

<https://doi.org/10.1038/s41538-025-00555-z>

# Reassessing fatty acid divergence: climatic and geographic constraints on origin traceability in global oil crops

Check for updates

Junyu Xu<sup>1,2,5</sup>, Yiqing Luo<sup>1,2,3,4,5</sup>, Lei Hu<sup>1,2,5</sup>, Kai Chen<sup>1,2,5</sup>, Shuang Cheng<sup>1</sup>, Mengting Xiao<sup>1,2</sup>, Haijing Xiao<sup>1,2,3</sup>, Shumin Hou<sup>1,2,3</sup>, Jiaxin Yang<sup>1,2</sup>, Jiao Yang<sup>1,3</sup>, Ying Wang<sup>4</sup>, Shaohua Zeng<sup>4</sup>✉ & Chunsong Cheng<sup>1,2,3</sup>✉

Global agricultural trade requires reliable tools to trace product origins and combat fraud. We propose two novel metrics—the Geographical Differentiation Index (GDI) and Environmental Heritability Index (EHI)—to quantify spatial variation in fatty acids and their environmental drivers. We systematically investigated the fatty acid profiles of four main oil-rich crops (olive, camellia, walnut, and peony seed) and revealed that fatty acid distributions follow elevation- and latitude-dependent patterns, with peony seed oils showing the strongest latitudinal sensitivity. Key fatty acids like stearic acid (C18:0) and linoleic acid (C18:2) correlated significantly with geographic factors globally, while the biomass of certain specific fatty acids varies significantly in high-altitude/low-latitude regions. These findings establish specific fatty acid signatures as a robust tool for geographic authentication. Our approach provides a chemically grounded framework for precision origin discrimination in global oil food systems.

Global agricultural trade has expanded markedly in recent decades, driven by shifting market dynamics and increasing interdependencies between regions<sup>1–3</sup>. Concurrently, consumer awareness of food safety and quality has intensified, with growing demand for transparent provenance information and compliance with regulatory standards<sup>4–6</sup>. This trend is exemplified by Geographical Indication (GI) labels, which hinge on verifiable traceability systems to prevent fraud and uphold product authenticity. Such systems are now critical for maintaining trust in global food supply chains, particularly for high-value crops<sup>7</sup>. Consequently, developing robust methodologies for crop origin verification has emerged as a pivotal research frontier in food science. GI agricultural products derive their unique quality characteristics from specific production regions, gaining both economic value and market competitiveness through this origin-linked identity. The European Union's stringent regulations (EC 510/2006) underscore how GI certification systems protect these products by verifying their geographical authenticity. Effective traceability systems are thus essential to safeguard product quality and reputation, requiring comprehensive monitoring from production through processing to final marketing<sup>7,8</sup>.

Current analytical techniques for origin verification, including infrared spectroscopy, Raman spectroscopy, fluorescence spectroscopy, and

ultraviolet visible spectroscopy, each present limitations, such as large sample requirements and variable accuracy<sup>9–11</sup>. While chemometric analysis of fatty acid composition offers a rapid approach for authenticating products like Tunisian olive oil, its dependency on extensive training datasets restricts broad applicability<sup>12</sup>. Similarly, methods integrating mineral elements and stable isotopes with machine learning, named convolutional neural networks, showed promise but face challenges in scaling across diverse geographical regions<sup>13</sup>. These constraints highlight an urgent need for innovative, precise, and scalable traceability technologies. Developing such tools is critical to advancing agricultural quality assurance and meeting the growing demand for transparent, reliable food supply chains. Fatty acids, fundamental structural and metabolic components of plants, exhibit dynamic compositional profiles shaped by the synergistic interplay of endogenous biosynthesis and exogenous environmental drivers. The *de novo* synthesis of fatty acids is governed by tightly regulated enzymatic cascades, notably involving acetyl-CoA carboxylase (ACCase) for the initial carboxylation of acetyl-CoA and fatty acid synthase (FAS) complexes for subsequent chain elongation<sup>14–18</sup>. Climatic variables — including precipitation seasonality, diurnal thermal amplitude, and altitudinal gradients — serve as critical modulators of these metabolic processes by altering photosynthetic carbon flux and enzyme

<sup>1</sup>College of Life Sciences in Nanchang University, Joint Plan of Postgraduate education with Jiangxi Key Laboratory for Sustainable Utilization of Chinese Material Medical Resources, Lushan Botanical Garden, Chinese Academy of Science, Jiujiang, Jiangxi, China. <sup>2</sup>Lushan Xinglin Institute for Medicinal Plants, Jiujiang Xinglin Key Laboratory for Traditional Chinese Medicines, Jiujiang, Jiangxi, China. <sup>3</sup>De'an Institute of Economic Plants, Lushan Botanical Garden, Chinese Academy of Science, Jiujiang, Jiangxi, China. <sup>4</sup>South China Botanical Garden, Chinese Academy of Science, Guangzhou, Guangdong Province, China. <sup>5</sup>These authors contributed equally: Junyu Xu, Yiqing Luo, Lei Hu, Kai Chen. ✉e-mail: [shhzeng@scbg.ac.cn](mailto:shhzeng@scbg.ac.cn); [chengcs@lsbg.cn](mailto:chengcs@lsbg.cn)

kinetics<sup>19</sup>. Concurrently, edaphic factors exert profound influence: rhizospheric microbial consortia participate in precursor synthesis through plant-microbe metabolic coupling<sup>20</sup>, while bioavailable micronutrients (e.g., Zn<sup>2+</sup> and Mn<sup>2+</sup>) function as essential cofactors for desaturases and elongases during fatty acid modification<sup>21,22</sup>. This multidimensional regulation ultimately dictates the nutritional and functional quality of plant-derived fatty acids, with significant implications for food authentication and GI certification systems.

Fatty acid profiles serve as potent biochemical markers for crop origin authentication, as their composition is intrinsically linked to both geographical origin and environmental growing conditions<sup>23</sup>. Advanced analytical techniques, including chromatographic separation and mass spectrometry, have enabled systematic extraction of origin-specific fatty acid signatures<sup>24</sup>. These approaches have proven particularly effective for three key applications: (I) quality grade classification, (II) adulteration detection, and (III) geographic traceability, as demonstrated by established protocols for olive oil authentication<sup>25–27</sup>. Current methodologies combine gas chromatography-mass spectrometry (GC-MS) for precise fatty acid quantification with multivariate statistical analysis (e.g., PCA, HCA) to develop robust origin discrimination models<sup>22,25,28</sup>. While research on oil-rich crops like olive and camellia has reached relative maturity, most studies remain constrained to single-product systems<sup>29–31</sup>. This leaves a significant knowledge gap regarding comprehensive traceability frameworks capable of handling diverse agricultural commodities<sup>32–34</sup>.

In particular, current knowledge failed to accurately describe the applicable geographical scale or the degree of geographical environmental differences for specific fatty acids or fatty acid combinations when detecting the origin of agricultural products. This study comprehensively synthesizes fatty acid data from prior research on oil-rich crops to systematically investigate their associations with environmental factors. The study employed an expanded sample collection strategy to ensure adequate data representation. The primary aim is to elucidate the regulatory influence of environmental variables—including climatic conditions and elevational gradients—on fatty acid compositional profiles. This study proposes two novel concepts: Geographical Differentiation Index (GDI) and Environmental Heritability Index (EHI). EHI effectively quantifies the relative contributions of extrinsic environmental conditions and intrinsic variations to plant fatty acid composition, while GDI efficiently identifies fatty acids exhibiting the most significant differences across geographical origins. First, spatial overlay techniques from Geographic Information Systems (GIS) were employed to delineate regions based on the fatty acid composition of different geographical origins and their corresponding environmental attributes. Subsequently, the EHI evaluation framework was applied to validate the rationality of these regional divisions. Using the GDI parameter, potential fatty acid geographic markers (C16:0, C18:0, C18:1, C18:2, C18:3) for oil-rich crops were identified, and their correlations with environmental parameters (particularly latitude and altitude) were evaluated through regression analysis (Origin 2021). These findings establish a theoretical framework for global origin tracing of oil-rich crops, offering valuable insights for agricultural traceability and quality assessment.

## Results

### Data collection and categorization

The four main crops we focused on were categorized into two primary groups based on their agricultural product characteristics: edible oils (including olive oil and camellia oil) and nuts/seeds (comprising walnuts and peony seeds). Data were collected from articles that investigated these crops across multiple continents, specifically South America, Europe, Africa, and Asia. Additionally, all samples were classified into different groups in this study based on latitude and specific environmental parameters, following the principle of environment and climate similarities. Table 1 presents the geostatistical details of data collection sources for the main oil-rich crops.

**Table 1 | Oil-rich crops data collection sources, geostatistics (The letters A–G represent the anthropogenic geographic origin of the crop in the table)**

