Fruit Trait Variation Among Five Ilex Taxa: Insights from Morphology, Color, and Electronic Nose-Based Volatile Profiles

Xiangxian Fan
Nanjing Forestry University
Meng Sun
jingling keji xuexiao: Jinling Institute of Technology
ye peng
pengye@njfu.edu.cn

Nanjing Forestry University https://orcid.org/0000-0001-7633-0012

Research Article

Keywords: Ilex, fruit morphology, electronic nose, principal component analysis, OPLS□DA.

Posted Date: February 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8701723/v1

Fruit Trait Variation Among Five *Ilex* Taxa: Insights from Morphology, Color, and Electronic Nose-Based Volatile Profiles

Xiangxian Fan[1], Meng Sun[2], Ye Peng[*]

1 The Co-Innovation Center for Sustainable Forestry in Southern China, College of Life Sciences, Nanjing Forestry University, Nanjing 210037, China

2 College of Horticulture, Jinling Institute of Technology, Nanjing 210038, China

*Corresponding author: Ye Peng

Address: College of Life Sciences, Nanjing Forestry University, No. 159 Longpan Road, Nanjing, 210037, Jiangsu Province, China

Email: pengye@njfu.edu.cn

Abstract:

Accurate discrimination of closely related Ilex taxa is essential for effective germplasm management and taxon-specific applications. In this study, five *Ilex* taxa (*Ilex rotunda* Thunb., *I. chinensis*, *I. cornuta* Lindl. et Paxt., *I. cornuta* 'Fortunei', and *I. latifolia* Thunb.) were evaluated using an integrated framework that combines fruit morphometrics and CIELAB color parameters with electronic-nose (E-nose) odor fingerprints and chemometric analyses. Fruit transverse diameter, longitudinal diameter, single-fruit weight, fruit shape index, and color parameters (L*, a*, b*, and chroma (C*)) differed significantly among taxa (one-way ANOVA, all $p < 0.001$). Notably, I. cornuta exhibited the largest fruits and the highest fruit weight, whereas *I. rotunda* had the smallest fruit diameters but the highest redness (a*) and chroma (C*). E-nose data from 10 metal-oxide semiconductor (MOS) sensors were analyzed using principal component analysis (PCA), hierarchical clustering, and supervised OPLS-DA. PCA explained 90.21% of the variance using the first two components (PC1, 71.8%; PC2, 10.5%) and revealed taxon-specific clustering. A targeted OPLS-DA model separating *I. rotunda* from the other four taxa showed strong goodness-of-fit and predictive ability ($R^2X = 0.785$, $R^2Y = 0.944$, $Q^2 = 0.935$). Permutation testing (n = 200) supported model validity ($p < 0.005$; 0 of 200 permutations exceeded the original model). Together, fruit phenotypes and E-nose chemometrics provided robust discrimination among the studied taxa, with *I. rotunda* consistently exhibiting the most divergent profile, consistent with the phylogenetic framework used in this study.

Keywords:

Ilex; fruit morphology; electronic nose; principal component analysis; OPLS-DA.

1 Introduction

The genus *Ilex* (Aquifoliaceae) is among the most species-rich genera of angiosperms and constitutes the largest assemblage of dioecious woody plants worldwide. Approximately 600 wild *Ilex* taxa occur globally across tropical, subtropical, and temperate regions in both hemispheres, with major diversity concentrated in Central/South America and tropical Asia (Manen et al.2010). China is a key center of *Ilex* diversity, harboring 204 taxa that are distributed mainly on the southern slopes of the Qinling Mountains, throughout the Yangtze River basin, and further south; these taxa have been classified into three subgenera, nine sections, and 21 series (Li et al.2007). Extensive natural variation, cultivars, and hybrids have been documented within *Ilex*, supporting its broad use as an ornamental, nectar, and medicinal resource (Zheng et al.2022). Phytochemical studies indicate that *Ilex* leaves and fruits are rich in triterpenoid saponins, flavonoid glycosides, and caffeoylquinic acid derivatives, compounds associated with anti-inflammatory, antioxidant, and antitumor activities (e.g., via modulation of NF-κB signaling and free-radical scavenging) (Dat et al.2025). In horticulture and landscaping, evergreen taxa such as *Ilex cornuta* Lindl. et Paxt. and *Ilex latifolia* Thunb. are widely planted as hedges and winter ornamentals because of their leathery, pruning-tolerant foliage and persistent, brightly colored fruits (Zou et al.2024).

Beyond their economic and ornamental value, *Ilex* taxa are typical ornithochorous plants; their brightly colored, fleshy fruits attract avian frugivores and facilitate seed dispersal (Zou et al.2024). Nevertheless, taxonomic identification within *Ilex* is often challenging because of pronounced morphological convergence. Traditional classifications rely heavily on macromorphological characters, yet substantial intraspecific variation and environmental plasticity can result in frequent misidentification (Manen et al.2004). Field observations suggest that frugivores do not consume all *Ilex* fruits equally. Our preliminary surveys indicate that *Ilex rotunda* Thunb. has a markedly higher fruit consumption rate than *I. cornuta* Lindl. et Paxt. and *I. chinensis.* This pattern cannot be fully explained by visible defensive traits alone, suggesting that olfactory cues mediated by fruit volatile organic compounds (VOCs) may interact with visual signals to influence foraging behavior (Schaefer et al.2008).

Accordingly, VOCs represent a promising and biologically meaningful basis for discriminating *Ilex* taxa. In recent years, electronic nose (E-nose) technology—designed to mimic olfaction using a sensor array coupled with pattern recognition—has been widely applied for rapid, non-destructive identification of plant materials (Röck et al.2008), including applications in *Zanthoxylum bungeanum* (Zhenping et al.2023) and soybean (Song et al.2024). When coupled with chemometric methods such as PCA, linear discriminant analysis (LDA), and orthogonal projections to latent structures–discriminant analysis (OPLS-DA), E-nose fingerprints can be translated into interpretable classification models with quantitative assessment and validation. However, VOC-based discrimination is strengthened when integrated with stable phenotypic descriptors. Fruit morphology and peel color are measurable traits relevant to taxonomy; in particular, CIELAB coordinates (L*, a*, b*) and derived indices such as chroma (C*) enable standardized, objective quantification of fruit coloration.

In this study, five representative *Ilex* taxa from eastern China were selected to quantify fruit morphological and colorimetric traits and to characterize fruit VOC fingerprints using a PEN3 electronic nose, followed by PCA, LDA, and OPLS-DA modeling. To contextualize our discrimination results within an evolutionary framework, we compiled a simplified phylogenetic tree (Figure 1) summarizing hypothesized species relationships from prior phylogenetic studies (Yang et al.2023;Chen et al.2021). In this topology, *I. cornuta* and its cultivar *I. cornuta* ‘Fortunei’ form a closely related sister pair, whereas *I. rotunda* occupies a more distant, early-diverging branch relative to the remaining taxa. We therefore

hypothesize that, as the most phylogenetically divergent taxon, *I. rotunda* will exhibit more pronounced separation in both fruit phenotypic and VOC fingerprint spaces than the other taxa.

2 Materials and Methods

2.1 Plant materials, study site, and instruments

2.1.1 Plant materials and sampling site

Fruits of five *Ilex* taxa (*Ilex rotunda* Thunb., *I. chinensis*, *I. cornuta* Lindl. et Paxt., *I. cornuta* 'Fortunei', and *I. latifolia* Thunb.) were collected from the campus of Nanjing Forestry University (Xuanwu District, Nanjing, Jiangsu Province, China; 32.08° N, 118.82° E). The site has a subtropical monsoon climate with four distinct seasons, a mean annual temperature of 15.4 °C, annual precipitation of 1,000–1,100 mm, and an annual frost-free period of approximately 237 days.

2.1.2 Instruments

Volatile fingerprints were acquired using a PEN3 (PEN III) portable electronic nose equipped with 10 metal-oxide semiconductor (MOS) sensors (AIRSENSE Analytics GmbH, Germany). Fruit color was measured using a TS7010 colorimeter (Shenzhen Threenh Technology Co., Ltd., China). Single-fruit fresh weight was measured using an electronic balance (JA5003; readability, 0.001 g; Shanghai LiangPing Instruments Co., Ltd., China). Fruit longitudinal and transverse diameters were measured using digital calipers.

2.2 Experimental procedures

2.2.1 Electronic nose measurement

For each taxon, mature fruits collected from multiple trees were pooled to obtain a composite sample (Chen et al.2020). A 2.0 g aliquot of fruit material was placed in a 50 mL headspace vial, sealed immediately (Laothawornkitkul et al.2008), and incubated for 30 min (Xia et al.2022) to allow headspace equilibration. Headspace volatiles were then analyzed using the PEN3 electronic nose.

Preliminary measurements indicated that sensor responses stabilized after approximately 70 s and remained essentially constant after 78 s; therefore, response values at 78 s were used as characteristic variables for subsequent analysis (Wang et al.2025). The instrument parameters were set as follows: flushing time, 80 s; zero-adjustment time, 5 s; sample preparation time, 5 s; headspace injection mode; injection time, 80 s; and dry air as the carrier gas at 300 mL/min (Grigioni et al.2004). Each taxon was measured in 10 technical replicates.

2.2.2 Determination of fruit quality traits

Healthy, mature fruits free of visible pest damage, disease symptoms, and mechanical injury were selected from trees with uniform vigor. For each taxon, >200 fruits were harvested initially; after screening for uniform maturity and the absence of defects, 90 fruits per taxon were retained for trait measurement (Fei et al.2023). Fruits were organized into three biological replicates (30 fruits per tree). Single-fruit fresh weight (FW, g) was measured using the JA5003 balance. Fruit longitudinal diameter (LD, mm) and transverse diameter (TD, mm) were measured using digital calipers. The fruit shape index (FSI) was calculated as (Zhang et al.2022):

$$\mathrm{FSI} = \frac{\mathrm{LD}}{\mathrm{TD}}$$

Peel color was measured using the TS7010 colorimeter on smooth, intact surfaces free of blemishes or

wrinkles, ensuring that the measuring aperture was held flush against the fruit surface (Xu et al.2022). The colorimeter was calibrated against black and white reference standards before measurements. CIELAB coordinates were recorded as L* (0 = black, 100 = white), a* (− green to + red), and b* (− blue to + yellow). For each biological replicate, mean L*, a*, and b* values were calculated from 30 fruits, yielding three replicate means per taxon. Chroma (C*) was calculated as (McGuire et al.1992):

$$C^* = \sqrt{a^{*2} + b^{*2}}$$

For visualization, mean L*, a*, and b* values were converted to RGB values using Python (da Silva Souza et al.2026).

