

Title	Gas chromatography-based metabolomics for authenticity evaluation of Asian palm civet coffee (Kopi Luwak)
Author(s)	Jumhawan, Udi
Citation	大阪大学, 2015, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/53962
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

Doctoral Dissertation

**Gas chromatography-based metabolomics for
authenticity evaluation of Asian palm civet coffee
(Kopi Luwak)**

Udi Jumhawan

July 2015

Graduate School of Engineering

Osaka University

Table of contents

Table of contents	2
Abstract	5
List of Abbreviations	6
Chapter 1	7
General introduction	
<i>1.1. Metabolomics</i>	7
1.1.1. General concept	7
1.1.2. Application to agricultural and food science	11
<i>1.2. Asian palm civet coffee: a case of perturbation in crop product</i>	16
1.2.1. History and processing	16
1.2.2. Recent research and approach in this study	19
<i>1.3. Research Objective</i>	21
<i>1.4. Thesis outline</i>	22
Chapter 2	23
Selection of discriminant markers for authentication of Asian palm civet coffee via metabolite profiling	
<i>2.1. Introduction</i>	23
<i>2.2. Experimental section</i>	27
2.2.1. Coffee materials, reagents and coffee preparation	27
2.2.2. Coffee beans extraction and derivatization	28
2.2.3. Metabolite profiling and quantification	30
2.2.4. Data preprocessing and metabolite identification	30
2.2.5. Multivariate data analysis	32
<i>2.3. Results and discussion</i>	33
2.3.1. GC/MS-based metabolite profiling of civet coffee	33

2.3.2. Discriminant analysis to select candidates for discriminant markers	35
2.3.3. Validation of the applicability of discriminant markers for authenticity assessment	38
2.4. <i>Conclusions</i>	41
Chapter 3	42
Application of gas chromatography-flame ionization detector (GC-FID)-based metabolite fingerprinting for authentication of Asian palm civet coffee	
3.1. <i>Introduction</i>	42
3.2. <i>Experimental section</i>	44
3.2.1. Coffee bean materials and chemicals	44
3.2.2. Coffee beans extraction and derivatization	44
3.2.3. Metabolite measurement with GC/MS and GC/FID	45
3.2.4. Data preprocessing and dataset construction	46
3.2.5. Multivariate data analysis	46
3.3. <i>Results and discussion</i>	47
3.3.1. Chromatographic data of GC/FID	47
3.3.2. Multivariate data analysis of coffee bean extracts by GC/FID-based metabolite fingerprinting	49
3.3.3. Validation of metabolite fingerprinting strategy to confirm authenticity of commercial coffee products	56
3.4. <i>Conclusions</i>	60
Chapter 4	61
Quantification of Asian palm civet coffee fraction in a coffee blend via metabolomics: a proof of concept	
4.1. <i>Introduction</i>	61
4.2. <i>Experimental section</i>	63
4.2.1. Coffee bean materials	63
4.2.2. Sample preparation, chromatographic and quantitative analyses with GC/MS	64
4.2.3. Data handling and statistical analysis	64
4.3. <i>Results and discussion</i>	66
4.3.1. Chromatographic profiles of coffee bean extract	66

4.3.1. Metabolite profiles of coffee bean extract	68
4.3.2. Mixing ratio-prediction model with OPLA	73
4.3.3. Quantification of important metabolites	78
<i>4.4. Conclusions</i>	<i>81</i>
Chapter 5 Conclusions and perspectives	82
References	85
Appendices	100
List of Publications	116
Acknowledgements	118

Abstract

Considering that fraudulent adulterations are now commonly practiced in various consumer sectors, development of quality standards through labeling and composition regulations and routine evaluation protocol is needed. Asian palm civet coffee is a specialty coffee produced by its passage through the digestive part of Asian palm civet (*Paradoxurus hermaphroditus*). The rarity, exotic processing and unique flavor have contributed to its premium price. However, there is no reliable and standardized protocol to ensure the authenticity of civet coffee. This thesis emphasizes development of a protocol for authenticity evaluation of civet coffee, a world-renowned priciest coffee that has notoriously subjected to fraudulent adulteration and its routine application in industry.

In Chapter 1, general introduction regarding the utility of metabolomics in food science, civet coffee and research background are presented. In Chapter 2, development of standardized protocol through GC/MS-based multimarker profiling of 21 coffee beans (*Coffea arabica* and *Coffea canephora*) from different cultivation areas was demonstrated to explore significant changes in the metabolite profiles as discriminant markers for authentication of civet coffee. Employing multivariate analyses, a set of significant metabolites, mainly organic acids, was selected for further verification by evaluating their differentiating abilities against various commercial coffee products. In Chapter 3, first practical application was presented by developing rapid, reliable and cost-effective analysis via GC coupled with universal detector, flame ionization detector (FID), and metabolite fingerprinting for discrimination 37 commercial and non-commercial civet coffee extracts. GC/FID provided higher sensitivity over a similar range of detected compounds than GC/MS and could successfully reproduce quality prediction from GC/MS for differentiation of commercial civet coffee, regular coffee and coffee blend with 50 wt % civet coffee content without prior metabolite details. In order to prevent illegal mixture of cheap coffee into civet coffee, the proof-of-concept of the utility of metabolomics approach and orthogonal projection to latent structures (OPLS) prediction technique to quantify the degree of coffee adulteration was demonstrated in Chapter 4. The prediction model exhibited accurate estimation of mixing ratio of known-unknown coffee samples. At last, general conclusion and future perspective are elaborated.