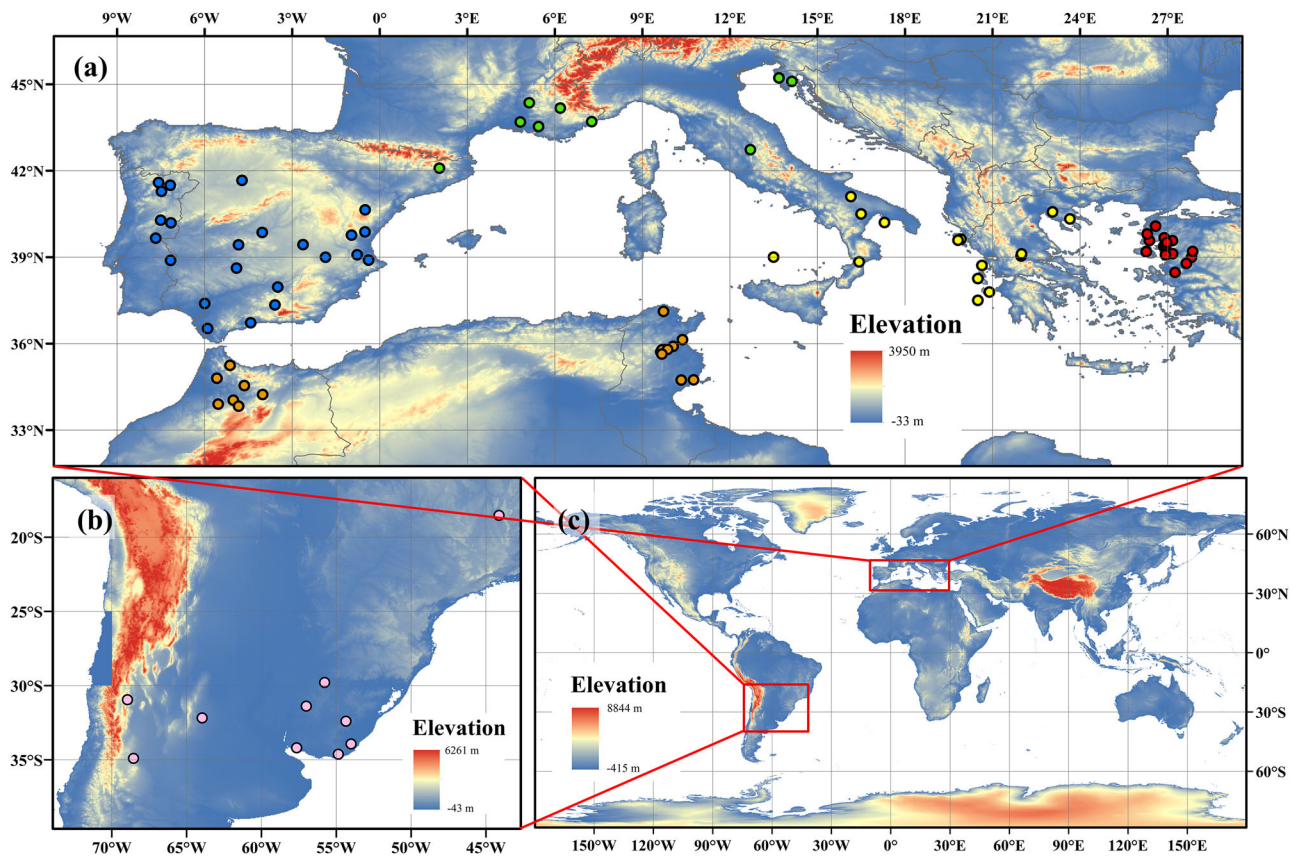
| Types of agricultural products | Regional grouping | Geographic information |               |              |
|--------------------------------|-------------------|------------------------|---------------|--------------|
|                                |                   | Latitude (degree)      | Elevation (m) | Climate zone |
| Olive oil                      | A                 | 38~41                  | 150.6         | MC           |
|                                | B                 | 33~38                  | 351.29        | MC           |
|                                | C                 | 35~42                  | 357.93        | MC           |
|                                | D                 | 42~46                  | 380.87        | MC           |
|                                | E                 | –35~–19                | 429.5         | TCC SHC      |
|                                | F                 | 36~42                  | 490.91        | TCC MC       |
| Camellia oil                   | A                 | 21~24                  | 77.29         | SMC          |
|                                | B                 | 25~33                  | 194.17        | TCC          |
|                                | C                 | 18~20                  | 278.2         | TRMC         |
|                                | D                 | 29~34                  | 334.04        | TCC SMC      |
|                                | E                 | 25~31                  | 375.38        | TCC          |
|                                | F                 | 24~31                  | 458.31        | TCC          |
|                                | G                 | 23~28                  | 1100.88       | TCC          |
| Walnuts                        | A                 | 36~39                  | 759.18        | TCC TEMC     |
|                                | B                 | 30~37                  | 853.08        | TCC SMC      |
|                                | C                 | 37~44                  | 1206          | TCC          |
|                                | D                 | 23~29                  | 1981.52       | SMC          |
| Peony seeds                    | A                 | 33~36                  | 170.1         | TCC TEMC SMC |
|                                | B                 | 27~30                  | 221           | SMC          |
|                                | C                 | 36~43                  | 1220.25       | TCC TEMC     |
|                                | D                 | 26~28                  | 2696.56       | SMC          |

'TCC' stands for 'Temperate continental climate', 'MC' stands for 'Mediterranean climate', 'SHC' stands for 'Subtropical humid climate', 'SMC' stands for 'Subtropical monsoon climate', 'TRMC' stands for 'Tropical monsoon climate', 'TEMC' for 'Temperate monsoon climate' see Supplementary Table 4 for details).

As shown in Fig. 1a, olive oils were classified into 6 groups, with Groups A, C, D, E, and F distributed around the Mediterranean Sea in different coastal directions. As shown in Fig. 1b, Group B was unconventional located in South America. Figure 1c is the olive data collection sourced for olive oil on the world map.

Figure 2a shows the camellia oil, walnuts and peony seeds data collection sources on the map of China. As shown in Fig. 2d, camellia oils were categorized into 7 groups distributed across southern China from Hainan Island (south) to Gansu Province (north), spanning a latitudinal range from 18° to 34°. As shown in Fig. 2c, walnuts were geographically classified into 4 groups, predominantly located in central China including Yunnan and Xinjiang provinces, latitudes range from 29° to 39° and the climate is temperate continental. As shown in Fig. 2b, peony seeds were geographically grouped into 4 groups, primarily located in Chinese mountainous areas, spanning latitudes from 26° to 36° across temperate continental, temperate monsoon, and subtropical monsoon zones. Among all the above classified groups, the maximum latitudinal difference intra-group was 77°, while the maximum latitudinal difference inter-group was 78°.

To validate the scientific basis and effectiveness of the geographical classification framework, this study developed an innovative EHI system. This index is derived from classical genetic theory: traditional heritability quantifies the contribution of genes to traits by calculating the ratio of additive genetic variance to phenotypic variance<sup>35</sup>. The EHI was formulated to assess the efficacy of geographic partitioning through spatial



**Fig. 1 | Geographic map of data collection sourced for olive oil.** Different colour of the data points represent different group. Red represents appellation A, orange represents appellation B, yellow represents appellation C, green represents appellation D, blue represents appellation E, pink represents appellation F. **a** Olive oil

samples sourced from the Mediterranean periphery; **b** olive oil samples sourced from South America; **c** Globally olive oil samples collected for geographical marker analysis using specific fatty acids.

heterogeneity analysis.

$$EHI = Var_w / Var_g \quad (1)$$

$$Var_w = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (2)$$

$$Var_g = \frac{\sum_{i=1}^n (x_i - \bar{u})^2}{N} \quad (3)$$

$Var_w$  is the within-group variance,  $Var_g$  is the global variance.  $n$  is the number of samples in the group,  $x_i$  is the  $i$ th sample value,  $\bar{x}$  is the sample mean.  $N$  is the total number of samples,  $x_i$  is the  $i$ th sample value,  $\bar{u}$  is the overall mean.

According to the EHI threshold law, an  $EHI > 1$  indicated that the variability within geographically group data surpassed the variability of the entire dataset (i.e., internal influencing factors dominated over external geo-environmental controls). Conversely, an  $EHI < 1$  represented the inverse relationship. The corresponding statistical outputs are summarized in Table 2.

A total of 12 distinct fatty acids were identified in olive oil across six origins. The cumulative proportion of EHI less than 1 was 81.1%. Camellia oil exhibited 12 distinct fatty acids across seven origins, and the total EHI less than 1 accounted for 65.9%. Walnut contained 6 distinct fatty acids with variability across four origins, and the cumulative proportion of EHI less than 1 was 65.6%. Peony seed oil had 6 distinct fatty acids with variability across four origins, and the total EHI less than 1 was 79.2%.

The EHI predominantly below 1 (81.1% for olive oil, 65.9% for camellia oil, 65.6% for walnuts, and 79.2% for peony seeds) indicates reduced environmental variation within groups compared to the overall population, demonstrating that the artificially defined geographic origins effectively represent distinct geoclimatic characteristics.

#### Differential fatty acid analysis of different origins and groups

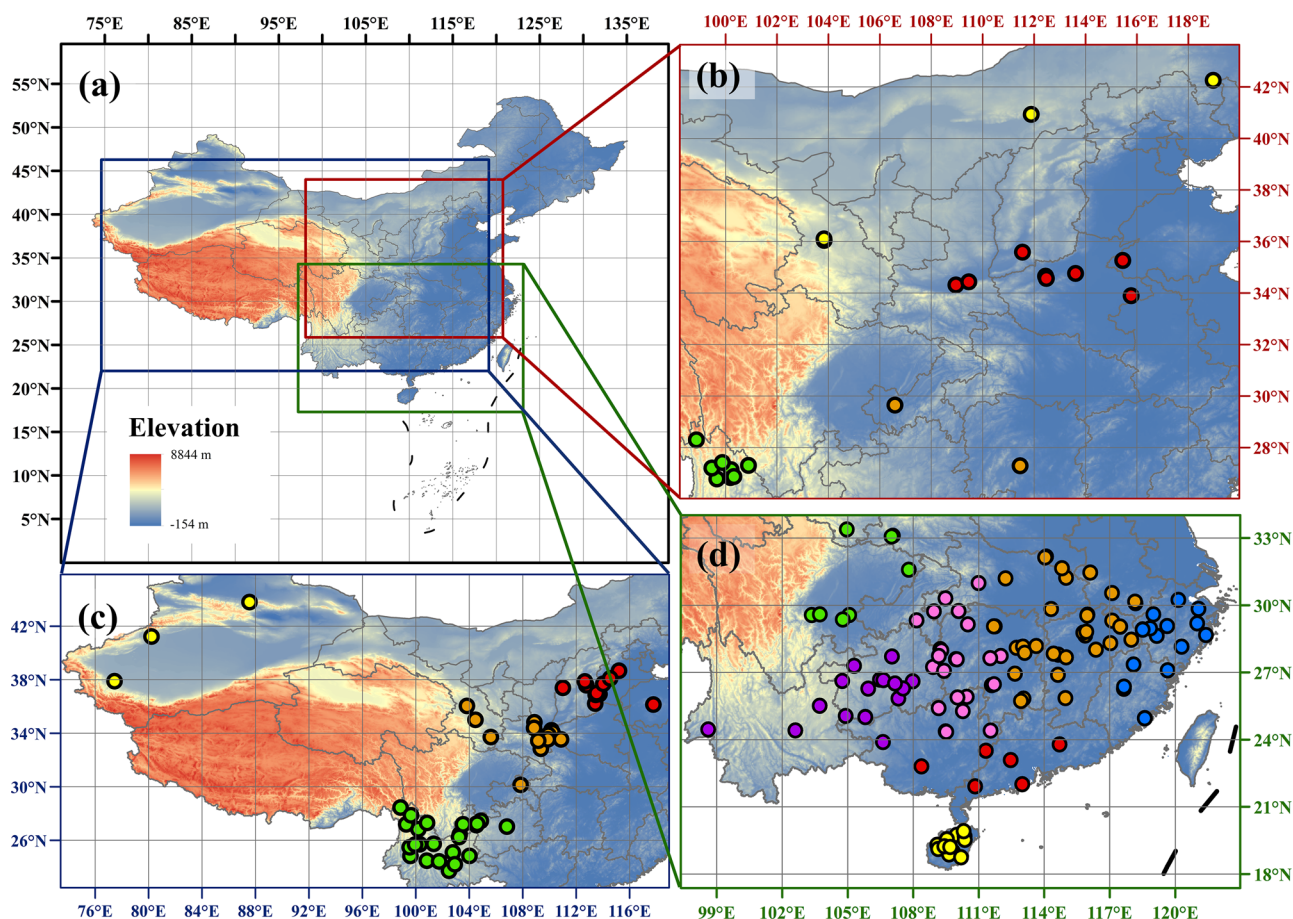
Based on the framework of geographical traceability research, this study further employed the differential fatty acid screening model to analyze the four types of crops, namely olive oil, camellia oil, walnuts, and peony seeds, and conduct a multi-dimensional analysis. The response of fatty acids to different geographical origins was quantified by constructing a GDI.

$$GDI = N_s / N_t \quad (4)$$

$N_s$  is the number of significant differences ( $P < 0.05$ ) in fatty acid profiles between a given origin and all other origins,  $N_t$  is the total number of origin.