2.3 Data processing and statistical analysis

Raw data were compiled in Excel 2019. Prior to multivariate analysis of electronic-nose data, the sensor response matrix was standardized using Z-score normalization (autoscaling) to remove scale effects among sensor variables (Baietto et al.2015). Here, x_{ij} is the original response of sensor j in sample i, and μ_j and σ_j are the mean and standard deviation of sensor j across samples, respectively. After normalization, PCA and LDA were performed using WinMuster (AIRSENSE)(Chen et al.2022). Figures were generated in Origin 2022. For univariate comparisons of fruit quality traits among taxa, statistical significance was evaluated using one-way ANOVA followed by the least significant difference (LSD) post hoc test at $p < 0.05$ (Steel et al.1981).

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

3 Results

3.1 Differences in fruit morphological and colorimetric traits

All measured fruit traits differed significantly among the five Ilex taxa, indicating substantial phenotypic divergence in size, shape, weight, and peel color. Table 2 summarizes eight quantitative traits, including transverse diameter (TD), longitudinal diameter (LD), single-fruit weight (FW), fruit shape index (FSI = LD/TD), CIELAB coordinates (L*, a*, b*), and chroma (C*). Different lowercase letters following the mean values denote significant pairwise differences among taxa (LSD post hoc test, $p < 0.05$).

3.1.1 Fruit size and weight

Fruit dimensions showed clear inter-taxon differentiation (Table 2). *I. cornuta* produced the largest fruits (TD, 7.62 ± 0.51 mm; LD, 8.83 ± 0.62 mm), significantly exceeding those of the other taxa. *I. chinensis* ranked second (TD, 7.09 ± 0.48 mm; LD, 8.50 ± 0.57 mm). In contrast, *I. rotunda* had the smallest fruits (TD, 5.64 ± 0.39 mm; LD, 5.86 ± 0.41 mm). The remaining taxa showed intermediate values: *I. cornuta* 'Fortunei' (TD, 6.79 ± 0.43 mm; LD, 7.91 ± 0.53 mm) and *I. latifolia* (TD, 6.84 ± 0.45 mm; LD, 7.08 ± 0.49 mm), with significant differences among groups indicated by the post hoc lettering. Single-fruit weight differed significantly among taxa ($p < 0.001$), ranging from 0.183 ± 0.019 g in *I. latifolia* to 0.313 ± 0.035 g in *I. cornuta* (Table 2). *I. cornuta* had the highest FW (0.313 ± 0.035 g), followed by *I. chinensis* (0.231 ± 0.028 g). *I. rotunda* (0.203 ± 0.022 g) and *I. cornuta* 'Fortunei' (0.196 ± 0.021 g) formed an intermediate-low group, whereas *I. latifolia* had the lowest FW. Notably, FW did not strictly mirror the diameter ranking across taxa: although I. latifolia showed intermediate TD and LD, it had the lowest mass, indicating that fruit size and mass were not perfectly coupled among the five taxa.

I. cornuta shows the highest central tendency (median ≈ 0.31 g), with most observations concentrated at higher FW values, whereas *I. latifolia* exhibits a lower and relatively narrow distribution centered at approximately 0.18–0.19 g. *I. cornuta* 'Fortunei' presents a broader violin and longer tails, suggesting greater within-taxon variability in fruit mass than the other taxa. Occasional high-FW observations are also evident in *I. cornuta*, consistent with its higher mean FW reported in Table 2.

3.1.2 Fruit shape index (FSI)

FSI (LD/TD) differed significantly among taxa ($p < 0.001$), reflecting systematic variation in fruit elongation (Table 2). *I. chinensis* exhibited the highest FSI (1.200 ± 0.062), indicating the most elongated fruits. *I. cornuta* (1.161 ± 0.055) and *I. cornuta* 'Fortunei' (1.169 ± 0.058) formed an intermediate group. By contrast, *I. rotunda* (1.038 ± 0.049) and *I. latifolia* (1.036 ± 0.047) had the lowest FSI, consistent with a more nearly spheroidal fruit shape.

The FSI raincloud plot corroborates these group-level differences while highlighting within-taxon dispersion. *I. chinensis* shows an upward-shifted distribution centered at approximately 1.20 with relatively concentrated density, indicating consistently elongated fruits. *I. rotunda* and *I. latifolia* cluster tightly around 1.03–1.05, suggesting relatively uniform fruit shape. By contrast, *I. cornuta* 'Fortunei' exhibits a broader violin and longer tails, indicating greater within-taxon heterogeneity in FSI and occasional extreme values, although its mean FSI remains within the intermediate group (Table 2).

3.1.3 Color parameters

Colorimetric traits differed markedly among taxa (all $p < 0.001$; Table 2). Lightness (L*) was highest in *I. chinensis* (45.26 ± 2.31) and lowest in *I. latifolia* (40.55 ± 2.18). Redness (a*) showed the

strongest differentiation: *I. rotunda* had the highest a* (52.90 ± 3.25), whereas *I. latifolia* had the lowest (35.92 ± 2.87). Yellowness (b*) followed a similar trend, with *I. rotunda* highest (33.24 ± 2.98) and *I. latifolia* lowest (22.16 ± 2.53). Consequently, chroma (C*), which reflects overall color saturation, was highest in *I. rotunda* (62.70 ± 3.52) and lowest in *I. latifolia* (42.21 ± 3.15). *I. chinensis* (55.83 ± 3.09) and *I. cornuta* (56.28 ± 3.12) formed an intermediate group, whereas *I. cornuta* 'Fortunei' (52.75 ± 3.15) showed lower chroma.

The chroma raincloud plot visually reinforces the pronounced inter-taxon gradient in color saturation. *I. rotunda* shows the highest C* values overall, whereas *I. latifolia* exhibits substantially lower C* values with a relatively tight distribution. *I. chinensis* and *I. cornuta* display compact violins with dense point clustering near their medians, indicating comparatively stable chroma within these taxa. By contrast, *I. rotunda* shows greater dispersion and a pronounced upper tail of high C* observations, suggesting higher within-taxon heterogeneity in saturation than the other taxa.

3.2 Electronic Nose Sensor Responses to Different *Ilex* Taxa

To visualize inter-taxon differences in volatile fingerprints at the sensor level, the steady-state responses of the 10 PEN3 MOS sensors—each primarily responsive (putatively) to specific classes of odor-active volatiles (Table 1)—were summarized using a radar plot (Figure 5A). Among the examined taxa, *I. rotunda* exhibited the most distinctive sensor-response pattern. Specifically, *I. rotunda* showed a pronounced elevation in W3C (aromatic and ammonia-related volatiles) and a higher W1C signal (aromatic compounds/benzenes), suggesting a greater relative contribution of these volatile classes to its headspace profile. Conversely, W2W (sulfur- and chlorine-containing organics) and W1W (sulfur-containing organics) exhibited lower responses for *I. rotunda*, implying a lower relative contribution of these compound classes to its headspace profile.

In contrast, *I. chinensis*, *I. cornuta*, *I. cornuta* 'Fortunei', and *I. latifolia* showed relatively flat and largely overlapping profiles across most sensor axes, including those putatively sensitive to nitrogen oxides (W5S), hydrogen-containing compounds (W6S), alkanes (W1S), short-chain alkanes and aromatics (W5C), aliphatic alkanes (W3S), and alcohols/aldehydes/ketones (W2S). This similarity suggests broadly comparable VOC compositions and relative abundances among these taxa. Notably, the response curves of *I. cornuta* and its cultivar *I. cornuta* 'Fortunei' were nearly superimposable across all sensors, consistent with their close phylogenetic relationship and indicating limited divergence in major volatile classes. Taken together, these sensor-response fingerprints highlight *I. rotunda* as the most chemically distinct taxon and provide an interpretable link between sensor outputs and compound classes, thereby informing subsequent chemometric discrimination.

To complement the radar plot, normalized multivariate sensor responses were visualized as a heatmap, in which color intensity encodes relative response magnitude. In Figure 5B, columns correspond to samples and rows correspond to sensors; each cell represents the normalized response of a given sensor for a given sample. Warmer colors denote higher normalized responses, whereas cooler colors denote lower responses. Consistent with the radar plot, *I. rotunda* exhibits a markedly different response pattern relative to the other taxa, indicating a distinct volatile signature and supporting the development of an identification model for this taxon.

3.3 Principal Component Analysis

Principal component analysis (PCA) is a widely used dimensionality reduction method for multivariate data. It linearly transforms high-dimensional variables into a smaller set of orthogonal

principal components (PCs), facilitating visualization while retaining most of the variance (Abdi et al.2010;Wold et al.1987). In this study, PCA was applied to the standardized E-nose sensor matrix to explore inter-taxon variation in VOC fingerprints.The PCA score plot revealed taxon-dependent clustering and separation patterns. *I. rotunda* was clearly separated from the other taxa, indicating pronounced differences in odor fingerprints and supporting discrimination based on electronic-nose responses. Moreover, *I. rotunda* samples formed a tight cluster, suggesting high repeatability and low within-taxon variation in odor fingerprints. Although *I. chinensis* did not overlap with *I. rotunda*, its confidence ellipse partially overlapped with those of *I. cornuta* and *I. cornuta* 'Fortunei', indicating similar odor fingerprints among these taxa. The confidence ellipses of *I. cornuta*, *I. cornuta* 'Fortunei', and *I. latifolia* showed substantial overlap, suggesting high similarity and limited discrimination by PCA alone. Because PCA is unsupervised and does not incorporate class labels, within-group dispersion remains evident (Figure 6) (Eriksson et al.2013). Therefore, additional supervised analyses were performed to improve discrimination among taxa.