List of Abbreviations

(in alphabetical order)

ANDI	Analytical Data Interchange Protocol
CE/MS	Capillary Electrophoresis/Mass Spectrometry
ESI	Electrospray Ionization
FFQ	Food Frequency Questionnaire
FT-ICR	Fourier Transform-Ion Cyclotron Resonance
FT-NIR	Fourier Transform-Near Infra Red
GC/FID	Gas Chromatography/Flame Ionization Detector
GC-Q/MS	Gas Chromatography-Quadrupole/Mass Spectrometry
ICP-AES	Inductively Coupled Plasma-Atomic Emission Spectroscopy
LOD	Limit of Detection
LOQ	Limit of Quantification
LC/MS	Liquid Chromatography/Mass Spectrometry
LDA	Linear Discriminant Analysis
MSTFA	<i>N</i> -methyl- <i>N</i> -(trimethylsilyl)trifluoroacetamide
ND	Not Detected
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
OPLS	Orthogonal Projection to Latent Structures
OPLS-DA	Orthogonal Projection to Latent Structures-Discriminant Analysis
PC	Principal Component
PCA	Principal Component Analysis
PLS	Projection to Latent Structures
QTOF	Quadrupole Time-of-Flight
RMSEE	Root Mean Squared Error of Estimation
RMSEP	Root Mean Squared Error of Prediction
RT	Retention Time
RI	Retention Index
RSD	Relative Standard Deviation
S/N	Signal to Noise
SPME	Solid-Phase Microextraction
UV (1)	Ultraviolet, <i>detector</i>
UV (2)	Unit Variance, <i>scaling method</i>
VIP	Variable Importance in Projection

Chapter 1

General introduction

In recent years, due to growing practice of frauds, interest among consumers over authenticity and quality of agricultural and food products have become one of the major issues in the food industry. Consequently, countless efforts to develop robust authentication systems have been deliberately accomplished by manufacturers and regulatory authorities to satisfy consumers' quality and safety agricultural and food requirements. However, from the perspective of analytical chemistry, those products can be fundamentally considered as complex mixtures and interactions involving various chemical compounds, so that evaluation of the product quality by specific parameter, *i.e.*, total acid, is challenging. Metabolomics, a new emerging 'omics' approach, provides comprehensive and high throughput analysis of set of metabolites, an advantageous feature for quality assessment of agricultural and food products. In this thesis, development of quality assessment of specialty coffee and its routine applications were presented employing metabolomics technology.

1.1. Metabolomics

1.1.1. General concept

Due to its role to mediate the flow of genetic information with the phenotypic feature, metabolomics – exhaustive profiling of set of metabolites within an organism – has gained a lot of attention in various fields from biomarker discovery of diseases to food and nutrition science. The capacity of metabolomics approach to enable snapshot of metabolic process can be linked to the gene function, providing corresponding analysis for those biological samples without genome information¹. Furthermore, transcripts and proteins are subjected to post-transcriptional and translational modifications thus far removed from the phenotype; different at these levels may not represent the result of metabolism. In contrary, metabolites are not

merely the final result of gene expressions but also actively involved and/or produced in the regulatory system of metabolism suggesting there is direct correlation between metabolome and the phenotype². Hence, metabolome is the best target to study metabolic and phenotypic alterations imposed by perturbation, which often found in the food fraud practices.

Metabolomics study is intended to be interconnected multidisciplinary fields that include biological context, analytical chemistry, chemometrics and bioinformatics (**Fig. 1-1**). The workflow enables straightforward manner to examine the working hypothesis based on the existing result³. Ontology and data exchange group worked to provide standardization for the list of descriptive and experimental terms and standard format for data exchange between user and data storage in metabolomics research.

Separation and detection of the metabolites have been considered the key step in the metabolomics technology. Therefore, recent progress of metabolomics has been greatly driven by historical use and continuous improvement of separation (chromatography) and detection science. Common detector systems are based on UV (Ultraviolet), NIR (Near Infra-Red), NMR (Nuclear Magnetic Resonance), or MS (Mass Spectrometry) techniques⁴⁻⁹. The former three techniques have provided fast, relatively non-laborious and non-destructive fingerprint analyses and with the capacity of NMR to allow identification of novel compound through structure elucidation. MS based detection system offers the best combination of chemical selectivity and sensitivity thus it has been used the most for food metabolomics (4). Coupling powerful MS detector with separation techniques such as liquid chromatography in its high performance (HPLC) or ultra performance (UPLC) forms, gas chromatography (GC), and capillary electrophoresis (CE) facilitates amendment of coverage of target compounds¹⁰. Furthermore, availability of plentiful mass spectral databases accelerates spectral matching to pinpoint metabolite identity. However, exhaustive profiling of metabolites within an organism is still challenging which require as many combinations of high-throughput techniques as possible and extremely depending on the availability of authentic chemical standard. Nevertheless, those challenges have opened tremendous prospect for future development of metabolomics technology.