As shown in Table 3, the variability degree represents the sensitivity of fatty acid distribution to geographical environments. The potential origin discrimination markers can be preliminarily screened by calculating the GDI values of different production areas of oil-rich crops and ranking them. First, calculate the total GDI value of each production area of SFA and UFA, respectively and rank them. Fatty acids with total GDI values greater than or equal to 1.5 in SFA and UFA were respectively selected as potential origin identification markers for the subsequent analysis of this agricultural product (if there are multiple fatty acids with a GDI index greater than or equal to 1.5, the first two should be selected for analysis). If the agricultural product's total GDI value was less than 1.5, select the fatty acid with the highest





**Fig. 2 | Geographic map of data collection sourced for camellia oil, walnuts and peony seeds.** Different colours of the data points represent different groupings. Red represents appellation A, orange represents appellation B, yellow represents appellation C, green represents appellation D, blue represents appellation E, pink represents appellation F, and purple represents appellation G. **a** a geographical presentation map of sample collection based on China's map to clarify the origin of all test sample data;

**b** peony seeds were geographically located in Chinese mountainous areas, spanning latitudes from 26° to 36° across temperate continental, temperate monsoon, and subtropical monsoon zones; **c** walnuts were predominantly located in central China including Yunnan and Xinjiang provinces; **d** Camellia oil samples sourced from the China from Hainan Island (south) to Gansu Province (north).

GDI value as a potential geographical discrimination marker. The saturated fatty acids in olive oil with a total GDI value greater than or equal to 1.5 are C18:0 (3.33) and C17:0 (2.98) (The GDI of C16:0 (2.01), C17:0(2.98) and C18:0 (3.33) of saturated fatty acids in olive oil are all greater than 1.5, then the top two with the largest values, namely 17:0 and C18:0, are selected). Among the unsaturated fatty acids, those with a total GDI value greater than or equal to 1.5 are C18:1 (1.67) and C18:2 (2.67). Of the saturated fatty acids in camellia oil, those with a total GDI value greater than or equal to 1.5 are C18:0 (2). No unsaturated fatty acid has a total GDI value greater than or equal to 1.5. Therefore, the unsaturated fatty acids with the highest total GDI values are selected as potential origin markers: C18:3 (1.42) and C20:1 (1.42). No fatty acid among the saturated fatty acids in walnuts has a total GDI value greater than or equal to 1.5. Consequently, the unsaturated fatty acid with the highest total GDI value, C18:0 (0.75), is selected as a potential origin marker. Among the unsaturated fatty acids, those with a total GDI value greater than or equal to 1.5 is C18:3 (1.5). Among the saturated fatty acids in peony seeds, those with a total GDI value greater than or equal to 1.5 are C16:0 (1.5) and C18:0 (1.5). Among the unsaturated fatty acids, those with a total GDI value greater than or equal to 1.5 are C18:1 (1.5) and C18:2 (2.5).

Whether a fatty acid can serve as a potential origin marker depends on its significance for origin discrimination and its identifiable, easily detectable characteristics. To this end, Tables S2 and S3 have been prepared, respectively showing the saturated and unsaturated fatty acid content of oil-rich crops to facilitate further screening of fatty acids (Supplementary Table 1 is a

statistical table of the saturated fatty acid content of oil-rich crops, and Supplementary Table 2 is a statistical table of the unsaturated fatty acid content of oil-rich crops). The basis for screening is whether the fatty acid ranks among the top three in terms of the percentage content of saturated or unsaturated fatty acids.

Cross-screening was conducted based on the preliminary screening principles of GDI and the ranking screening method of saturated fatty acid percentage content. The following qualified fatty acids were obtained. In olive oil, the percentage content of C18:0 ranks second (C18:0: 1.1–3.734%). In camellia oil, the percentage content of C18:0 ranks second (C18:0: 1.38–6.316%). Among walnuts, the content of C18:0 also ranks second (C18:0: 1.48–6.434%). Among peony seeds, the saturated fatty acid contents of C16:0 and C18:0 rank first and second, respectively (C16:0: 3.91–7.72%; C18:0: 0.74–2.87%).

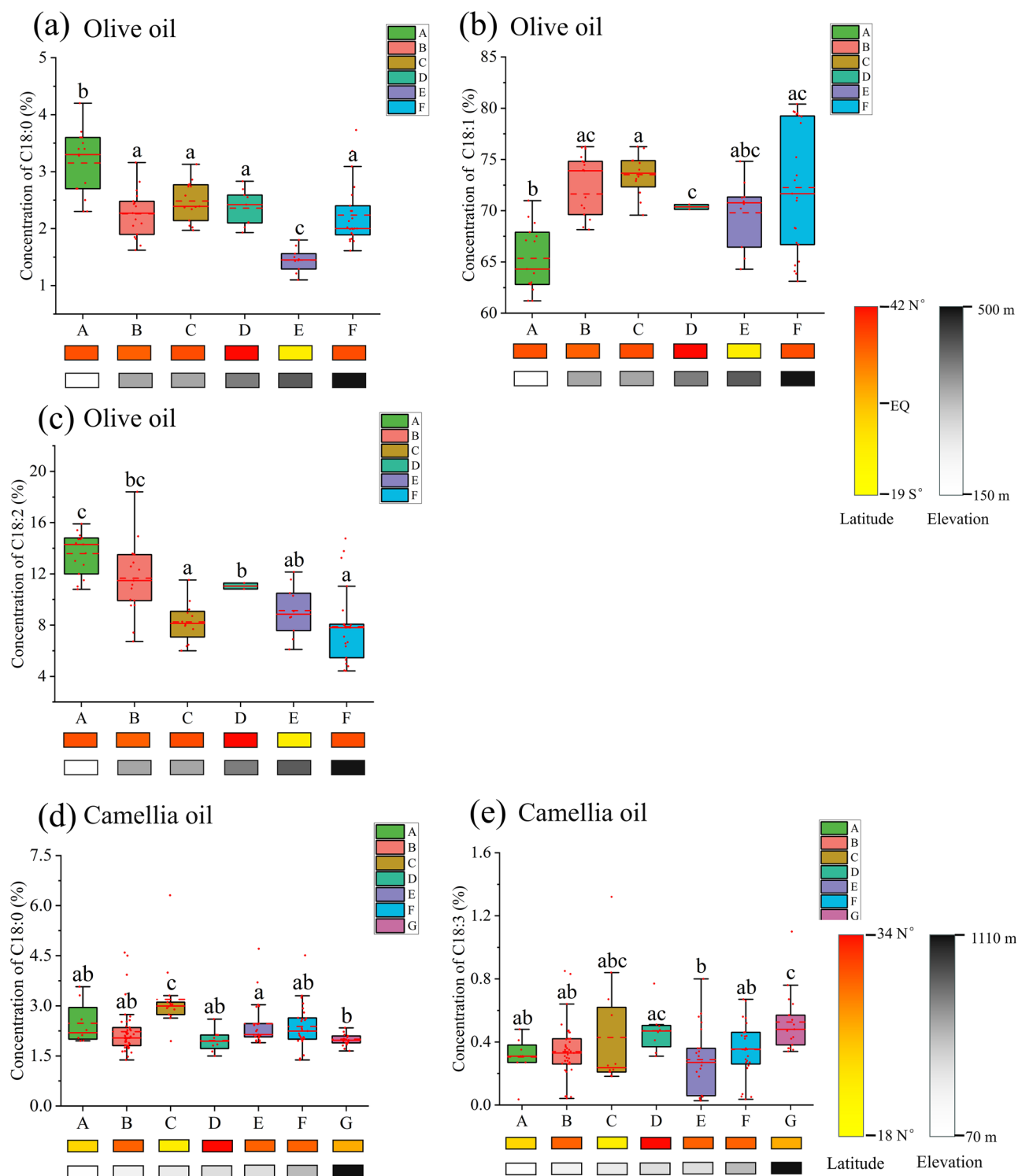
Unsaturated fatty acids are also screened in accordance with the screening methods for potential origin markers in saturated fatty acids. The results obtained are as follows: In olive oil, the percentage contents of C18:1 and C18:2 rank first and second, respectively. (C18:1: 54.8–79.62%; C18:2: 4.42–18.4%). In camellia oil, the percentage content of C18:3 ranks third (C18:0: 0.028–1.32%). *t* ranks third in terms of C18:3 content among walnuts (C18:3: 1.32–16.8%). Among peony seeds, the percentage contents of C18:1 and C18:2 rank first and second, respectively. (C18:1: 20.88–35.94%; C18:2: 13.64–28.67%). It is worth noting that after the screening of fatty acid percentage content, C17:0 in olive oil was eliminated because it excluded the top three in the ranking of saturated fatty acid percentage content (C17:0:



Table 3 | Geographical Differentiation Index (GDI) of oil-rich crops

| Types of agricultural products | Fatty acid               |       |            |             |             |       |       |                            |      |       |       |             |             |             |             |      |      |
|--------------------------------|--------------------------|-------|------------|-------------|-------------|-------|-------|----------------------------|------|-------|-------|-------------|-------------|-------------|-------------|------|------|
|                                | Saturated fatty acid (%) |       |            |             |             |       |       | Unsaturated fatty acid (%) |      |       |       |             |             |             |             |      |      |
|                                | Regional grouping        | C14:0 | C16:0      | C17:0       | C18:0       | C20:0 | C22:0 | C24:0                      | SFA  | C16:1 | C17:1 | C18:1       | C18:2       | C18:3       | C20:1       | PUFA | UFA  |
| Olive oil                      | A                        | —     | 0.17       | 0.17        | 0.83        | 0.67  | 0     | 0.33                       | 0    | 0.33  | 0.17  | 0.5         | 0.33        | 0.5         | 0.67        | 0    | —    |
|                                | B                        | —     | 0.17       | 0.5         | 0.33        | 0.5   | 0.33  | 0                          | 0.33 | 0.17  | 0.5   | 0.17        | 0.33        | 0.5         | 0.67        | 0    | —    |
|                                | C                        | —     | 0.17       | 0.67        | 0.33        | 0.5   | 0     | 0                          | 0    | 0     | 0     | 0.33        | 0.5         | 0           | 0.17        | 0    | —    |
|                                | D                        | —     | 0.17       | 0.17        | 0.33        | 0.33  | 0.33  | 0.17                       | 0    | 0     | 0     | 0.33        | 0.5         | 0           | 0           | 0    | —    |
|                                | E                        | —     | 0.67       | 0.17        | 0.83        | 0.83  | 0.33  | 0.17                       | 0.17 | 0     | 0.17  | 0           | 0.17        | 0.17        | 0.17        | 0    | —    |
|                                | F                        | —     | 0          | 0.33        | 0.33        | 0.5   | 0.33  | 0                          | 0.17 | 0.67  | 0.33  | 0.17        | 0.5         | 0.17        | 0.17        | 0    | —    |
|                                | total                    | 1.35  | 2.01       | <b>2.98</b> | <b>3.33</b> | 1.32  | 0.67  | 0.67                       | 1.35 | 1.17  | 1.33  | <b>1.67</b> | <b>2.67</b> | 1.01        | 1.35        | 0    | —    |
| Camellia oil                   | A                        | 0.29  | 0          | 0           | 0.14        | 0     | —     | —                          | 0    | 0     | 0.14  | 0           | 0           | 0.14        | 0.14        | —    | 0.43 |
|                                | B                        | 0     | 0          | 0           | 0.14        | 0     | —     | —                          | 0.29 | 0.14  | 0.14  | 0           | 0           | 0.14        | 0.14        | —    | 0.14 |
|                                | C                        | 0     | 0.29       | 0           | 0.86        | 0     | —     | —                          | 0.57 | 0.43  | 0.14  | 0.43        | 0           | 0           | 0           | —    | 0.71 |
|                                | D                        | 0.14  | 0.14       | 0           | 0.14        | 0.14  | —     | —                          | 0.14 | 0     | 0.57  | 0           | 0           | 0.14        | 0.14        | —    | 0.29 |
|                                | E                        | 0     | 0          | 0           | 0.29        | 0.29  | —     | —                          | 0.29 | 0.14  | 0.14  | 0.14        | 0           | 0.29        | 0.29        | —    | 0.29 |
|                                | F                        | 0.14  | 0          | 0           | 0.14        | 0     | —     | —                          | 0.43 | 0.14  | 0     | 0.14        | 0           | 0.14        | 0.14        | —    | 0.14 |
|                                | G                        | 0     | 0.14       | 0           | 0.29        | 0.14  | —     | —                          | 0.29 | 0     | 0     | 0.14        | 0           | 0.57        | 0.57        | —    | 0.29 |
|                                | total                    | 0.57  | 0.57       | 0           | <b>2</b>    | 0.57  | 0     | 0                          | 2.01 | 0.85  | 1.13  | 0.85        | 0           | <b>1.42</b> | <b>1.42</b> | 0    | 2.29 |
| Walnuts                        | A                        | —     | 0          | —           | 0.25        | —     | —     | —                          | 0    | 0     | —     | 0           | 0           | 0.25        | 0.25        | 0.25 | —    |
|                                | B                        | —     | 0          | —           | 0.25        | —     | —     | —                          | 0    | 0.25  | —     | 0           | 0           | 0.5         | 0.5         | 0    | —    |
|                                | C                        | —     | 0          | —           | 0.25        | —     | —     | —                          | 0    | 0.25  | —     | 0           | 0           | 0.25        | 0.25        | 0    | —    |
|                                | D                        | —     | 0          | —           | 0           | —     | —     | —                          | 0    | 0     | —     | 0           | 0           | 0.5         | 0.5         | 0.25 | —    |
|                                | total                    | 0     | 0          | 0           | <b>0.75</b> | 0     | 0     | 0                          | 0    | 0.5   | 0     | 0           | 0           | <b>1.5</b>  | —           | 0.5  | —    |
|                                | A                        | —     | 0.25       | —           | 0.25        | 0.25  | —     | —                          | —    | 0.25  | —     | 0.25        | 0.5         | —           | —           | —    | —    |
|                                | B                        | —     | 0.25       | —           | 0.25        | 0     | —     | —                          | —    | 0     | —     | 0.25        | 0.5         | —           | —           | —    | —    |
| Peony seeds                    | C                        | —     | 0.25       | —           | 0.25        | 0.5   | —     | —                          | —    | 0.5   | —     | 0.25        | 0.75        | —           | —           | —    | —    |
|                                | D                        | —     | 0.75       | —           | 0.75        | 0.25  | —     | —                          | —    | 0.25  | —     | 0.75        | 0.75        | —           | —           | —    | —    |
|                                | total                    | 0     | <b>1.5</b> | —           | <b>1.5</b>  | 1     | —     | —                          | —    | 1     | —     | <b>1.5</b>  | <b>2.5</b>  | —           | —           | —    | —    |

(The "total" column is bolded to indicate: the GDI of this fatty acid is greater than or equal to 1.5, or the GDI value of this fatty acid ranks first among the high-fat crops).

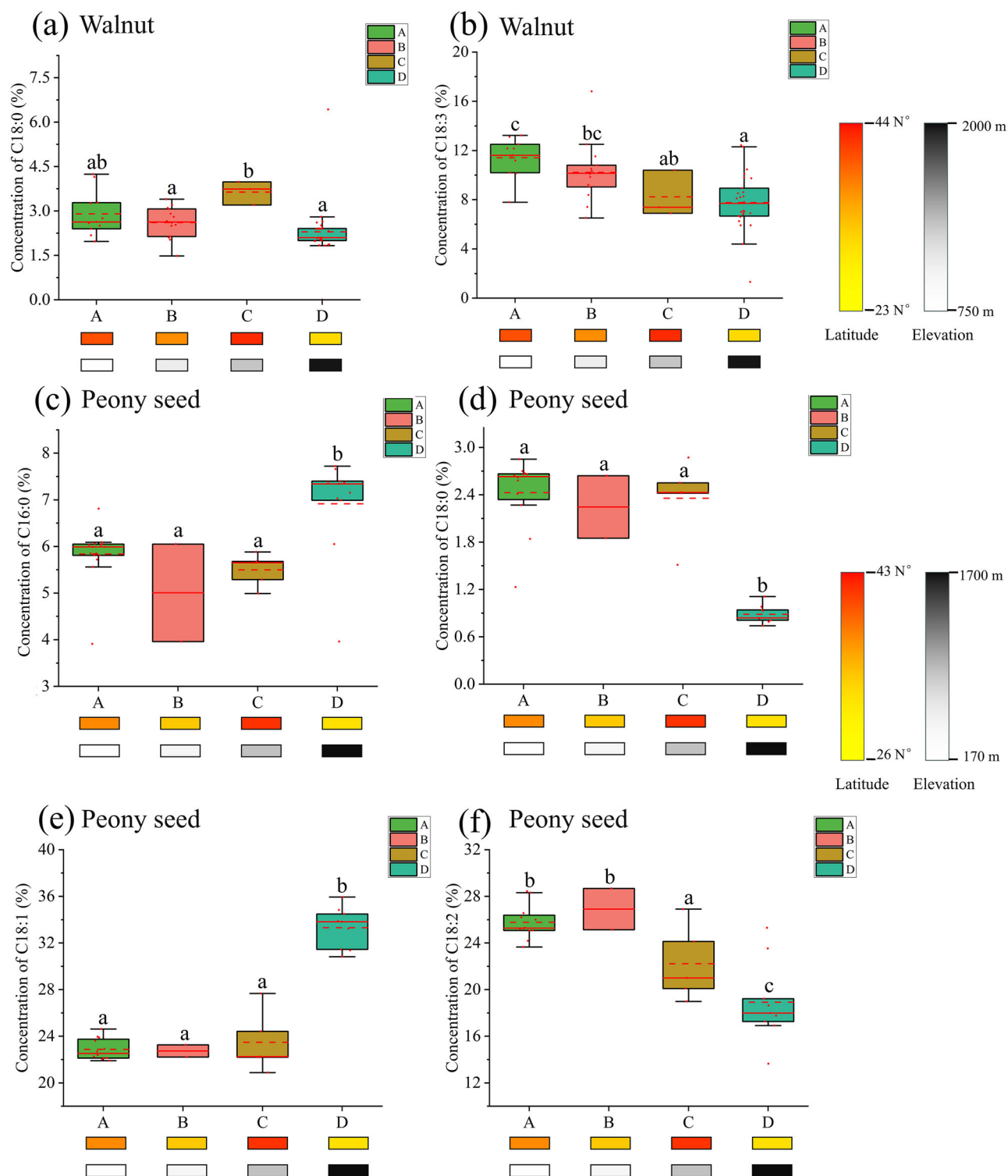


**Fig. 3 | Box plots statistical charts of olive oil and camellia oil.** a–c are box plots of the fatty acid percentage composition in olive oil, d and e are box plots of the fatty acid percentage composition in camellia oil (different lowercase letters represented significant differences at 0.05 levels).

0.03–0.26%). Similarly, C20:1 (0.004–0.75%) was also excluded from the analysis because, ranked by percentage content, it was not among the top three unsaturated fatty acids.