3.4 Linear Discriminant Analysis

Linear discriminant analysis (LDA) was applied to electronic-nose sensor response data (10 sensors) from five *Ilex* taxa. LDA is a supervised method that maximizes between-class separation while minimizing within-class dispersion in the projected space. Specifically, LDA projects samples into a low-dimensional discriminant space such that within-class samples cluster tightly and class centroids are well separated (Fisher et al.1936; Marsili et al.2001). Figure 7 shows the LDA score plot based on electronic-nose volatile fingerprints for the five *Ilex* taxa. The first three linear discriminants (LD1, LD2, and LD3) accounted for ~100% of the discriminant variance (71.83%, 10.50%, and 17.87%, respectively), indicating that the 3D space captured essentially all discriminatory information among taxa. Compared with unsupervised PCA, LDA incorporating class labels further improved separation between *I. rotunda* and the remaining taxa. In addition, *I. rotunda* samples formed a tight cluster within the confidence ellipse, indicating high within-taxon consistency in volatile fingerprints. Overall, the electronic-nose–LDA approach enabled reliable discrimination of *I. rotunda* from the other taxa and improved separation among closely related *Ilex* taxa, supporting its use as a rapid, non-destructive identification tool.

3.5 Construction of OPLS-DA Identification Model

To develop a parsimonious identification model for *I. rotunda*, orthogonal projections to latent structures–discriminant analysis (OPLS-DA) was applied for supervised multivariate classification. Using the preprocessed electronic-nose response matrix (X) and a binary class vector (Y; *I. rotunda* vs. non-target taxa), we built an OPLS-DA model to discriminate *I. rotunda* from the other taxa. OPLS-DA is a supervised extension of partial least squares (PLS) commonly used for class discrimination. It decomposes X into predictive components correlated with Y and orthogonal components uncorrelated with Y (Bylesjö et al.2006). This decomposition enhances interpretability by emphasizing class-related variation and reducing structured noise, thereby improving group separation (Yang et al.2024).

The OPLS-DA score plot (Figure 8) shows that *I. rotunda* and non-target taxa formed two clearly separated clusters, primarily along the predictive component p1 ($R^2X = 0.657$). The orthogonal component o1 captured non-class-related variation ($R^2X = 0.128$), which may reflect systematic effects unrelated to class assignment (e.g., sampling or batch variation). The 95% confidence ellipses did not overlap, indicating robust multivariate separation between classes. The *I. rotunda* cluster was more compact, suggesting lower within-class variability in its volatile fingerprint. Because the non-target class

comprised multiple taxa, it showed greater dispersion; nevertheless, it remained clearly separated from *I. rotunda* along the predictive axis, supporting the specificity of the *I. rotunda* sensor signature under multi-taxon conditions.

The OPLS-DA model included one predictive component (p1) and one orthogonal component (o1), separating class-related variation from structured variation unrelated to class assignment (Trygg et al.2002). Model performance was assessed using R^2X, R^2Y, and the cross-validated predictive ability (Q^2). Permutation testing (n = 200) was used to assess model robustness and potential overfitting. Here, Q^2 reflects predictive performance, whereas R^2 reflects explained variance. Q^2 was calculated by cross-validation as defined below (Bylesjö et al.2006):

$$R^2X = 1 - \frac{\mathrm{SS}(\boldsymbol{E})}{\mathrm{SS}(\boldsymbol{X}_{\mathrm{cnee}})}, R^2Y = 1 - \frac{\mathrm{SS}(\boldsymbol{F})}{\mathrm{SS}(\boldsymbol{Y}_{\mathrm{cnee}})}$$

$$Q^2 = 1 - \frac{\mathrm{SS}(\boldsymbol{F}_{\mathrm{C}})}{\mathrm{SS}(\boldsymbol{Y}_{\mathrm{cnee}})}$$

Cross-validation indicated strong model fit and predictive performance (R^2X = 0.785, R^2Y = 0.944, Q^2 = 0.935). The predictive component p1 explained 65.7% of X variation and 86.8% of Y variation, whereas o1 explained 12.8% of X variation and 7.57% of Y variation, indicating that discrimination was driven primarily by the predictive component and that non-class-related variation was effectively isolated. In addition, $R^2Y - Q^2$ = 0.009 (< 0.2), indicating no strong indication of overfitting and good agreement between fit and predictive ability.

Model robustness was evaluated using 200 random permutations of the class labels. The original model (R^2Y = 0.944; Q^2 = 0.935) showed higher performance than all permuted models. The permuted models exhibited low R^2Y and Q^2 values (with some negative Q^2), and permutation testing indicated $p < 0.005$. These results indicate that the observed separation is unlikely to arise by chance and that the model captures reproducible multivariate differences in sensor fingerprints between I. rotunda and the other taxa, with low evidence of overfitting.

In the OPLS-DA model, the variable importance in projection (VIP) quantifies the contribution of each predictor variable to class discrimination. Variables with VIP > 1 are commonly considered major contributors, and larger VIP values indicate greater influence on model discrimination. VIP values were calculated using the standard definition as follows (Trygg et al.2002):

$$\mathrm{VIP}_j = \sqrt{\frac{p}{h}\sum_{f=1}^{h}\left(w_{jf}^2 \cdot \mathrm{SS}(\boldsymbol{T}_f)\right) / \sum_{f=1}^{h}\mathrm{SS}(\boldsymbol{T}_f)}$$

As shown in Figure 11, VIP scores were computed for all 10 sensors. Using VIP > 1 as the criterion, six sensors were identified as primary contributors, whereas the remaining four showed lower contributions. The primary contributors were W5C (1.21), W3C (1.20), W1C (1.20), W2W (1.19), W3S (≈1.14–1.15), and W1W (≈1.08–1.10), with VIP values ranging from 1.08 to 1.21. The remaining sensors (W1S, W5S, W2S, and W6S) had VIP < 1 and contributed less to discrimination. These results suggest that discrimination of I. rotunda relies on a multivariate sensor signature rather than a single sensor response. The combination of higher- and lower-contributing sensors forms a characteristic fingerprint that may improve robustness compared with relying on any single feature. These high-VIP sensors provide candidates for developing targeted marker features in future studies.

The statistical robustness of the OPLS-DA model for *I. rotunda* discrimination was further evaluated using 200 random permutations of the class labels (Figure 12). As label similarity decreased from 1 (original) to 0 (random permutation), permuted models showed R^2Y concentrated around 0–0.5

and Q^2 values < 0, indicating markedly reduced performance compared with the original model. These results indicate that model performance is unlikely to be driven by chance and supports the presence of reproducible multivariate differences in sensor fingerprints between *I. rotunda* and the other taxa. Overall, these findings support the statistical reliability of the OPLS-DA model and provide methodological support for targeted identification of *I. rotunda* using high-VIP sensors.

4 Conclusion

An integrated framework combining fruit morphometrics with electronic-nose chemometrics reliably discriminated five *Ilex* taxa from eastern China, and the trait-space patterns strongly corroborated the preset molecular phylogenetic framework (Figure 1). Phenotypically and in VOC profiles, *I. rotunda* showed the greatest divergence from other taxa—consistent with its early-diverging position in the phylogeny—while *I. cornuta* and its cultivar *I. cornuta* 'Fortunei' exhibited minimal differentiation in fruit traits (e.g., shape index) and nearly superimposable sensor responses, echoing their close evolutionary relationship. All fruit traits differed significantly among taxa ($p < 0.001$): *I. cornuta* had the largest fruits and highest weight, *I. chinensis* the most elongated shape, and *I. rotunda* the highest redness (a*) and chroma (C*). E-nose fingerprints complemented morphological data: PCA explained 90.21% of variance, LDA improved separation of closely related taxa, and the OPLS-DA model for *I. rotunda* achieved strong performance ($R^2X = 0.785$, $R^2Y = 0.944$, $Q^2 = 0.935$) with no overfitting (permutation test, $p < 0.005$). Key VIP sensors (W5C, W3C, W1C, W2W, W3S, W1W) formed a multivariate signature for reliable *I. rotunda* identification. This rapid, non-destructive approach supports germplasm management and field surveys; future work should expand taxon sampling, integrate molecular validation, and annotate discriminatory VOCs to develop targeted biosensors and clarify trait divergence mechanisms in *Ilex.*

Author Contributions

Ye Peng conceived the study. Meng Sun acquired the fund. Xiangxian Fan conducted the sampling and conducted experiments. Xiangxian Fan carried out bioinformatics analysis and drafted the manuscript. Ye Peng and Meng Sun revised the manuscript. All authors approved the final manuscript.

Funding:

Study on the Breeding and Propagation of Ilex Varieties（BY20240060）

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459. https://doi.org/10.1002/wics.101

Baietto, M., & Wilson, A. D. (2015). Electronic-nose applications for fruit identification, ripeness and quality grading. Sensors, 15(1), 899-931. https://doi.org/10.3390/s150100899

Barsan, N., Röck, F., & Weimar, U. (2008). Electronic nose: current status and future trends. Chemical reviews, 108(2), 705-725. https://doi.org/10.1021/cr068121q

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics: A Journal of the Chemometrics Society, 20(8-10), 341-351. https://doi.org/10.1002/cem.1006

Chen, H., Chong, X., Zhou, Y., Sun, L., Zhang, F., Lu, X., ... & Li, Y. (2021). Characterization and phylogenetic analysis of the complete chloroplast genome of Ilex rotunda, traditional Chinese medicine plant. Mitochondrial DNA Part B, 6(10), 2791-2792. https://doi.org/10.1080/23802359.2021.1967804

Chen, S., Qin, W., Guo, Z., Li, R., Ding, C., Zhang, S., & Tan, Z. (2020). Metabonomics study of fresh bruises on an apple using the gas chromatography–mass spectrometry (GC–MS) method. European Food Research and Technology, 246(1), 201-212. https://doi.org/10.1007/s00217-019-03386-x

Chen, Y., Fu, Y., Li, P., Xi, H., Zhao, W., Wang, D., ... & Xie, J. (2022). Characterization of traditional Chinese sesame oil by using headspace solid-phase microextraction/gas chromatography–mass spectrometry, electronic nose, sensory evaluation, and RapidOxy. Foods, 11(22), 3555. https://doi.org/10.3390/foods11223555

da Silva Souza, J., Uneda-Trevisoli, S. H., Dalla Lana, F., & Fritsche-Neto, R. (2026). A Low-Cost RGB-Based Image Processing Method for High-Throughput Assessment of Rice Grain Chalkiness. Rice Science, 24. http://www.ricesci.org/CN/

Dat Le, D., Truong, V., Dang, T., Yu, S., Dinh, T., Thao, N. P., & Lee, M. (2025). Triterpenoids From Ilex rotunda and Their Anti-Inflammatory Effects by Interfering With NF-κB and MAPK Activation. Archiv der Pharmazie, 358(5), e12003. https://doi.org/10.1002/ardp.202500025

Eriksson, L., Byrne, T., Johansson, E., Trygg, J., & Vikström, C. (2013). Multi-and megavariate data analysis basic principles and applications (Vol. 1). Umetrics Academy.