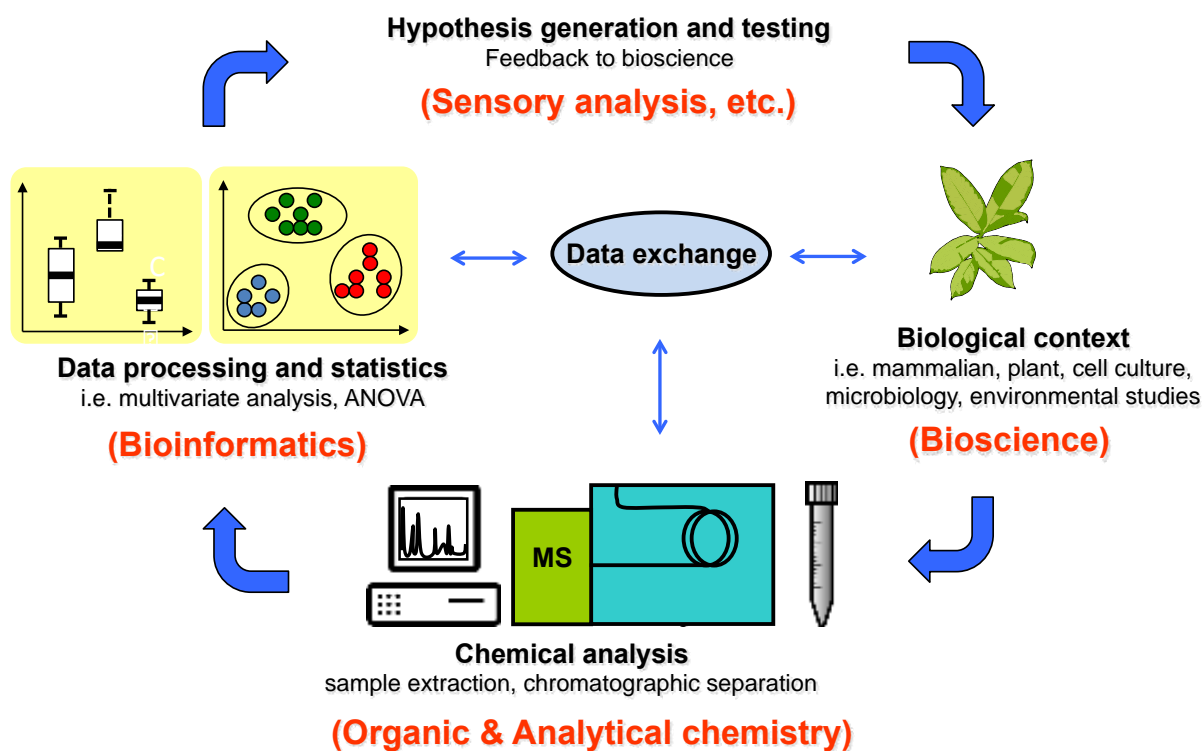


Fig. 1-1. Metabolomics workflow involves multidisciplinary fields (adapted and modified from the metabolomics standards initiative (MSI); Fiehn et al.³).

Metabolomics technology can be conventionally classified into four major approaches; (i) targeted analysis, (ii) metabolite profiling, (iii) non-targeted analysis, and (iv) metabolite fingerprinting (**Fig. 1-2**)^{3,4,11,12}. Targeted analysis involves precise identification and quantification of specific set of metabolites with most cases requiring high level of extraction and purification of metabolites. Metabolites profiling, also widely known as semi targeted approach, focus on broader analysis coverage for both identified and unidentified metabolites related to specific pathway or cellular phenomenon. Non-targeted analysis and metabolite fingerprinting can be essentially referred as detection of the utmost number of metabolites as biochemical fingerprint for rapid classification without necessarily identifying the metabolites. Yet, identification and quantification are often needed for non-targeted analysis¹³⁻¹⁵. To examine global metabolic perturbation caused by illicit action to food and agricultural product, metabolite profiling is the most suitable approach that it can provide depth-of-coverage analysis without a priori target compounds and reliance to availability of authentic chemical standard thus marking possibility to capture novel metabolites as

biomarker. Meanwhile, metabolite fingerprinting offers profile of total metabolite, a fingerprint, without necessarily characterizing metabolite identity and their biochemical function which beneficial for routine application in industry.

Based on specific goal and data processing, Cevallos-Cevallos et al.⁴ divided metabolomics strategies into; informative, discriminative and predictive (**Fig. 1-2**). Informative analysis commonly covers principal work of metabolomics involving identification and quantification of metabolites to provide fundamental knowledge of the biological sample. Informative analysis has been utilized for various applications such as development of metabolite databases, functionality studies through metabolic flux analysis, metabolomics-assisted strain improvement and many others¹⁵⁻¹⁷. Metabolomics analysis has tended to be discriminative when applying statistical models for differentiation of two or more samples populations. The discriminative approach is commonly associated with biomarker identification^{18,19}. The use of statistical models is also necessary for predictive metabolomics. Prediction of variables connecting a huge metabolite data and phenotypic trait that is problematic by other approaches can be done through statistical prediction model. When dealing with illegal attempts toward food and agricultural product, metabolomics reports have been generally discriminative and predictive.

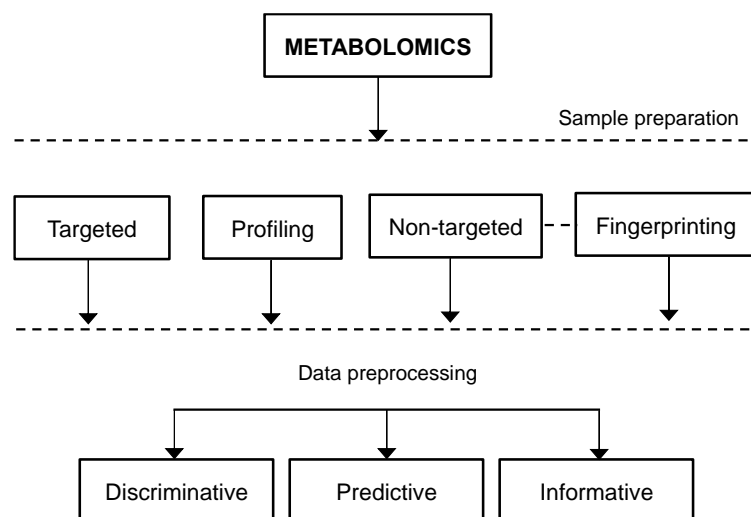


Fig. 1-2. Classifications of metabolomics strategies (adapted and modified from Cevallos-Cevallos et al.⁴).