The components used as markers of agricultural product origin must differ between origins and possess identifiable and easily detectable characteristics. After two rounds of screening, the potential origin discrimination markers obtained for the four high-oil crops are as follows: olive oil: C18:0, C18:1 and C18:2; camellia oil: C18:0 and C18:3; walnut: C18:0 and

C18:3; peony seeds: C16:0, C18:0, C18:1 and C18:2. Among saturated fatty acids, it is worth noting that C18:0 has been identified as a potential geographical marker in all four high-oil crops. C18:0 could be an important fatty acid for distinguishing between high-fat crops from different production areas. The comprehensive analysis revealed that the marker combination specific to olive oil, C18:0(SFA)/C18:1, C18:2 (UFA), in conjunction with analogous profiles observed in three other crops, constitutes a core parameter set for metabolomics-based geotracing systems. This approach



**Fig. 4 | Box plots statistical charts of Walnuts and peony seeds.** a–d are box plots of the fatty acid percentage composition in walnuts, e, f are box plots of the fatty acid percentage composition in peony seeds (Different lowercase letters represented significant differences at 0.05 levels).

provides molecular-level quantitative evidence for the certification of crop origin.

#### Differential fatty acid analysis

In this study, the differences in geographical distribution of fatty acid composition across oil and nut crops were systematically elucidated through a comprehensive analysis of GDI characteristics (For details, please refer to Figs. 3 and 4. (The main text displays the fatty acids retained after GDI

screening, and the box plots of the remaining fatty acids are shown in Supplementary Fig. 1–4). (The fatty acid content mentioned here refers to the percentage content. The same applies to the fatty acid content mentioned in the following text).

As shown in Fig. 3a–c presents box plots of the most differentiated fatty acids in olive oil from distinct geographical origins, while Fig. 3d and e present box plots of camellia oil. The horizontal axis is arranged in order of progressively increasing elevation from left to right. Specifically, Fig. 3a



illustrates the relationship between C18:0 content in olive oil and elevation/latitude, revealing the highest mean C18:0 content in origin A (3.15; the content here refers to percent, and the contents mentioned below are all percent) and the lowest in origin E (1.45). The average C18:0 content in origin A was 2.17-fold higher than that in origin E, indicating a statistically significant disparity. However, if the latitude and elevation differences between the different origin areas are small, there is no significant difference ( $p > 0.05$ ) between the fatty acids of the different origin areas. As shown in Fig. 3b, the fatty acid content of C18:1 in olive oil exhibited a positive correlation with elevation gradient (indicated by a colour gradient from white to black) across origin areas A to C. A similar trend was observed for appellations D to F, where C18:1 content increased with rising elevation. However, a decline in C18:1 content was noted between origin areas C and D. In contrast, C18:2 displayed an inverse relationship with elevation: its content progressively decreased as elevation increased (Fig. 3c). These trends align with findings from studies by Rey-Giménez<sup>36</sup>, Revelou<sup>37</sup>, Dehghan Nayeri<sup>38</sup>, and Dar<sup>39</sup>, which attribute the inverse correlation between C18:1 and C18:2 to the enzymatic activity of fatty acid desaturase (FAD2). Specifically, FAD2 catalyzes the conversion of oleic acid (C18:1) to linoleic acid (C18:2) during olive ripening via double bond insertion<sup>36,40</sup>. Notably, a variation in C18:2 levels occurred between origin areas C and D, despite their overall declining trend across both A~C and D~F origins with increasing elevation. This anomaly is explained by the combined influence of elevation and latitude: the higher latitude range (42°~46°) in origin area D promotes C18:2 synthesis, resulting in an unexpected surge in C18:0 content amidst the general decreasing pattern.

As illustrated in Fig. 3d, the C18:0 content of camellia oil showed no significant variation with elevation or latitude across most origin areas. However, Area C exhibited a significantly elevated C18:0 content (3.19) compared to other origins (A: 2.47, B: 2.23, D: 1.96, E: 2.46, F: 2.38, G: 2.00,  $p < 0.05$ ), indicating that the unique tropical monsoon island climate could significantly promote the synthesis of C18:0 in camellia<sup>23</sup>. In Fig. 3d, C18:3 fatty acid content demonstrated a progressive increase with elevation gradient across origin Area A to D. A similar trend was observed in Areas E to G, where C18:3 levels also rose with increasing elevation. Notably, C18:3 content declined between Areas D and E. Indicated that the D and E origin's subtropical monsoon climate likely affects C18:3 synthesis, leading to a markedly lower mean C18:3 content in Area E (0.28) compared to other origins (A: 0.31, B: 0.34, C: 0.43, D: 0.47, F: 0.35, G: 0.53,  $p < 0.05$ ).

As depicted in Fig. 4a, C18:0 content in walnuts displayed marked variation across origin areas. Area C exhibited a significantly higher mean C18:0 content (3.64) compared to Areas A, B, and D (2.90, 2.61, and 2.30, respectively,  $p < 0.05$ ). The results showed that in the sample range (37°~44°), the higher the latitude, the easier C18:0 was to accumulate in walnuts. Notably, C18:0 content in Areas A, B, and D remained stable despite increasing elevation gradients. In contrast, Fig. 4b reveals a progressive decline in C18:3 fatty acid content with rising elevation gradients, suggesting that higher elevations may inhibit the synthesis of C18:3 in walnuts.

As illustrated in Fig. 4c, peony seeds' C16:0 content exhibited distinct variability across origin areas. Area D demonstrated a significantly higher mean C16:0 content (7.22) compared to Areas A, B, and C (2.84, 5.01, and 5.67, respectively). This divergence is attributed to Area D's unique geographical conditions: the lowest latitude and highest elevation among all origins. Low latitude and high elevation, as combined environmental factors, may inhibit the synthesis of C16:0 in peony seeds. Notably, minimal differences in C16:0 content were observed between origin areas with small latitude and elevation gradients. The content of C18:0 in peony seeds exhibited a similar trend (Fig. 4c, d). When the geographical differences (latitude and elevation) between origin areas were minor, C18:0 levels showed no significant variations ( $p > 0.05$ ). However, the combined influence of low latitude and high elevation markedly suppressed C18:0 accumulation, particularly in Area D, where the average content (0.89) was significantly lower than in Areas A, B, and C (2.43, 2.45, and 2.36, respectively,  $p < 0.05$ ) (Fig. 4d). These results further confirm that the interaction

of low latitude and high elevation affects the synthesis of C18:0 in peony seeds. As shown in Fig. 4e, the C18:1 content in peony seed oil exhibited a distribution pattern analogous to that observed for C16:0. Area D demonstrated a significantly higher mean C18:1 concentration (33.313) compared to Areas A, B, and C (17.327, 17.770 and 23.484, respectively) ( $p < 0.05$ ). This discrepancy can be attributed to Area D's distinctive geographical characteristics—the lowest latitude and highest elevation among all sampling sites. The combined effect of low latitude and high elevation conditions may promote C18:1 accumulation in peony seeds. As shown in Fig. 4f, the mean C18:2 content in peony seeds from Areas C and D (22.22 and 18.92, respectively) was significantly lower than in Areas A and B (25.77 and 26.91, respectively,  $p < 0.05$ ). This low content level of C18:2 in area D can be explained by the combined inhibitory effects of high elevation and low latitude on the synthesis of C18:2. In contrast, Area C's decreased C18:2 content likely stems from extremely high-latitude conditions, which similarly inhibit the synthesis of C18:2.

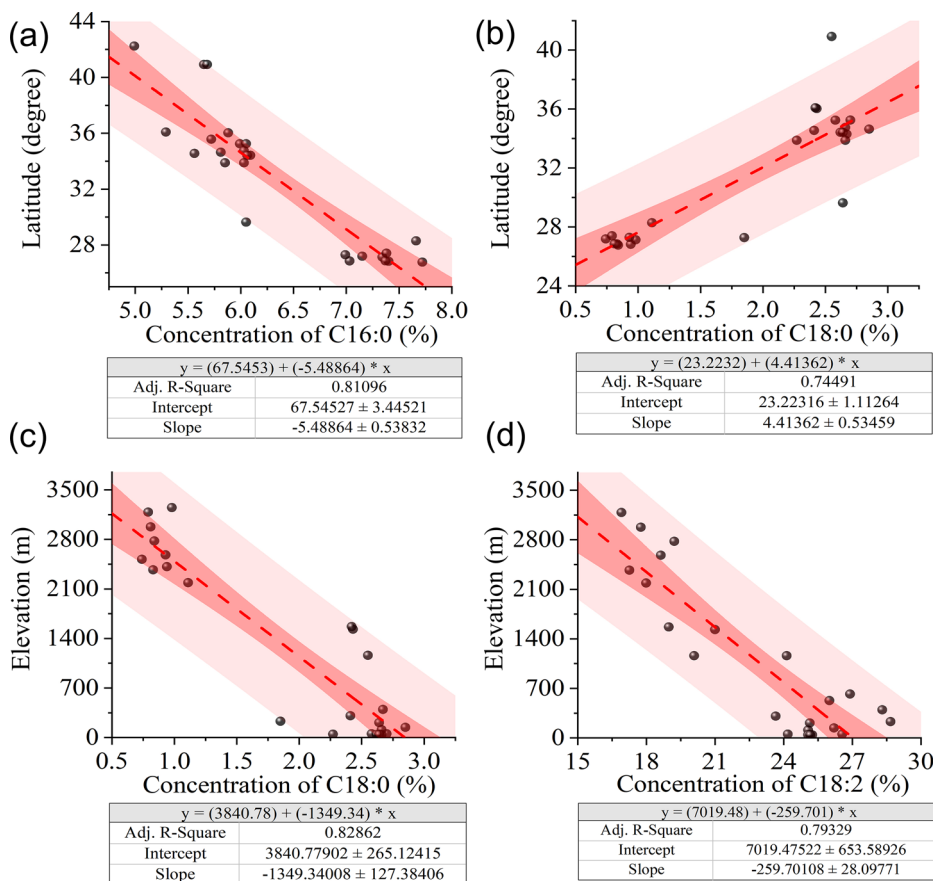
Linear regression analyses were performed using Origin 2021 to evaluate the relationships between geographic parameters (x: latitude and altitude) and fatty acid contents (y) in four oil-rich crop samples. Scatter plots were generated to visualize the data distribution, followed by simple linear regression via the Simple Fit tool. The fitted models provided regression equations ( $y = \text{Slope} \cdot x + \text{Intercept}$ ), along with Pearson's correlation coefficient ( $r$ ), adjusted R-squared (Adj.  $R^2$ ). The results showed that the fatty acid data of olive oil, camellia oil and walnuts had a weak correlation with geographical parameters ( $R^2 < 0.7$ ), as detailed in Supplementary Fig. 5–22. (Supplementary Figs. 5–10 is linear regression analysis of olive oil, Supplementary Figs. 11–16 is camellia oil, Supplementary Figs. 17–20 is walnuts, Supplementary Figs. 21, 22 is peony seeds). Only the linear fitting results of the three fatty acids of peony seed C16:0, C18:0, C18:2 were more satisfactory ( $R^2 > 0.7$ ), as detailed in Fig. 5. From the above, the fatty acid composition of peony seeds showed the strongest correlation with geographical parameters. Peony seeds revealed significant latitude-dependent correlations for C16:0 and C18:0 acids ( $R^2 = 0.81$  and  $0.74$ , respectively) in above study, with C16:0 decreasing and C18:0 increasing at higher latitudes. Conversely, C18:0 and C18:2 contents exhibited strong negative correlations with elevation ( $R^2 = 0.82$  and  $0.79$ ), showing suppressed accumulation at elevated elevations. These analyses results revealed the specific variation patterns of fatty acid composition in peony seeds across diverse geographical environments.