Fei, L. I., Peng, Z. H. O. U., Jing, H. U. A. N. G., Min, Z. H. A. N. G., & Jing, W. U. (2023). Fruit color and its coloring mechanism of two Ilex verticillata cultivars. Journal of Zhejiang Forestry Science and Technology, 43(4), 34-41. https://dx.doi.org/10.3969/j.issn.1001-3776.2023.04.005

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of eug enics, 7(2), 179-188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Grigioni, G., Carduza, F., Irurueta, M., & Pensel, N. (2004). Flavour characteristics of Ilex par aguariensis infusion, a typical Argentine product, assessed by sensory evaluation and electr onic nose. Journal of the Science of Food and Agriculture, 84(5), 427-432. https://doi.org/1 0.1002/jsfa.1670

Laothawornkitkul, J., Moore, J. P., Taylor, J. E., Possell, M., Gibson, T. D., Hewitt, C. N., & Paul, N. D. (2008). Discrimination of plant volatile signatures by an electronic nose: a pot ential technology for plant pest and disease monitoring. Environmental science & technolo gy, 42(22), 8433-8439. https://doi.org/10.1021/es801738s

Li, J. (2007). Flora of China. Harvard Papers in Botany, 13(2), 301-302. https://doi.org/10.3100 /1043-4534-13.2.301

Manen, J. F. (2004). Are both sympatric species Ilex perado and Ilex canariensis secretly hybri dizing? Indication from nuclear markers collected in Tenerife. BMC Evolutionary Biology, 4(1), 46. https://doi.org/10.1186/1471-2148-4-46

Manen, J. F., Barriera, G., Loizeau, P. A., & Naciri, Y. (2010). The history of extant Ilex spe cies (Aquifoliaceae): evidence of hybridization within a Miocene radiation. Molecular phylo genetics and evolution, 57(3), 961-977. https://doi.org/10.1016/j.ympev.2010.09.006

Marsili, R. (2001). Combining mass spectrometry and multivariate analysis to make a reliable a nd versatile electronic nose. In Flavor, Fragrance, and Odor Analysis (pp. 365-390). CRC Press.

McGuire, R. G. (1992). Reporting of objective color measurements.

Röck, F., Barsan, N., & Weimar, U. (2008). Electronic nose: current status and future trends. Chemical reviews, 108(2), 705-725. https://doi.org/10.1021/cr068121q

Schaefer, H. M., McGraw, K., & Catoni, C. (2008). Birds use fruit colour as honest signal of dietary antioxidant rewards. Functional Ecology, 22(2), 303-310. https://doi.org/10.1111/j.136 5-2435.2007.01363.x

Song, X., Qian, L., Zhang, D., Wang, X., Fu, L., & Chen, M. (2024). Effectiveness of Differe ntiating Mold Levels in Soybeans with Electronic Nose Detection Technology. Foods, 13(2 4), 4064. https://doi.org/10.3390/foods13244064

Steel, R. G. D., & Torrie, J. H. (1981). Principles and procedures of statistics, a biometrical a pproach (No. Ed. 2, p. 633pp).

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics: A Journal of the Chemometrics Society, 16(3), 119-128. https://doi.org/10.1 002/cem.695

Wang, B., & Zhang, L. (2025). An electronic nose system coupled with an effective deep lear ning method to identify the different storage periods of peanuts. Journal of Food Science and Technology, 1-9. https://doi.org/10.1007/s13197-025-06293-0

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9

Xia, Y., Zha, M., Liu, H., Shuang, Q., Chen, Y., & Yang, X. (2022). Novel insight into the f ormation of odour—active compounds in sea buckthorn wine and distilled liquor based on GC–MS and E–nose analysis. Foods, 11(20), 3273. https://doi.org/10.3390/foods11203273

Xu, K. W., Wei, X. F., Lin, C. X., Zhang, M., Zhang, Q., Zhou, P., ... & Duan, Y. F. (2022). The chromosome-level holly (Ilex latifolia) genome reveals key enzymes in triterpenoid saponin biosynthesis and fruit color change. Frontiers in Plant Science, 13, 982323. https://doi.org/10.3389/fpls.2022.982323

Yang, X., Li, L., Yu, X., Liu, Y., Xie, S., Zhu, G., ... & Zhao, P. (2024). An untargeted metabolomics analysis of the components of heartwood and sapwood in 4 fast-growing Chinese fir (Cunninghamia lanceolata (Lamb.) Hook) clones. Trees, 38(2), 339-356. https://doi.org/10.1007/s00468-023-02486-2

Yang, Y., Jiang, L., Liu, E. D., Liu, W. L., Chen, L., Kou, Y. X., ... & Peng, H. (2023). Time to update the sectional classification of Ilex (Aquifoliaceae): New insights from Ilex phylogeny, morphology, and distribution. Journal of Systematics and Evolution, 61(6), 1036-1046. https://doi.org/10.1111/jse.12935

Zhenping, L. I. U., ZHANG, D., & HUANG, Y. (2023). Identification for Hanyuan Zanthoxylum bungeanum by electronic nose technique combined with stoichiometric analysis. Shipin gongye ke-ji, 44(16), 306-312.

Zheng, X., Wang, J., Zhou, J., & Wang, D. (2022). The extract of ilex cornuta bark promotes bone healing by activating adenosine A2A receptor. Drug Design, Development and Therapy, 2569-2587. https://doi.org/10.2147/DDDT.S362238

Zou, Y., Huang, Y., Zhang, D., Chen, H., Liang, Y., Hao, M., & Yin, Y. (2024). Molecular insights into a non-lethal yellow bud mutant in Ilex×‘Nellie R. Stevens’. Scientia Horticultu rae, 329, 113033. https://doi.org/10.1016/j.scienta.2024.113033

Zou, Y., Huang, Y., Zhang, D., Chen, H., Liang, Y., Hao, M., & Yin, Y. (2024). Molecular mechanisms of chlorophyll deficiency in Ilex× attenuata ‘Sunny Foster’mutant. Plants, 13(10), 1284. https://doi.org/10.3390/plants13101284

Tables:

Table 1. Types and performance of electronic nose sensors

Table 2. Fruit morphological and colorimetric traits of five *Ilex* taxa.

Figures:

Figure 1. Simplified phylogenetic relationships among the five *Ilex* taxa

Figure 2. Raincloud plot of single fruit weight (g) across five Ilex taxa.

Figure 3. Raincloud plot of fruit shape index (LD/TD).

Figure 4. Raincloud plot of chroma (C*).

Figure 5. A. Radar plot of electronic-nose sensor response fingerprints for five Ilex taxa. B. Heatmap with hierarchical clustering of normalized electronic-nose sensor responses.

Figure 6. PCA score plot of electronic nose sensor response data for different Ilex taxa

Figure 7. LDA score plot of electronic-nose sensor responses for the five Ilex taxa.

Figure 8. OPLS-DA score plot for discrimination of Ilex rotunda and other Ilex taxa. Samples form two distinct clusters primarily along the predictive component p1 ($R^2X = 0.657$), whereas o1 captures orthogonal variation ($R^2X = 0.128$). The separation indicates stable multivariate differences in electronic-nose sensor fingerprints between I. rotunda and non-target taxa.

Figure 9. Summary plot of cross-validation for the OPLS-DA model discriminating Ilex rotunda from non-target taxa. The model comprises one predictive component (p1) and one orthogonal component (o1). For p1, $R^2X = 0.657$, $R^2Y = 0.868$, and $Q^2 = 0.865$; for o1, $R^2X = 0.128$, $R^2Y = 0.0757$, and $Q^2 = 0.0698$. Overall, the results indicate strong explained variance and high cross-validated predictive performance.

Figure 10. Permutation test plot of the OPLS-DA model discriminating Ilex rotunda from other Ilex taxa ($n = 200$). The plot shows the distribution of R^2Y and Q^2 for permuted models after random shuffling of class labels, with vertical markers indicating the original model. The original model ($R^2Y = 0.944$; $Q^2 = 0.935$) outperformed all permuted models ($p < 0.005$), supporting model robustness ($R^2X = 0.785$).

Figure 11. Variable importance in projection (VIP) plot of the OPLS-DA model for Ilex rotunda discrimination. Variables are sorted by VIP scores; higher and lower contributions indicate inter-group response trends based on model coding. Features with VIP > 1 (e.g., W5C, W3C, W1C, W2W, W3S, W1W) are major contributors and can serve as candidate marker features for targeted identification.

Figure 12. Permutation test results of the OPLS-DA model ($n = 200$).

490 **Table 1.** Types and performance of electronic nose sensors.