1.1.2. Application to agricultural and food science

Metabolomics provides novel approach to understand a thorough and comprehensive molecular snapshot of what gives certain food and agricultural products their phenotypic features such as taste, aroma, or color²⁰. In food and agricultural products originates from plant, those features are likely defined from interactions of more than 10000 detectable metabolites, with more than 2000 of them are nutrient related metabolites, with environmental factors²¹. Due to its capacity for simultaneous profiling of large numbers of metabolites in various biological matrices, metabolomics technology has recently risen as a potential tool for manufacturing, safety and quality evaluation of food and agricultural products.

Metabolomics in food component analysis and processing

From the perspective of metabolomics, most food and agricultural products are chemical pools consisting of thousands of metabolites from very broad groups such as protein, sugars, amino acids, fats, fiber, vitamins, and trace elements, and its interaction in a form of solid, semi-solid and liquid matrix²⁰. The number and variety of chemical constituents in food and agricultural products are dependable on the level of processing and the natural diversity itself. Some processed products only consist of up to 20 different chemical constituents (soft drinks, artificial juices, vegetable oils) while some others may contain hundreds and thousands of compounds (raw fruits, meats, and most prepared food)²⁰. Changes in the food components through food processing including physical and chemical modification can be also captured by metabolomics. The production of soy sauce, traditional fermented soybean beverage, industrial pasta and beer are among the study cases of metabolomics application in food manufacturing²²⁻²⁴.

The advent of metabolomics technology has been contributed to the impressive collection of 15000+ food related metabolites²⁵. GC/MS and LC/MS remained to be the most popular platform used for most agricultural products, *i.e.*, grapes, tomato, potato, rice, melon, green tea, coffee and herbal plants, due to its selectivity and sensitivity^{6,18,19,26-35}. The use of NMR and FT-NIR non-destructive methods has also been progressively adapted for component analyses of green tea, milk, tomato juices, and beer³⁶⁻⁴⁰. For food science, metabolomics-based food component analysis will provide detailed knowledge in association with the effect of different food preparation (frying and baking; steaming and boiling) and

storage. Moreover, these studies help to improve the quality through selective breeding of agricultural products as well as livestock.

Metabolomics in diets and food consumption monitoring

Considering that diseases brought on by appalling dietary lifestyle have become a rising concern among consumers, monitoring food consumption is one of the most important issues worldwide. While monitoring food consumption through a dietary program impacted physical health⁴¹, most of the nutritional scientists have relied on the use of questionnaire based epidemiological studies, termed as food frequency questionnaire (FFQ) during pre-metabolomics era, which unfortunately do not allow for measurement of biological fluids to corroborate direct association between certain foods or diet with its consequences²⁵.

The onset of metabolomics have been concomitantly supported, or even replaced the practice of epidemiological study of large populations since biological circumstances, represented by circulating nutrients in plasma or blood, may vary from person to person⁴². The capacity of metabolomics to capture spectral alterations occurring from diet or food-induced changes presenting a better interpretation of the physiological effects of a certain food or diet and identification of biomarkers that potentially beneficial for health. Those changes are present in the blood or urine for a couple hours and some of them last as long as 48 hours⁴³, marking the significance of both biological liquids for nutritional studies. Majority of biomarkers that can be used as food consumption are typically polar metabolites (chlorogenic acid for coffee, epicatechin for tea) while non-polar compounds dominating the biomarkers of physiological reaction to food or diet, *i.e.*, triacylglycerol and total cholesterol increased risk of cardio vascular disease²⁰.

Metabolomics in quality and authenticity evaluation of food and agricultural products

Despite of annual fluctuation in the global production due to environmental and socio-economy, agricultural crops and livestock remain to be most valuable commodities and starting material to yield various products from food to fuel. In recent years, concerns regarding safety, quality and authenticity of food and some valuable crops have become the major challenge in food science. Adulterations in many consumer sectors are commonly practiced, both intentionally and accidentally. Whereas intentional adulteration involve illicit

act to enhance the product values, accidental or non-intentional adulteration often associated with contamination and lack of proper hygiene during the overall processing, resulting in a downgrading of product values. Detection of fraudulent adulteration is essential for industry and legislative authority. Both parties are accountable to establish quality standards through labeling and composition regulations and routine evaluation protocol to circumvent unfair competition among manufacturers as well as to ensure safety, quality and authenticity of the product for consumers^{44,45}.

According to Dennis⁴⁴, authenticity evaluation of food and agricultural products can be generalized into these following issues.

- Species of origin

Authenticity issue arise when one species possess superior qualities due to differences in genetic makeup, physical and morphological properties, compared to others. Differentiation of panax (*Panax ginseng* and *Panax quinquefolium*) and coffee (*Coffea arabica* and *Coffea canephora*) are among the well-known incidents in which species of origin plays a significant benchmark for authentication. The discrepancy in the sale price of panax species is due to its medical value, while Arabica commanding selling prices up to three times higher than Robusta, market name of *Coffea canephora*, because the former has finer taste and aroma. The consequence of the price gap in both incidents are growing illegal attempt to mix one species with similar material from a cheaper species.