## Discussion

This study was conducted within a global geographical context to identify agricultural biomarkers that are responsive to cross-latitudinal and elevation factors. Given the challenges associated with acquiring global single-plural sample data and the limitations in geographical traceability accuracy, a multi-source, cross-species integration strategy was implemented. Four representative oil crop and nut crop, widely distributed across both hemispheres, were selected to establish the sample framework.

Data collection for camellia oil, walnuts and peony seeds was focused on China and these sample collections comprehensively covered the latitude gradient (23 N°~43 N°) and elevation range (278~2696 m). For example, the data collection for walnuts covered several latitudinal bands. Sample data were mainly collected from the Mourning Mountains in Yunnan (23 N°) to the Ali River Valley in Xinjiang (43 N°), creating a sample database across several latitudinal zones. Meanwhile, camellia oil data collection covered several elevation gradients. From the southeastern coast to the Yunnan-Guizhou Plateau, our dataset of fatty acids covers diverse geographic regions, including Anhui Dabie Mountain, Hainan Wuzhi Mountain, Guangdong Nanling Mountains, and Yunnan Hengduan Mountains. At the same time, due consideration was given to the breadth of the data collection and data were also collected for countries other than China. An example is the data collection on olive oil. We focused on the effect of latitudinal heterogeneity in the main olive-producing origins, such as the Mediterranean coast (42 N°) and the South American Andes (33 S°). Finally, the above data were statistically collected to create a database containing several climatic zones and spanning several latitudinal bands.

**Fig. 5 | Linear regression analysis of peony seeds ( $R^2 > 0.7$ ). a and b show linear regression analyses of latitude versus peony seeds C16:0 and C18:0; c and d show linear regression analyses of elevation versus peony seeds C18:0 and C18:2.**



Meanwhile, we also discovered and summarised some correlations between the fatty acid content of four oil-rich crops and latitude and altitude. As demonstrated in the previous section, the content of three fatty acids (C18:0, C18:1 and C18:2) in olive oil is influenced by elevation. When the elevation difference between producing origins is minimal, the variation in fatty acid composition is also limited. However, a stronger correlation between fatty acid levels and elevation was observed in origins with extremely low latitudes (area A, mean elevation: 150 m). These findings also supported the previous reports by Aguilera<sup>41</sup>, indicated that the saturated fatty acids were found at higher levels and greater concentrations in low-elevation regions. Furthermore, the impact of low elevation varies among different types of fatty acids. For instance, the content of C18:1 in area A was significantly lower than in other origins, whereas the levels of C18:0 and C18:2 were notably higher. This trend is consistent with the findings of Di Vaio<sup>42</sup>, who reported that olive oils from higher elevations exhibited elevated contents of monounsaturated fatty acid C18:1. Similarly, Zhang<sup>43</sup> documented a positive correlation between oleic acid content and latitude. In addition to elevation, the variation in C18:0 content was also influenced by latitude (Fig. 3a), with the effect intensifying as latitudinal differences between producing origins expanded. For example, in origin E (South America, latitude:  $-35^{\circ}$ – $19^{\circ}$ ), the synthesis of C18:0 was suppressed by its extremely low latitudinal position, resulting in significantly lower contents compared to other origins ( $p < 0.05$ ).

Analysis of camellia oil fatty acid revealed that the content of C18:0 and C18:3 varies minimally when producing origins demonstrate small latitudinal and elevation differences. However, under extreme environmental conditions—specifically, very high elevations and very low latitudes—these fatty acids show stronger correlations with both geographic factors. For instance, low latitude promoted the synthesis of C18:0, resulting in significantly higher C18:0 content ( $p < 0.05$ ) in origin C (latitudinal range:  $18^{\circ}$ – $20^{\circ}$ , the lowest among all study origins). Conversely, high elevation

facilitated C18:3 synthesis, leading to substantially elevated C18:3 content in origin G (1100 m mean elevation, the highest of the sampled origins). Moreover, numerous studies have substantiated the influence of environmental factors on the fatty acid composition of camellia oil. Liu<sup>44</sup> reported that saturated fatty acid (SFA) content in cultivated camellia oil increased with rising latitude, while unsaturated fatty acid (UFA) levels declined. Similarly, Gao<sup>45</sup> identified a latitudinal gradient effect on C16:0 and C18:1 profiles in camellia oil, further corroborating the role of geographic factors in modulating fatty acid biosynthesis.

Fatty acid analysis of walnuts revealed that geographical factors and climatic conditions are critical determinants of their fatty acid composition<sup>46</sup>. Specifically, increasing elevation was found to suppress the synthesis of C18:3, whereas higher latitudes significantly enhanced the accumulation of C18:0. Fatty acid composition in peony seeds was significantly influenced by geographical factors<sup>47,48</sup>. Analysis revealed that C16:0, C18:0, C18:1 and C18:2 exhibited minimal variation in composition when origin areas showed small latitudinal and elevation differences. In contrast, under extreme low-latitude, high-elevation conditions (origin D:  $26^{\circ}$ – $28^{\circ}$ N, mean elevation 2696.56 m), these fatty acids demonstrated markedly stronger responses to both geographic variables. Notably, the synergistic effects of low latitude and high elevation diverged among fatty acids: they promoted the accumulation on C16:0 and C18:1, but inhibited the synthesis of C18:0 and C18:2. These findings align with Chang<sup>48</sup>, who further demonstrated that high-latitude and low-elevation environments similarly alter fatty acid accumulation patterns in peony seeds.

Therefore, the extreme environmental differences exert a pronounced impact on oil-rich crops. However, when environmental variations between origin areas are minimal, the correlation between fatty acid content and environmental factors becomes less pronounced. It is hypothesized that under such conditions, other factors—such as genotype—may also

influence fatty acid composition. This hypothesis is supported by previous studies. Cheng reported that fatty acid content and composition in camellia oil vary depending on origin, cultivar, and growing conditions, complicating the establishment of a unique fatty acid fingerprint<sup>49,50</sup>. Similarly, Navas-López demonstrated that genotype is the primary source of variation in the major fatty acids of olive oil<sup>51</sup>. Furthermore, PCA results from Sakar revealed that Moroccan olive oil physicochemical properties are influenced by multiple factors, including variety, agroclimatic conditions, and extraction techniques<sup>52</sup>. These findings suggest that when environmental differences between origin areas are small, environmental factors alone may not exclusively determine the fatty acid composition of oil-rich crops, as genetic and agrotechnical factors also play significant roles.

In conclusion, this study first constructed a cross-species origin tracing framework based on EHI and GDI by integrating geo-climatic characteristics and fatty acid metabolomics data. The EHI effectively evaluates the influence of extrinsic environmental conditions and intrinsic variations on plant fatty acid profiles, whereas the GDI efficiently identifies fatty acids exhibiting the most pronounced differences among various production regions. By integrating these two innovative evaluation index, environmental zones for oil crop cultivation were effectively delineated. Additionally, fatty acid differences among these regions were quantitatively assessed using the indices, laying a foundation for further theoretical analysis. The study demonstrated that all four high oil-rich crop species possess characteristic fatty acid profiles that, when under extreme environmental conditions, exhibit significant responsiveness to geographical gradients (latitudinal and elevation variations). Notably, the fatty acid profiles of peony seeds exhibited a pronounced linear correlation with latitudinal and elevation gradients. Furthermore, the regulation of fatty acids by geographical parameters followed a gradient transition pattern: as latitude or elevation increased, specific fatty acid synthesis pathways progressively shifted from being predominantly driven by micro-environmental factors to being regulated by macro-geographical factors. These findings establish a theoretical foundation for the development of metabolic fingerprint-based traceability systems in agricultural products and broaden the scope of research on plant environmental adaptation mechanisms.

Finally, this study has several limitations that need to be acknowledged. Analyses of the available dataset revealed that the internal variability of fatty acid profiles in specific geographically grouped samples was significantly higher than the overall intergroup heterogeneity ( $EHI > 1$ ), indicating that sub-environmental differences or genetic background heterogeneity within origins may exert non-negligible regulatory effects on fatty acid metabolism. First, this phenomenon could be attributed to species-specific differences: in the same original group, variations in fatty acid profiles between different species (e.g., due to genetic or enzymatic factors) might dominate over geographic influence. Second, sub-environmental heterogeneity likely plays a role, as a subset of geographically grouped areas exhibited pronounced environmental gradients, thereby amplifying intra-group data variability. These findings underscore the need for future studies to establish a geo-genetic biaxial classification framework to enhance the robustness of fatty acid traceability models.

## Methods

### Information sources and search strategy

PubMed, Web of Science, Science Direct, and CNKI (China National Knowledge Infrastructure) databases were searched for dates ranging from December 2003 to December 2023, utilizing a structured search strategy guided by keywords and Boolean arithmetic. The study selection was guided by a structured search strategy using subject terms, keywords, and Boolean operator arithmetic. The keywords were classified into five categories, each representing a unique concept. The search keywords were (a) Identification of the geographical origin of fatty acids, (b) Olive oil, (c) Camellia oil, (d) Walnut, (e) Peony seed. The actual search terms were as follows: ('Olive Oil' OR 'Oil, Olive' OR 'Oils, Olive' OR 'Olive Oils') AND ('Fatty Acids' OR 'Aliphatic Acid') AND ('Geographical origin'); ('Camellia oil') AND ('Fatty

Acids' OR 'Aliphatic Acid') AND ('Geographical origin'); ('Juglans' OR 'Walnut') AND ('Fatty Acids' OR 'Aliphatic Acid') AND ('Geographical origin'); ('Acids' OR 'Aliphatic Acid') AND ('geographical origin'); ('Peony seed') AND ('Fatty Acids' OR 'Aliphatic Acid'); ('Identification') AND ('Geographical origin') AND ('Fatty Acids' OR 'Aliphatic Acid').