Number	Sensor	Response-sensitive flavor compound categories
1	W1C	Aromatic compounds (benzenes)
2	W5S	Nitrogen oxides
3	W3C	Aromatic compounds; ammonia-related compounds
4	W6S	Hydrogen-containing compound
5	W5C	Short-chain alkanes; aromatic compounds
6	W1S	Alkanes
7	W1W	Sulfur-containing organic compounds
8	W2S	Alcohols, aldehydes and ketones
9	W2W	Sulfur-containing, chlorine-containing organic compounds
10	W3S	Aliphatic alkane compounds

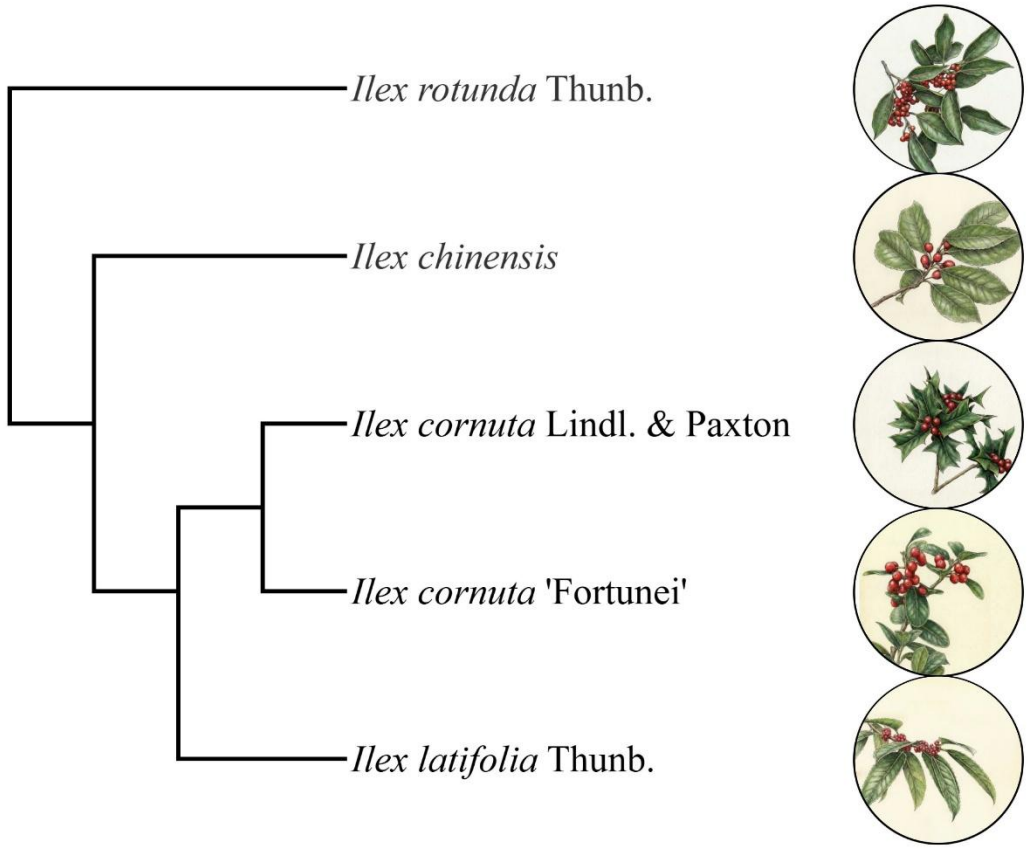
491

Table 2. Fruit morphological and colorimetric traits of five *Ilex* taxa.

	I. rotunda Thunb.	*I. chinensis*	*I. cornuta* Lindl. et Paxt.	*I. cornuta* 'Fortunei'	*I. latifolia* Thunb.	p-value
Transverse diameter(mm)	5.64±0.39d	7.09±0.48b	7.62±0.51a	6.79±0.43c	6.84±0.45c	<0.001
Longitudinal diameter(mm)	5.86±0.41e	8.50±0.57b	8.83±0.62a	7.91±0.53c	7.08±0.49d	<0.001
Single fruit weight(g)	0.203±0.022c	0.231±0.028b	0.313±0.035a	0.196±0.021c	0.183±0.019d	<0.001
Fruit shape index	1.038±0.049c	1.200±0.062a	1.161±0.055b	1.169±0.058b	1.036±0.047c	<0.001
Lightness (L)	41.82±2.21c	45.26±2.31a	42.58±2.25b	43.26±2.19b	40.55±2.18d	<0.001
Redness (a)	52.90±3.25a	46.89±2.41c	48.23±2.36b	45.32±2.38d	35.92±2.87e	<0.001
Yellowness (b)	33.24±2.98a	29.35±2.23b	28.86±2.15b	27.68±2.19c	22.16±2.53d	<0.001
Chroma (C)	62.70±3.52a	55.83±3.09b	56.28±3.12b	52.75±3.15c	42.21±3.15d	<0.001

Values are presented as mean ± SD. Different lowercase letters within a row indicate significant differences among taxa (LSD post hoc test, $p < 0.05$). All traits showed significant inter-taxon differences (one-way ANOVA, $p < 0.001$).

Figure 1. Simplified phylogenetic relationships among the five Ilex taxa



Simplified phylogenetic relationships among the five studied Ilex taxa, compiled from published phylogenetic inferences. This schematic tree illustrates the hypothesized relationships: I. cornuta and its cultivar I. cornuta ‘Fortunei’ form a sister clade, whereas I. rotunda occupies a relatively distant, early-diverging branch. Branch lengths are schematic and do not represent actual evolutionary distances.

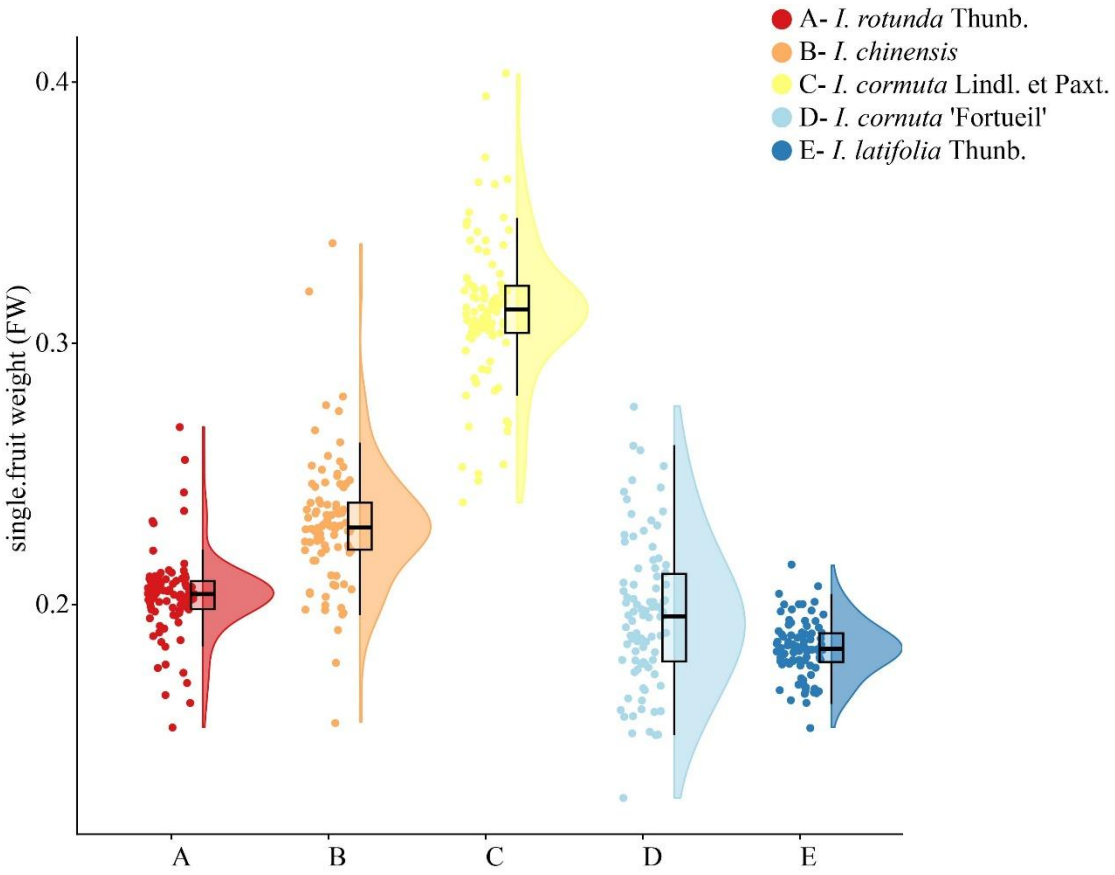


Figure 2. Raincloud plot of single fruit weight (g) across five Ilex taxa.

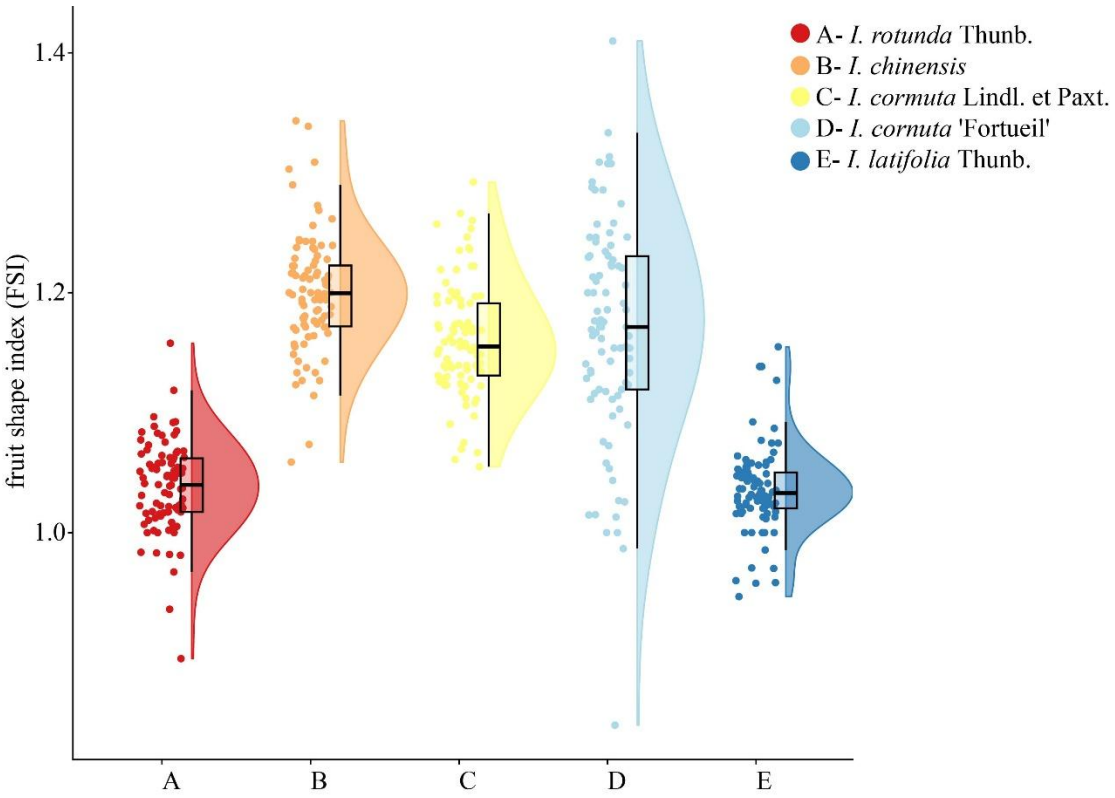


Figure 3. Raincloud plot of fruit shape index (LD/TD).