- Geographical region of origin

Several foodstuffs and beverages such as cheeses, sausages, olive oil, wine and beer are very common to be associated with geographical origin, particularly in European countries. Protocols to discriminate Japanese and Chinese herbal medicine have also been reported in attempt to detect adulteration¹⁹. The need to establish protection to those products on the account of geographical origin is mainly motivated because of environmental and heritage factors.

- Commercial treatment

Several commercial treatments are considered to be the source of adulteration including cold pressed olive oil, pasteurization and food irradiation. Unlawful mixing or addition of water into olive oil, milk, wine and beer remains a common extender of those products (although it is indirect with commercial treatment). This particular issue led to attempt for protecting originality of brands and its quality, *i.e.*, Scotch whisky.

Particularly for agricultural crops such as coffee and tea, quality evaluation has been conventionally assessed on the basis of human sensory perception. Consumers' preferences are represented with panel of assessors, on whom the products are evaluated utilizing human senses (sight, smell, taste and touch). However, this method tended to be highly subjective with only up to 20% precision⁴⁶. Moreover, because it takes years of training to obtain specialty in sensory analysis, hiring highly trained specialists to assess product quality is unavoidably expensive. Therefore, it would be advantageous to develop an alternative method from nonhuman measurement.

Metabolite, target in metabolomics technology, can be directly connected with phenotype that is sensitively affected by any type of perturbation or stress. The capacity of metabolomics to measure metabolites as well as to identify food components favor its usefulness for detection of adulterated crops or food products. Generally, technical approaches for detection of adulteration assisted by metabolomics can be largely divided into three categories⁴⁵. The first approach focuses on determination of the ratio of some chemical constituents, specifically the ones that are likely influenced by perturbation, with assumption these ratios are constant in authentic product. The use of multivariate analysis is often required to deal with differentiation of large sample set. Seeking specific marker in the product that emerges in association with fraud can be posed as second approach. Comparative analysis utilizing statistical analyses can be easily classified adulterated and authentic samples. The last possibility for detection of counterfeit in food products involves global examination of the products applying combination of physical and chemical (metabolomics) analysis for observing the outcomes of adulteration to the food products thoroughly. The availability of numerous analytical platforms, data processing algorithms, and multivariate analysis highly embolden the application of metabolomics technology to counteract adulteration.

Coffee is one of the most valuable internationally traded crop products for an estimated 530 million bags in December 2014⁴⁷ and hence is very susceptible to any kind of adulterations. Motivated by its economical values, different types of coffee adulteration have been frequently found in the market. The most common encountered adulteration cases as summarized by Briandet et al⁴⁸ are as follow:

- Adulteration with coffee substitutes such as chicory, malt, figs, cereal, caramel, starch, maltodextrins, glucose and coffee husks.
- Unlawful replacing or mixing of two species. Among many coffee species, only two of them are commercially profitable, Arabica (*C. arabica*) and Robusta (*C. canephora*). Robusta has been branded as a weak-flavored coffee with low acidity and a strong bitterness, whereas Arabica provide a milder, intense aroma and finer taste.
- Mixing of expensive coffee from a particular cultivation area with cheap beans grown in another region or mixing of premium coffee with superior features with common beans.

In relation to application of metabolomics for coffee discrimination and detection of adulteration, there are several studies were reported utilizing wide range of analytical systems (Table 1-1).

Table 1-1. Metabolomics technology for coffee science

Objective	Technique(s)	Discriminant marker	Ref
Discrimination between Arabica and Robusta green coffee	Raman spectroscopy	Chlorogenic acid and lipid content	46
Discrimination of mature – immature Arabica beans	LC/MS	Tryptophan (immature beans)	49
Differentiation of coffee owing to geographic growing origins	ICP-AES	Not reported	50

Table 1-1. (Continued)

Objective	Technique(s)	Discriminant marker	Ref
Discrimination of Arabica and Robusta in instant coffee	FT-IR	Chlorogenic acid and caffeine (Robusta)	51
Discrimination of defective – non-defective among Brazilian Arabica (1) and between green Arabica – Robusta beans	SPME-GC/MS, ESI-MS	Benzene acetic acid, 1H-pyrrole, 4-Methylthiazole, methyl ester, etc. (defective beans), High sucrose (non-defective), low sucrose (defective), higher content of phenolic compounds, such as chlorogenic acid for Robusta	52,53
Detection of coffee blend between Arabica and Robusta	Raman spectroscopy, FT-ICR/MS, QTOF	Kahweol, three diterpene glycosides (Arabica), caffeoyltryptophan and caffeoyltyrosine (Robusta)	54,55
Classification of green coffee beans according to variety and origin	NMR	Sucrose, caffeine, chlorogenic acids, choline, amino acids and trigonelline	56

1.2. Asian palm civet coffee: a case of perturbation in crop product

1.2.1. History and processing

Asian palm civet coffee (in short, civet coffee) or Kopi Luwak, Indonesian words for coffee and civet, respectively, is a premium coffee prepared from coffee berries that have been eaten by Asian palm civet (*Paradoxurus hermaphroditus*), a small mammal native to southern and northern Asia, widely spread from India, Bangladesh, Sri Lanka to central Indonesia (**Fig. 1-3**). It is listed as one of the world's priciest coffee with a price tag of 3000 Japanese yen per 100 gr (or US\$ 150 to 220 per pound), a hundred times higher than common beans. Approximately, only 300 kg to 1 ton of civet coffee are traded and exported annually to specialty coffee consumer in South Korea, Japan, USA, Australia and European countries

(personal communication with Indonesian Coffee and Cocoa Research Institute). The transient fermentation inside the civet's gut hypothetically adds a distinct flavor to the coffee beans. However, its rarity as well as the coffee's exotic and unique production process ultimately accounts for its high selling price. Civet also inhabit African continent with the most well known species, *Civettictus civetta*, are historically the main species from which was obtained a musky scent used in perfumery⁵⁸.