This study obtained 1054 articles, and after removing 214 duplicates, 840 relevant papers were accepted for the next analysis. The titles, keywords, and abstracts of the collected articles were first screened based on our inclusion criteria that the research objects were crops originating from diverse sources worldwide. A total of 146 articles were selected for the next study. Subsequently, these articles were further checked to ensure that the data reported the fatty acid content of the research objects using GC-MS or other similar chemical analysis methods. Ultimately, A total of 44 effective articles with comparative study and statistical analysis were included for this re-statistical analysis. An overview of these 44 articles can be found in Supplementary Table 3<sup>53–77</sup>.

### Classification method of climatic features

In this study, a multi-source geographic information integration framework was constructed, and a crop geographic archive was established through the integration of literature data mining and digital platform validation. The basic geographic data were extracted using a three-stage validation mechanism. First, the literature traceability method was employed, with crop origin information extracted from the "Materials and Methods" section of the incorporated literature. Second, when the literature records were incomplete, a plant database cross-validation procedure was initiated. Authoritative plant distribution data from Plants of the World Online, locality information of specimens from the Digital Herbarium of China, and geographic metadata of China's national specimen resources were searched. For specimens with missing latitude and longitude information, reverse geocoding was performed using the Google Map (<http://www.gditu.net>) coordinate picking system, and coordinate calibration was conducted based on the dual constraints of administrative areas and terrain features. For the acquisition of spatial environmental variables, habitat information corresponding to sample points was obtained from the WorldClim Global Climate Database (<http://www.worldclim.org>). The ecological parameters of all sample points were extracted in batch using the Spatial Analyst module of ArcGIS 10.4. The data processing procedure included the following steps: (1) constructing topological groupings based on plant species and mapping their spatial distribution using the geographic visualization function of ArcGIS 10.4; and (2) establishing a hierarchical grouping system based on three-dimensional geographic attributes, including latitude gradient, latitude and longitude coordinates, and climatic zoning codes, using the K-means clustering algorithm. This hierarchical grouping system lays the foundation for the subsequent multidimensional analysis of ecological sites and model construction.

### Data analysis and statistics

The IBM SPSS Statistics 25 software was employed as the primary analytical framework. One-way ANOVA was conducted to assess statistical differences in fatty acid percentages among geographical groups, with the significance level set at  $\alpha = 0.05$ . This analysis incorporated chi-square tests and corrections for multiple comparisons. Distribution patterns of fatty acids were visualized using Origin 2021 box plots, which detailed central tendencies, dispersion, and outliers across original clusters.

### Systematic nomenclature for fatty acids

Fatty acids: C14:0, myristic acid (tetradecanoic acid), C16:0, palmitic acid (hexadecanoic acid), C16:1, palmitoleic acid (hexadecenoic acid), C17:0, margaric acid (heptadecanoic acid), C17:1, heptadecenoic acid (no common name), C18:0, stearic acid (octadecanoic acid), C18:1, oleic acid (octadecenoic acid), C18:2, linoleic acid (octadecadienoic acid), C18:3,  $\alpha$ -linolenic acid (octadecatrienoic acid), C20:0, arachidic acid (eicosanoic acid), C20:1, gondoic acid (eicosenoic acid), C22:0, behenic acid (docosanoic acid),



C22:1, erucic acid (docosanoic acid), C24:0, lignoceric acid (tetracosanoic acid), Saturated Fatty Acids (SFA), Monounsaturated Fatty Acids (MUFA), Polyunsaturated Fatty Acids (PUFA), Unsaturated Fatty Acids (UFA).

## Data availability

No datasets were generated or analysed during the current study.

Received: 15 May 2025; Accepted: 18 August 2025;

Published online: 30 September 2025

## References

- Hew, P. S. et al. Quality and safety of food product - Current assessment, issues, and metabolomics as a way forward. *Food Chem. Adv.* **4**, 100632 (2024).
- Janjarasskul, T. & Suppakul, P. Active and intelligent packaging: The indication of quality and safety. *Crit. Rev. Food Sci. Nutr.* **58**, 808–831 (2018).
- Li, Y., Man, S., Ye, S., Liu, G. & Ma, L. CRISPR-Cas-based detection for food safety problems: Current status, challenges, and opportunities. *Compr. Rev. Food Sci. Food Saf.* **21**, 3770–3798 (2022).
- Begum, M., Alam, M. J., Parikh, P. & De Steur, H. Understanding food safety knowledge, attitude, and practices of consumers and vendors: an Umbrella review. *Food Control* **171**, 111094 (2025).
- Ding, H., Xie, Z., Yu, W., Cui, X. & Wilson, D. I. Artificial intelligence enhances food testing process: A comprehensive review. *Food Biosci.* **68**, 106404 (2025).
- Aung, M. M. & Chang, Y. S. Traceability in a food supply chain: Safety and quality perspectives. *Food Control* **39**, 172–184 (2014).
- Yu, Z. et al. Smart traceability for food safety. *Crit. Rev. Food Sci. Nutr.* **62**, 905–916 (2022).
- Katerinopoulou, K., Kontogeorgos, A., Salmas, C. E., Patakas, A. & Ladavos, A. Geographical origin authentication of agri-food products: A review. *Foods* **9**, 489 (2020).
- Fu, X. & Ying, Y. Food safety evaluation based on near infrared spectroscopy and imaging: A review. *Crit. Rev. Food Sci. Nutr.* **56**, 1913–1924 (2016).
- Wang, L., Sun, D. W., Pu, H. & Cheng, J. H. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. *Crit. Rev. Food Sci. Nutr.* **57**, 1524–1538 (2017).
- Zaroual, H., Chénè, C., El Hadrami, E. M., Romdhane, M. & Karoui, H. Application of new emerging techniques in combination with classical methods for the determination of the quality and authenticity of olive oil: a review. *Crit. Rev. Food Sci. Nutr.* **62**, 4526–4549 (2022).
- Laroussi-Mezghani, S. et al. Authentication of Tunisian virgin olive oils by chemometric analysis of fatty acid compositions and NIR spectra. Comparison with Maghrebian and French virgin olive oils. *Food Chem.* **173**, 122–132 (2015).
- Fu, J. et al. Application of stable isotope and mineral element fingerprint in identification of Hainan Camellia oil producing area based on convolutional neural networks. *Food Control* **150**, 109744 (2023).
- Cassagne, C. et al. Biosynthesis of very long chain fatty acids in higher plants. *Prog. Lipid Res.* **33**, 55–69 (1994).
- Zhukov, A. & Popov, V. Synthesis of C20–38 fatty acids in plant tissues. *Int. J. Mol. Sci.* **23**, 4731 (2022).
- Zhang, H., Feng, T. & Chang, Q. Impact of molecular regulation on plant oil synthesis. *Plant Sci.* **354**, 112428 (2025).
- He, M., Qin, C. X., Wang, X. & Ding, N. Z. Plant unsaturated fatty acids: Biosynthesis and regulation. *Front. Plant Sci.* **11**, 390 (2020).
- Ohlrogge, J. & Browse, J. Lipid biosynthesis. *Plant Cell* **7**, 957–970 (1995).
- Hilditch, T. P. Biosynthesis of unsaturated fatty acids in ripening seeds. *Nature* **167**, 298–301 (1951).
- Batsale, M. et al. Biosynthesis and functions of very-long-chain fatty acids in the responses of plants to abiotic and biotic stresses. *Cells* **10**, 1284 (2021).
- Deiana, P. et al. Effect of pedoclimatic variables on analytical and organoleptic characteristics in olive fruit and virgin olive oil. *Eur. J. Agron.* **148**, 126856 (2023).
- Sakar, E. H. et al. Combined effects of cultivar, extraction technology, and geographic origin on physicochemical traits of Moroccan olive oil as revealed by multivariate analysis. *J. Food Compos. Anal.* **133**, 106375 (2024).
- Singer, S. D., Zou, J. & Weselake, R. J. Abiotic factors influence plant storage lipid accumulation and composition. *Plant Sci.* **243**, 1–9 (2016).
- Dou, X. et al. Mass spectrometry in food authentication and origin traceability. *Mass Spectrom. Rev.* **42**, 1772–1807 (2023).
- Hu, B. et al. Chemometrics analysis of camellia oil authenticity using LF NMR and fatty acid GC fingerprints. *J. Food Compos. Anal.* **133**, 106447 (2024).
- Shi, T., Wu, G., Jin, Q. & Wang, X. <sup>1</sup>H NMR and chemical analysis to characterize camellia oil obtained by different extraction methods: A comparative study using chemometrics. *Food Chem.* **137**, 106926 (2025).
- Issaad, F. Z. et al. Classification of Algerian olive oils: Physicochemical properties, polyphenols and fatty acid composition combined with machine learning models. *J. Food Compos. Anal.* **125**, 105812 (2024).
- Shi, T., Dai, T., Wu, G., Jin, Q. & Wang, X. Camellia oil grading adulteration detection using characteristic volatile components GC-MS fingerprints combined with chemometrics. *Food Control* **169**, 111033 (2025).
- Spangenberg, J. E. & Lantos, I. Fingerprinting of monovarietal olive oils from Argentina and Uruguay via stable isotope, fatty acid profile, and chemometric analyses. *Food Chem.* **449**, 139194 (2024).
- Shi, T., Wu, G., Jin, Q. & Wang, X. Camellia oil authentication: A comparative analysis and recent analytical techniques developed for its assessment. A review. *Trends Food Sci. Technol.* **97**, 88–99 (2020).
- Deng, Z. et al. Geographical origin identification of Hainan camellia oil based on fatty acid composition and near infrared spectroscopy combined with chemometrics. *J. Food Compos. Anal.* **125**, 105730 (2024).
- Caño-Carrillo, I., Gilbert-López, B., Ruiz-Samblás, C., Molina-Díaz, A. & García-Reyes, J. F. Virgin olive oil authentication using mass spectrometry-based approaches: A review. *Trends Anal. Chem.* **181**, 118029 (2024).
- Rashed, M. M. A. et al. Traceability, authentication, and quality control of food-grade lavender essential oil: A comprehensive review. *Adv. Colloid Interface Sci.* **340**, 103466 (2025).
- Wadood, S. A., Boli, G., Xiaowen, Z., Hussain, I. & Yimin, W. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin. *Microchem. J.* **152**, 104295 (2020).
- Bulmer, M. G., The mathematical theory of quantitative genetics. Oxford: Clarendon Press: 1980.
- Rey-Giménez, R. & Sánchez-Gimeno, A. C. Effect of cultivar and environment on chemical composition and geographical traceability of Spanish olive oils. *Food Chem.* **101**, 371–382 (2024).
- Revelou, P. K. et al. Chemometric study of fatty acid composition of virgin olive oil from four widespread Greek cultivars. *Molecules* **26**, 4151 (2021).
- Nayeri, F. D. & Yarizade, K. Bioinformatics study of delta-12 fatty acid desaturase 2 (FAD2) gene in oilseeds. *Mol. Biol. Rep.* **41**, 5077–5087 (2014).
- Dar, A. A., Choudhury, A. R., Kancharla, P. K. & Arumugam, N. The FAD2 gene in plants: Occurrence, regulation, and role. *Front. Plant Sci.* **8**, 1789 (2017).
- Kritioti, A., Menexes, G. & Drouza, C. Chemometric characterization of virgin olive oils of the two major Cypriot cultivars based on their fatty acid composition. *Food Res. Int.* **103**, 426–437 (2018).