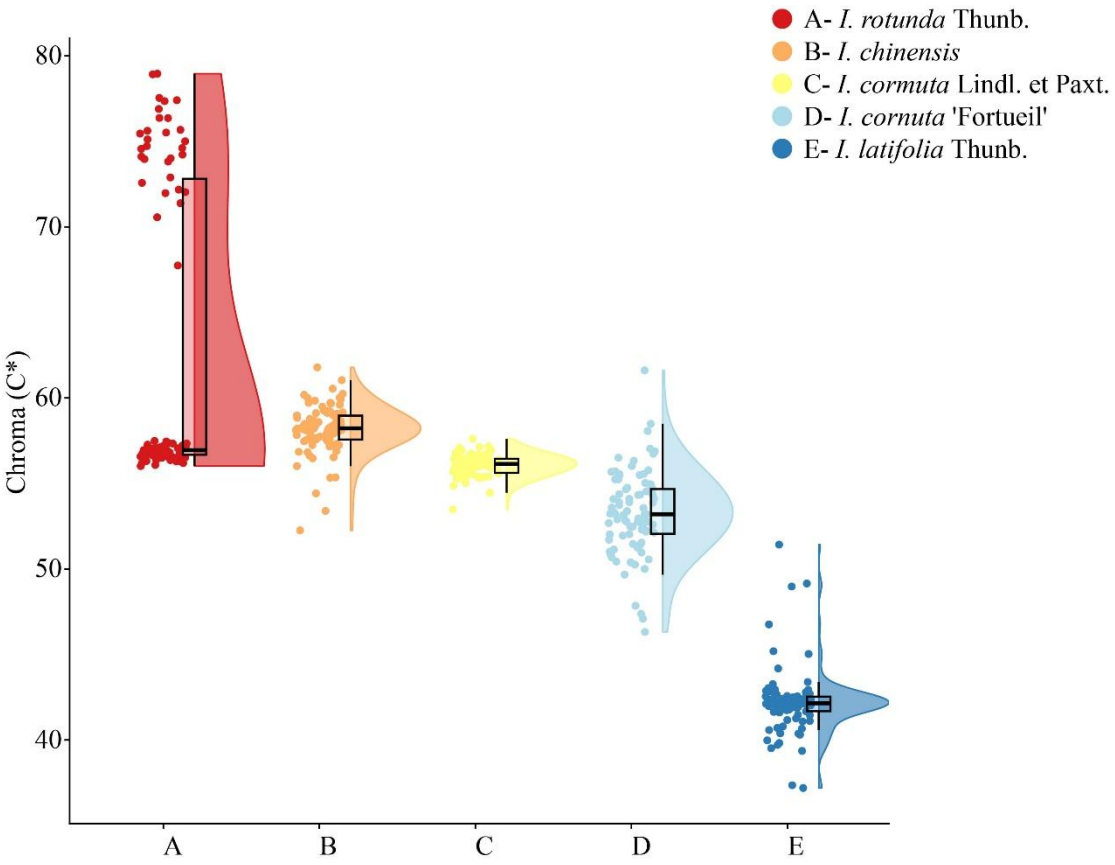


Figure 4. Raincloud plot of chroma (C*).

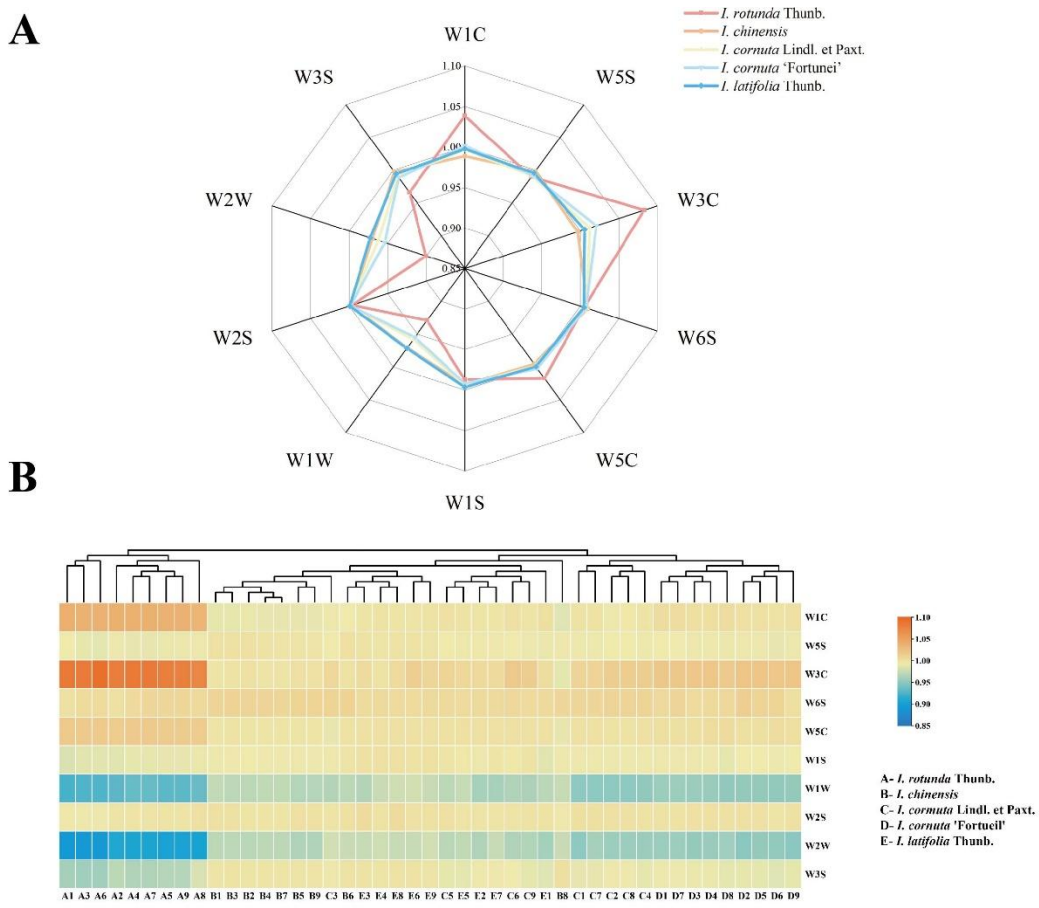


Figure 5 A. Radar plot of electronic-nose sensor response fingerprints for five Ilex taxa. B. Heatmap with hierarchical clustering of normalized electronic-nose sensor responses.

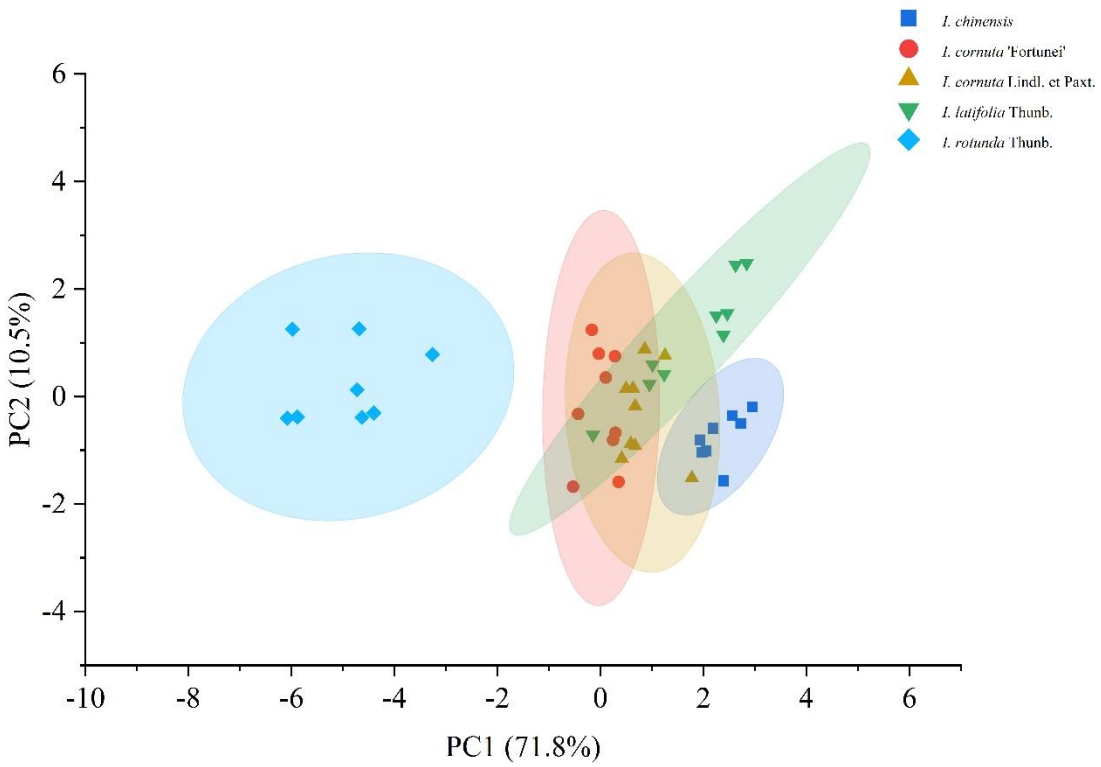


Figure 6. PCA score plot of electronic nose sensor response data for different Ilex taxa.