Fig. 1-3. Asian palm civet or *Paradoxurus hermaphroditus* (right) and its native habitat (left native in green, introduced in red color)

Civets are primarily frugivorous, a fruit eater, feeding on berries and pulpy fruits. They also eat small vertebrates and insects⁵⁷. This 3 – 10 pound animal normally inhabit primary tropical forest, showing nocturnal activity patterns with peaks between late evening until after midnight and often emit a nauseating secretion from their anal scent gland as a chemical defense⁵⁸. Based on IUCN (International Union for Conservation in Nature), Asian palm civet was classified as ‘Least Concern’ animal, as it is tolerant of a broad range of habitats, and is widely distributed with large populations that are unlikely to be declining (IUCN Red List of Threatened Species, 2008). It was reported that civet coffee has its origins in the Dutch coffee plantation estates during Dutch colonization in Indonesia (before 1945). The knowledge of existence of Luwak coffee was first reported in the March 1981 issue of National Geographic, which mentioned civet coffee by name in one of the feature articles entitled “The Bonanza Bean – Coffee”⁵⁷.

The civet is an expert tree climber. They climb coffee trees at night and instinctively select the ripest coffee cherries. The coffee cherry is sweet and coffee pericarp is completely digested whereas the beans are excreted along with their feces. Transient fermentation, within a day, inside civets' digestive tract, may alter the morphology and intrinsic features of coffee beans, proven with smoother surface, reddish color, and exhibit unique flavor described as earthy, musty, syrupy, smooth, low acidity, low bitterness and rich with chocolate undertone⁵⁷, after beans are digested by civet.

There are two common methods of civet coffee production, traditional and conventional method. Traditional method allows civet roam freely within defined boundaries of coffee plantation or forest, unconsciously picking coffee cherries from the tree, compelling the farmer to collect defecated beans from the nature. In contrary, conventional method includes intensive farming where civet are put inside cage and fed with coffee beans on a daily basis. The latter, that is adopted by many farmers nowadays, has raised ethical concerns about animal cruelty from international organizations. Eventually, Indonesian coffee and cocoa research institute, one of pioneer of civet coffee production in Indonesia, initiated a '*less cruel, more friendly*' conventional method where the battery cage systems are only permitted during coffee harvest season. Afterward, the civets are released to its nature (or within defined boundaries), switching into traditional method and the cycles are recurring by capturing the wild civet prior to harvest period.

After processed by digestive system of civet, the beans are excreted along with the feces, having kept their shape and still covered with some of fleshy pulp fruit. Bean clumps are then collected, cleaned, sorted into individual bean, wet-fermented, sun-dried, and further treated by roasted (**Fig. 1-4**). Civet coffee is considerably safe for consumption after extensive washing under running water, drying and roast at high temperature. The level of harmful microorganisms (enteric and coliform microorganisms) was significantly reduced comparing to untreated coffee beans. Upon roasting, reduction of colony occurred to near undetectable levels. Extensive wash and roasting would decrease microbial colony and lower their overall counts⁵⁷.