41. Aguilera, M. P. et al. Characterisation of virgin olive oil of Italian olive cultivars: 'Frantoio' and 'Leccino', grown in Andalusia. *Food Chem.* **89**, 387–391 (2005).
42. Di Vaio, C., Nocerino, S., Paduano, A. & Sacchi, R. Influence of some environmental factors on drupe maturation and olive oil composition. *J. Sci. Food Agric.* **93**, 1134–1139 (2013).
43. Zhang, Y. et al. Effect of geographical location and soil fertility on main phenolic compounds and fatty acids compositions of virgin olive oil from Leccino cultivar in China. *Food Res. Int.* **157**, 111207, <https://doi.org/10.1016/j.foodres.2022.111207> (2022).
44. Liu, X., Jia, L., Gao, Y., Li, B. & Tu, Y. Anti-inflammatory activity of total flavonoids from seeds of *Camellia oleifera* Abel. *Acta Biochim. Biophys. Sin.* **46**, 920–922 (2014).
45. Gao, S. et al. Variation in fruit morphology and seed oil fatty acid composition of *Camellia oleifera* collected from diverse regions in Southern China. *Horticulturae* **8**, 818 (2022).
46. Rébua, C., Artaud, J. & Le Dréau, Y. Walnut (*Juglans regia* L.) oil chemical composition depending on variety, locality, extraction process and storage conditions: A comprehensive review. *J. Food Compos. Anal.* **110**, 104534 (2022).
47. Liu, F. et al. Fatty acid composition analysis of peony seed oil at different elevations and areas. *J. Bionosci.* **11**, 455–459 (2017).
48. Chang, M. et al. Characterization of fatty acids, triacylglycerols, phytosterols and tocopherols in peony seed oil from five different major areas in China. *Food Res. Int.* **137**, 109416 (2020).
49. Zhang, F. et al. Composition, bioactive substances, extraction technologies and the influences on characteristics of *Camellia oleifera* oil: A review. *Food Res. Int.* **156**, 111159 (2022).
50. Cheng, X. et al. New method for effective identification of adulterated *Camellia* oil basing on *Camellia oleifera*-specific DNA. *Arab. J. Chem.* **11**, 815–826 (2018).
51. Navas-López, J. F., Cano, J., de la Rosa, R., Velasco, L. & León, L. Genotype by environment interaction for oil quality components in olive tree. *Eur. J. Agron.* **119**, 126115 (2020).
52. Hassan Sakar, E. I., Aalam, Z., Khtira A., Uluta, S., Durmaz, G. & Gharby, S. Combined effects of cultivar, extraction technology, and geographic origin on physicochemical traits of Moroccan olive oil as revealed by multivariate analysis. *J. Food Compos. Anal.* **133**, 106375, <https://doi.org/10.1016/j.jfca.2024.106375>.
53. Peršurić, Ž., Saftić, L., Mašek, T. & Kraljević Pavelić, S. Comparison of triacylglycerol analysis by MALDI-TOF/MS, fatty acid analysis by GC-MS and non-selective analysis by NIRS in combination with chemometrics for determination of extra virgin olive oil geographical origin. A case study. *LWT* **95**, 326–332 (2018).
54. Ouni, Y. et al. Volatile compounds and compositional quality of virgin olive oil from Oueslati variety: Influence of geographical origin. *Food Chem.* **124**, 1770–1776 (2011).
55. Longobardi, F. et al. Characterisation of the geographical origin of Western Greek virgin olive oils based on instrumental and multivariate statistical analysis. *Food Chem.* **133**, 169–175 (2012).
56. Longobardi, F. et al. Instrumental and multivariate statistical analyses for the characterisation of the geographical origin of Apulian virgin olive oils. *Food Chem.* **133**, 579–584 (2012).
57. Borges, T. H. et al. Characterization of Arbequina virgin olive oils produced in different regions of Brazil and Spain: Physicochemical properties, oxidative stability and fatty acid profile. *Food Chem.* **215**, 454–462 (2017).
58. Baccouri, B. et al. Application of solid-phase microextraction to the analysis of volatile compounds in virgin olive oils from five new cultivars. *Food Chem.* **102**, 850–856 (2007).
59. Piscopo, A., De Bruno, A., Zappia, A., Ventre, C. & Poiana, M. Characterization of monovarietal olive oils obtained from mills of Calabria region (Southern Italy). *Food Chem.* **213**, 313–318 (2016).
60. Aranda, F., Gómez-Alonso, S., Rivera del Álamo, R. M., Salvador, M. D. & Fregapane, G. Triglyceride, total and 2-position fatty acid composition of Cornicabra virgin olive oil: Comparison with other Spanish cultivars. *Food Chem.* **86**, 485–492 (2004).
61. Bajoub, A. et al. Quality and chemical profiles of monovarietal north Moroccan olive oils from 'Picholine Marocaine' cultivar: Registration database development and geographical discrimination. *Food Chem.* **179**, 127–136 (2015).
62. Ollivier, D., Artaud, J., Pinatel, C., Durbec, J.-P. & Guérère, M. Differentiation of French virgin olive oil RDOs by sensory characteristics, fatty acid and triacylglycerol compositions and chemometrics. *Food Chem.* **97**, 382–393 (2006).
63. Kosma, I., Badeka, A., Vatavali, K., Kontakos, S. & Kontominas, M. Differentiation of Greek extra virgin olive oils according to cultivar based on volatile compound analysis and fatty acid composition. *Eur. J. Lipid Sci. Technol.* **118**, 849–861 (2016).
64. Rodrigues, N. et al. Geographical discrimination of olive oils from Cv. 'Galega Vulgar'. *Food Chem.* **398**, 133945 (2023).
65. Liu, W. et al. Discrimination of geographical origin of extra virgin olive oils using terahertz spectroscopy combined with chemometrics. *Food Chem.* **251**, 86–92 (2018).
66. López-Cortés, I., Salazar-García, D. C., Velázquez-Martí, B. & Salazar, D. M. Chemical characterization of traditional varietal olive oils in East of Spain. *Food Res. Int.* **54**, 1934–1940 (2013).
67. Wang, P. X. The impact of different production areas and tree ages in Hainan Province on the economic traits of Vietnamese *camellia oleifera*. Unpublished doctoral dissertation, *Hainan University*, Haikou, China. <https://doi.org/10.27073/d.cnki.ghadu.2022.001059> (2022).
68. Liu, H. et al. Effects of different *camellia* species and origins in Guizhou on the quality of *camellia* seed oil. *China Oils Fats* **47**, 145–152 (2022).
69. Wang, Y. P. et al. Principal component and cluster analysis of fatty acids and triglycerides in *camellia* seed from different origins. *China Oils Fats* **46**, 112–119 (2021).
70. Chen, X. L. et al. Comparison of oils and nutritional components from wild *camellia* seeds with different single-grain masses from ten regions. *Food Sci.* **40**, 227–234 (2019).
71. Ye, Z. C. et al. A comparative analysis of the antioxidant activities of extracts from *camellia* seed oils of different origins and their nutritional evaluation. *Biotechnol. Bull.* **35**, 80–88 (2019).
72. Yang, Y., Zhang, P., Xi, R. C. & Huang, R. R. Variability in oil content and fatty acid composition of *camellia oleifera* fruits from different production areas in Gaozhou. *J. Econ. For. Res.* **36**, 144 (2018).
73. He, X. S. et al. Comparative study on fatty acid content of tea oil from different regions in Jiangxi. *J. South. For. Sci.* **46**, 6–9 (2018).
74. Feng, Q. Q., Wu, M. M., Wang, Y. C. & Ma, Y. J. Comparative analysis of the physicochemical properties and components of pressed *camellia* oil from different origins. *Grain Oil Sci. Technol.* **30**, 39–42 (2017).
75. Liu, J. L. et al. Research on the nutritional quality of major *camellia oleifera* seed oils in Hunan Province. *J. Chin. Cereals Oils Assoc.* **39**, 184–190 (2024).
76. Fan, H. Q. et al. Oil content and fatty acid composition of ten different *camellia* seed kernels. *J. Guizhou Univ.* **41**, 59–63 (2024).
77. Wang, Y. F. & Dai, X. J. Evaluating the quality of *camellia* oil from different producing areas using grey relational analysis. *J. China Univ. Metrol.* **33**, 116–122 (2022).

## Acknowledgements

The authors gratefully acknowledge financial support from Jiangxi Province Double Thousand Talent-Leader of Natural Science Project (JXSQ2023101038), Jiangxi Province Urgently Overseas Talent Project (2022BCJ25027), The Key Research Projects in Jiangxi Province (20223BBH8007&20232BBG70014), Youth Innovation Promotion Association CAS (2015286), and Jiangxi Provincial Natural Science Foundation (20242BAB20275). This work was also funded by the key research programme of Jiujiang City (S2023ZDYFN96, S2024ZDSYS037, S2024KXJJ0001).

## Author contributions

Junyu Xu: Conceptualization, Investigation, Writing – original draft. Yiqing Luo: Conceptualization, Formal analysis, Methodology. Lei Hu: Data curation, Writing – original draft. Kai Chen: Software. Shuang Cheng: Formal analysis. Mengting Xiao: Resources. Haijing Xiao: Formal analysis. Shumin Hou: Investigation. Jiaxin Yang: Writing – original draft. Jiao Yang: Visualization. Ying Wang: Conceptualization, Investigation. Shaohua Zeng: Formal analysis, Data curation, Writing – review & editing. Chunsong Cheng: Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41538-025-00555-z>.

**Correspondence** and requests for materials should be addressed to Shaohua Zeng or Chunsong Cheng.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025