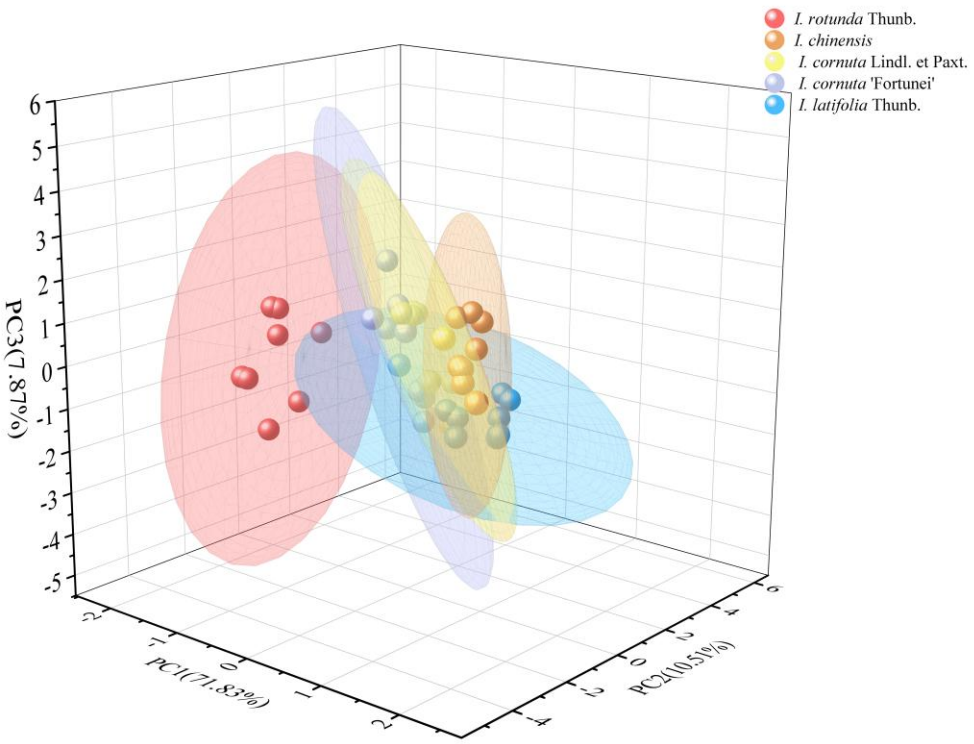
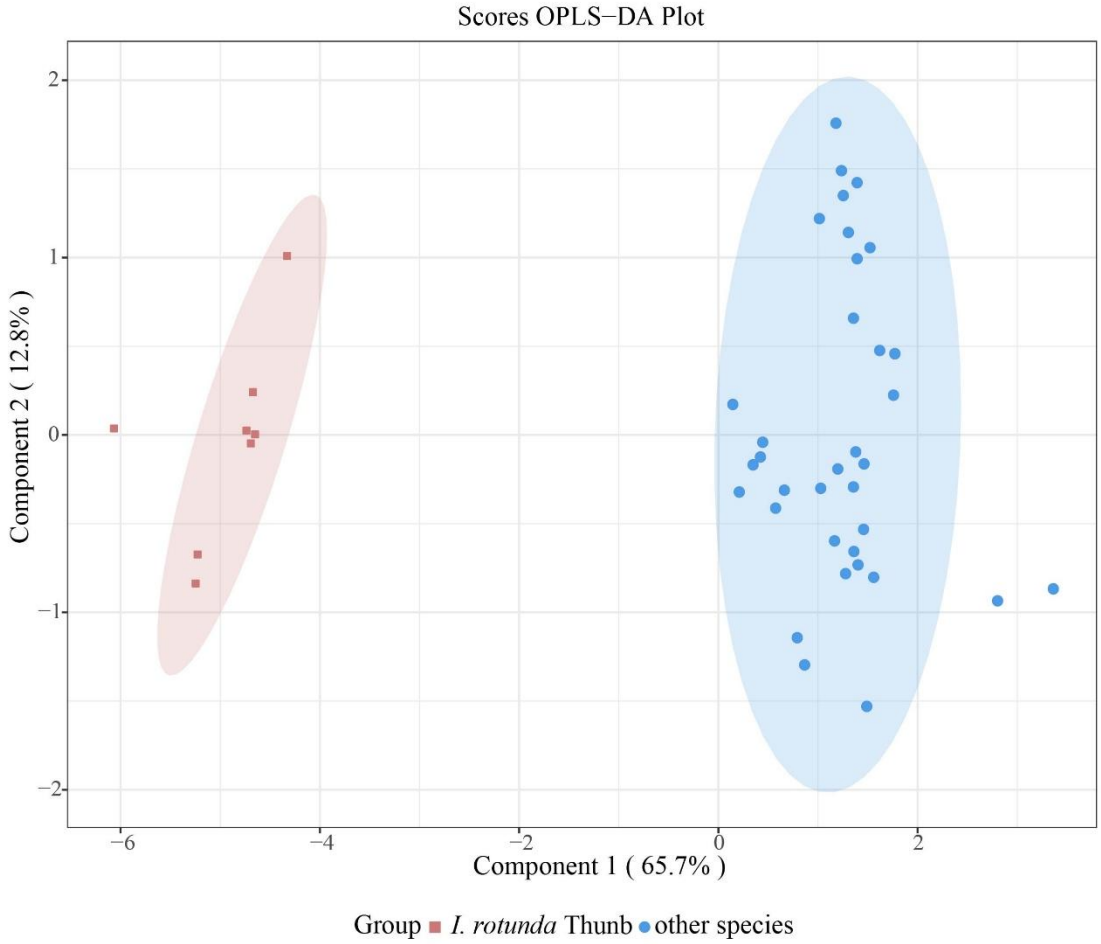


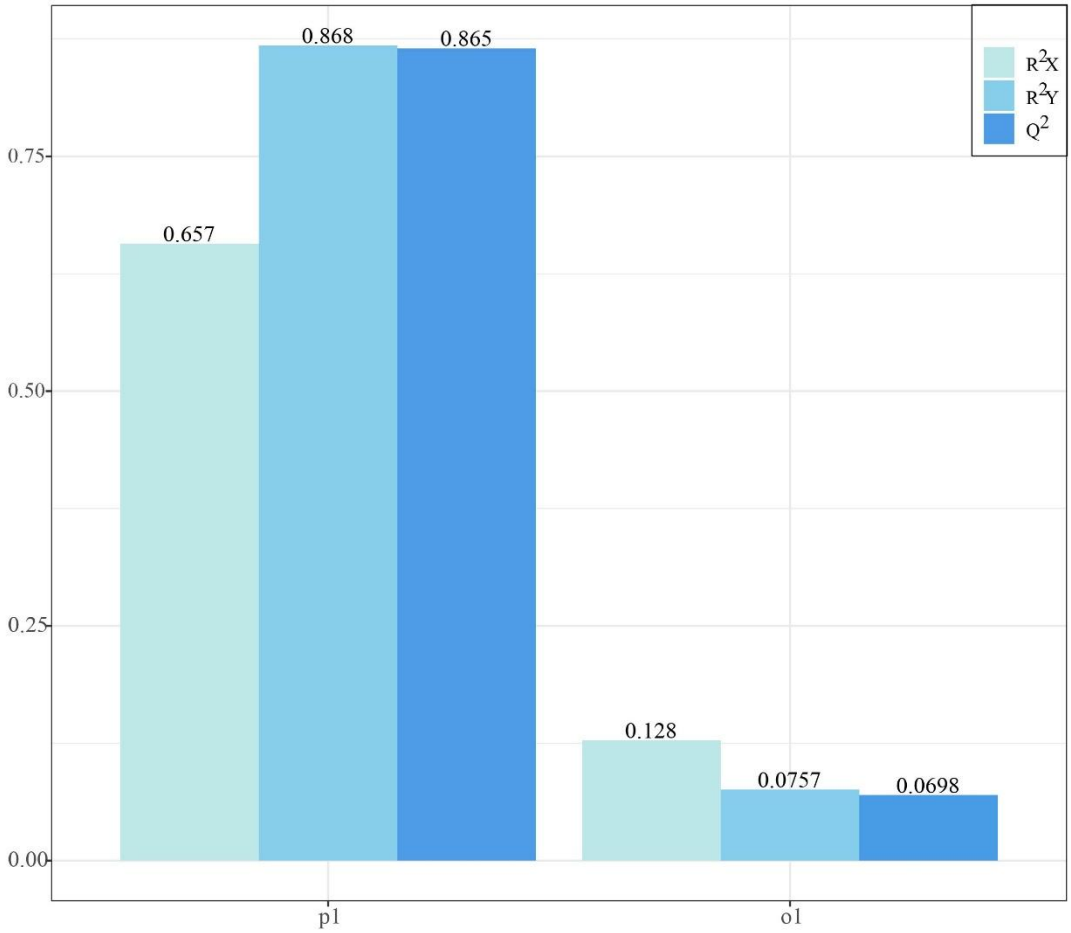
Figure 7. LDA score plot of electronic-nose sensor responses for the five Ilex taxa.

Figure 8. OPLS-DA score plot for discrimination of Ilex rotunda and other Ilex taxa.