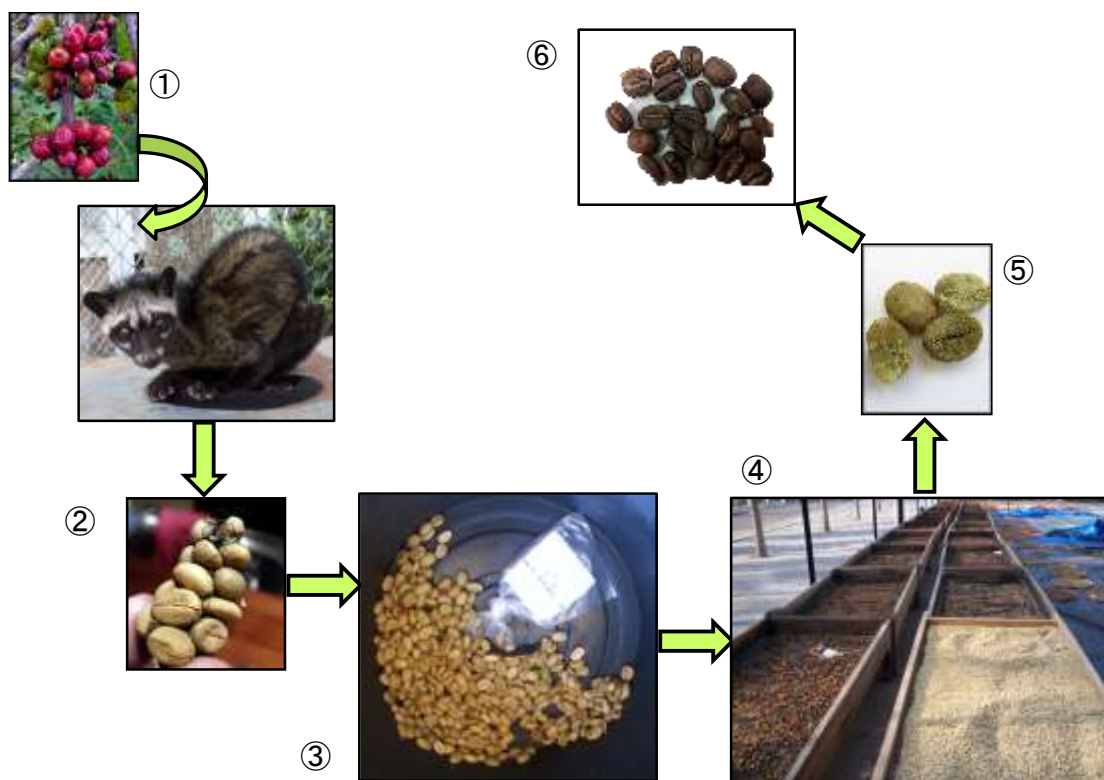


Fig. 1-4. The civet coffee production starts with digestion of coffee cherries by civet ①, internal fermentation removes the pericarps and resulting in a bean clumps ②, followed with cleaning with running water – sorting – wet fermentation ③, sun-drying ④ to obtain proper green beans ⑤ and treated with roasting ⑥.

1.2.2. Recent research and approach in this study

Despite of its uniqueness to be one of few human food produced by its transitory route to digestive tract of animal, there is lack of scientific information on this exotic coffee. The reported studies, including one research article⁵⁷ and one proceeding⁵⁹, mainly focuses on discrimination between civet coffee from specific production areas and ‘regular coffee’ (not eaten by civet). The first report employed physiochemical characterization involving color, surface area, total protein, aroma, and cupping test to reveal dissimilarities between those coffees. Owing to physical and enzymatic treatment during passage to civet’s gut, civet coffee expectedly exhibited smoother appearance and reddish – darker color comparing control bean. Substantial degradation of storage protein was also observable confirming the penetration of proteolytic enzymes into coffee beans. Furthermore, cupping test and electronic nose showed

that civet and regular coffee were rather distinct each other based on their aroma, body and acidity. The latter report utilized electronic nose and GC/MS to distinguish civet from the control beans and to disclose that aroma and volatiles profiles of civet coffee were specific to varietal and production area. However, both studies emphasized a mere discrimination analysis without providing information of candidates for authentication. It is therefore challenging to standardize the promising approaches, *i.e.*, electronic nose and SDS-PAGE for future application.

In this study, metabolomics technology is employed as an effective approach to pinpoint changes in the metabolite profile of coffee beans triggered by animal digestion, due to its direct association with phenotypic features. Semi targeted analysis, or metabolite profiling, and metabolite fingerprinting assisted by GC/MS and GC/FID are used to provide a wider coverage of metabolites to screen candidates of discriminant marker as well as to establish rapid protocol of civet coffee quality evaluation in industry. To thoroughly screen and isolate significant metabolites and to select proper candidates of discriminant marker, three types of metabolomics strategies, informative, discriminative and predictive are applied, represented by the use of multivariate analyses, such as PCA (Principal Component Analysis), OPLS (Orthogonal Projection to Latent Structures) and its combination with discriminant analysis (OPLS-DA) (**Fig. 1-5**).

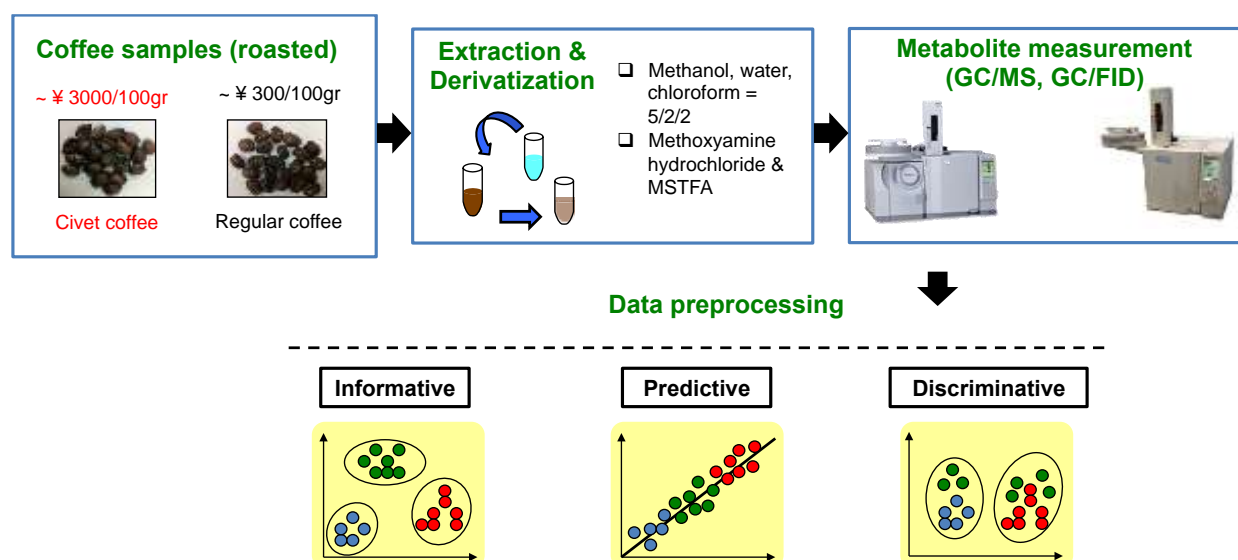


Fig. 1-5. Experimental workflow and strategies in this study

1.3. Research Objective

In the recent years, owing to extensive fraud attempts that have been commonly found in various consumer sectors, authentication of food and crop products is currently indispensable to ensure consumers' safety and satisfaction. Aside from being one of the food product produced by animal perturbation, civet coffee also has been susceptible to fraudulent adulteration due to its uniqueness and economic value. As described in section 1.2.2, the scientific information of this exotic coffee is remained scarce. Consequently, there is no reliable, standardized protocol to evaluate authenticity of civet coffee.

