	26.64	20.23
SO7	0.52	1.55	0.44	3.39	1.36	9.46	11.77	1.15	7.11	3.11	13.62	27.47	19.05
SO8	0.47	1.51	0.38	3.43	1.27	8.94	11.74	1.14	8.73	2.86	12.68	27.14	19.68
SO9	0.52	1.56	0.44	3.35	1.39	9.48	11.79	1.12	6.94	3.13	13.60	27.59	19.10
SO10	0.46	1.47	0.38	3.37	1.27	8.77	11.48	1.32	8.91	3.17	12.81	27.08	19.50

Triacylglycerol composition of refined sunflower oil (standard deviation)

	PaOlPa	PaLiPa	PaOlSt	PaOlOl	PaLiSt	PaLiOl	PaLiLi	StOlOl	OlOlOl	StLiOl	OlLiOl	LiLiOl	LiLiLi
SO1	0.00	0.00	0.01	0.01	0.00	0.03	0.04	0.01	0.02	0.02	0.01	0.03	0.14
SO2	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.01	0.00	0.04	0.03	0.03	0.09
SO3	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.02	0.01	0.02	0.04	0.08
SO4	0.00	0.01	0.01	0.02	0.00	0.04	0.05	0.02	0.01	0.03	0.02	0.08	0.16
SO5	0.00	0.01	0.01	0.01	0.00	0.03	0.01	0.02	0.02	0.05	0.02	0.01	0.03
SO6	0.00	0.01	0.00	0.02	0.00	0.03	0.04	0.01	0.01	0.01	0.03	0.05	0.08
SO7	0.00	0.01	0.01	0.02	0.03	0.05	0.00	0.00	0.02	0.04	0.02	0.03	0.01
SO8	0.00	0.01	0.01	0.00	0.03	0.00	0.01	0.03	0.00	0.00	0.02	0.06	0.03
SO9	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.02	0.01	0.00
SO10	0.01	0.02	0.00	0.08	0.04	0.18	0.23	0.23	0.22	0.38	0.12	0.13	0.27

Notes: LOQ= 0.05 %; The following triacylglecerol were analyzed but not detectable: StOlSt

Tocochromanol composition of refined sunflower oil (mean values in %)

	α-Tocopherol	β-Tocopherol	γ-Tocopherol	Plastochromanol-8	γ-Tocotrienol
SO1	96.59	2.24	0.43	0.63	<LOQ
SO2	96.23	2.26	0.88	0.52	<LOQ
SO3	96.49	2.06	0.72	0.57	<LOQ
SO4	96.68	2.18	0.45	0.55	<LOQ
SO5	96.79	2.06	0.43	0.61	<LOQ
SO6	96.47	2.26	0.49	0.65	<LOQ
SO7	96.98	2.13	0.28	0.53	<LOQ
SO8	96.12	2.28	0.84	0.61	<LOQ
SO9	96.99	2.22	<LOQ	0.63	<LOQ
SO10	95.98	2.35	0.89	0.64	<LOQ

Tocochromanol composition of refined sunflower oil (standard deviation)

	α-Tocopherol	β-Tocopherol	γ-Tocopherol	Plastochromanol-8	γ-Tocotrienol
SO1	0.05	0.05	0.07	0.09	-
SO2	0.11	0.01	0.06	0.06	-
SO3	0.18	0.03	0.06	0.07	-
SO4	0.03	0.00	0.00	0.00	-
SO5	0.16	0.01	0.04	0.08	-
SO6	0.12	0.02	0.08	0.01	-
SO7	0.07	0.02	0.03	0.02	-
SO8	0.03	0.00	0.00	0.02	-
SO9	0.06	0.02	-	0.02	-
SO10	0.07	0.01	0.01	0.03	-

Notes: LOQ= 2mg/kg oil; The following tocochromanol were analyzed but not detectable: α-Tocotrienol, β-Tocotrienol, δ-Tocopherol, δ-Tocotrienol