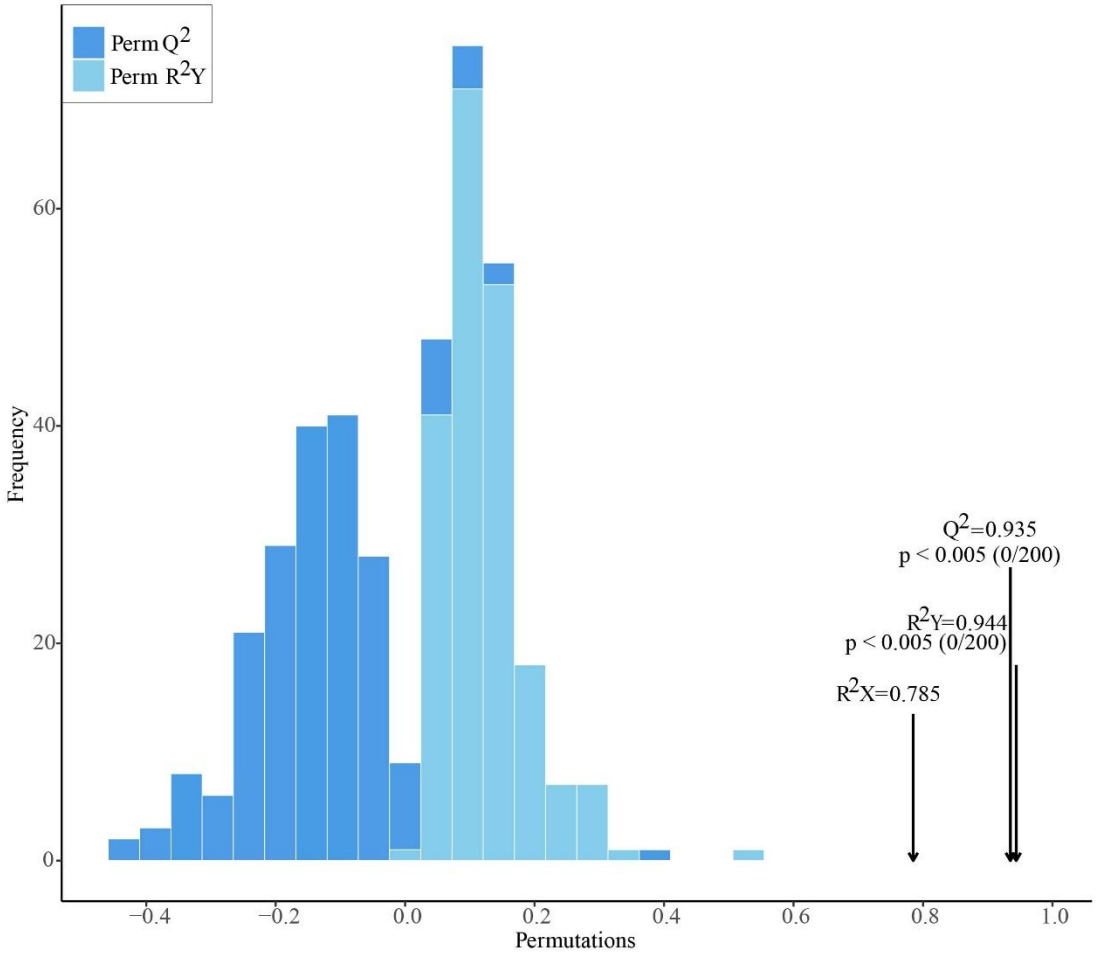
Samples form two distinct clusters primarily along the predictive component p1 ($R^2X = 0.657$), whereas o1 captures orthogonal variation ($R^2X = 0.128$). The separation indicates stable multivariate differences in electronic-nose sensor fingerprints between I. rotunda and non-target taxa.

Figure 9. Summary plot of cross-validation for the OPLS-DA model discriminating Ilex rotunda from non-target taxa.



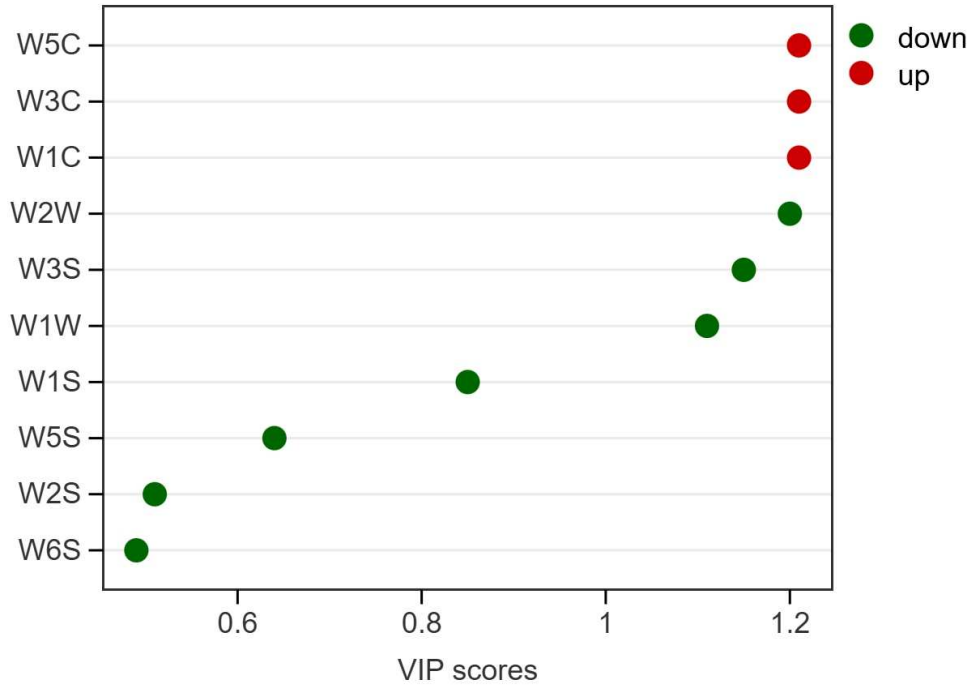
The model comprises one predictive component (p1) and one orthogonal component (o1). For p1, $R^2X = 0.657$, $R^2Y = 0.868$, and $Q^2 = 0.865$; for o1, $R^2X = 0.128$, $R^2Y = 0.0757$, and $Q^2 = 0.0698$. Overall, the results indicate strong explained variance and high cross-validated predictive performance.

Figure 10. Permutation test plot of the OPLS-DA model discriminating Ilex rotunda from other Ilex taxa (n = 200).



The plot shows the distribution of R^2Y and Q^2 for permuted models after random shuffling of class labels, with vertical markers indicating the original model. The original model (R^2Y = 0.944; Q^2 = 0.935) outperformed all permuted models ($p < 0.005$), supporting model robustness (R^2X = 0.785).

Figure 11. Variable importance in projection (VIP) plot of the OPLS-DA model for Ilex rotunda discrimination.



Variables are sorted by VIP scores; higher and lower contributions indicate inter-group response trends based on model coding. Features with VIP > 1 (e.g., W5C, W3C, W1C, W2W, W3S, W1W) are major contributors and can serve as candidate marker features for targeted identification.

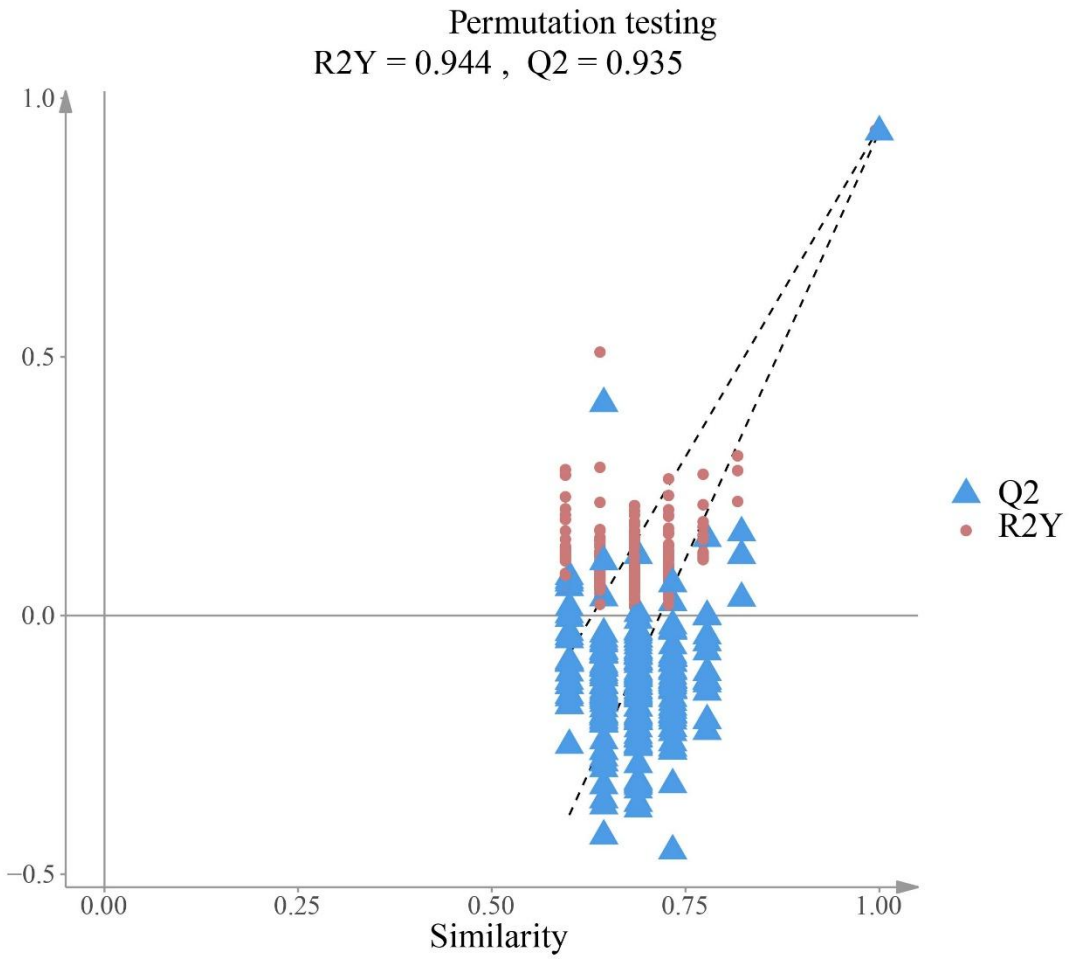


Figure 12. Permutation test results of the OPLS-DA model (n = 200).