Metabolites are the ultimate readouts of metabolism, representing a closest link to phenotypic features, and thus are the most proper target to seek changes in their profiles resulted from animal perturbation as candidates of discriminant marker for quality evaluation of civet coffee. The objective of this study is to develop reliable protocol for authenticity and quality assessment of civet coffee as well as to expand its application for rapid, routine evaluation in industry. To accomplish this overall objective, the following strategies are needed:

- 1) Seek and identify discriminant marker candidates to differentiate civet coffee and regular coffee on the regard of animal perturbation through metabolite profiling
- 2) Verify the applicability of discriminant marker candidates for evaluation of commercial coffee products
- 3) Expand the usefulness of metabolomics for rapid analysis utilizing knowledge of discriminant markers via metabolite fingerprinting coupled with cost-effective analytical platform and construction of prediction model to determine degree of coffee adulteration for industrial applications.

1.4. Thesis outline

This thesis consist five chapters emphasizing development of a protocol for authenticity evaluation of civet coffee, a world-renowned priciest coffee that has notoriously subjected to fraudulent adulteration and its routine application in industry. In Chapter 1, general introduction regarding the utility of metabolomics in food science, civet coffee and research background are presented. In Chapter 2, GC/MS-based multimarker profiling of 21 coffee beans (*Coffea arabica* and *Coffea canephora*) from different cultivation areas was demonstrated to explore significant changes in the metabolite profiles as discriminant markers for authentication of civet coffee. Employing multivariate analyses, a set of significant metabolites was selected for further verification by evaluating their differentiating abilities against various commercial coffee products. In Chapter 3, first practical application was presented by developing rapid, reliable and cost-effective analysis via GC coupled with universal detector, flame ionization detector (FID), and metabolite fingerprinting for discrimination 37 commercial and non-commercial civet coffee extracts. GC/FID provided higher sensitivity over a similar range of detected compounds than GC/MS and could successfully reproduce quality prediction from GC/MS for differentiation of commercial civet coffee, regular coffee and coffee blend with 50 wt % civet coffee content without prior metabolite details. Considering that illegal mixture of cheap coffee into civet coffee is a growing concerns among consumers, the proof-of-concept of the utility of metabolomics approach and orthogonal projection to latent structures (OPLS) prediction technique to quantify the degree of coffee adulteration was demonstrated in Chapter 4. The prediction model exhibited accurate estimation of mixing ratio of known-unknown coffee samples. At last, general conclusion and future perspective are elaborated.

Chapter 2

Selection of discriminant markers for authentication of Asian palm civet coffee via metabolite profiling

2.1. Introduction

As described in previous chapter, fraudulent adulterations are currently found in every consumer sectors including crops and livestock and forcing government as authorized regulator and manufacturers to provide verification vial products' labeling. Different labeling and compositional regulations from one food to another, from country to country, has opened a challenge to seek appropriate method, technology and/or biomarker for food quality and safety evaluation⁴⁴. With up to 530 million bags are exported across the globe during period of December 2014⁴⁷, coffee is undoubtedly one of the most valuable and traded agriculture products. The prominence of coffee market and its globalization have resulted in drawbacks including countless attempts for fraud. Adulteration with coffee substitutes and illegal mixing/replacing prestigious coffee species with the cheaper kind are commonly practiced in the market.

Among all known coffee species, the most traded are *C. arabica* and *C. canephora* (Robusta) and among many varieties that are grown across the globe, a few, such as *Jamaican Blue Mountain* and *Hawaiian Kona*, have been recognized for its remarkable flavor thus, they are branded as premium or specialty coffee⁵⁷. Of these premium coffees, civet coffee is considered to be the most expensive. The actions of civet's gastro-intestinal tract and digestive enzymes modify the chemical composition of these coffees, thus yielding a unique flavor. It is however, the exotic process and short supply are the principal reasons for its high price, which is approximately 2 – 3 times higher than other premium coffees⁴⁷.

Despite its profitable prospects, there is no reliable, standardized method for determining the authenticity of civet coffee. Moreover, there is limited scientific information on this exotic coffee. Recently, type 3 adulteration (commercial treatment), involving regular coffee that was treated enzymatically and physically (cold storage) and marketed with the tag of “Kopi Luwak”, was reported in the coffee market⁶⁰. This poses serious concern among consumers over the authenticity and quality of the products currently available in the market. Discrimination between civet coffee and regular coffee has been achieved based on its aroma content via electronic nose data and total protein content using SDS-PAGE⁵⁷. However, the selection of a discriminant marker for authentication was not addressed. Moreover, 1D SDS-PAGE had a low dynamic range where 1 single band represents hundred of protein and commercial library of electronic nose data currently is not available, suggesting a challenge for standardization. The methods currently employed by civet coffee producers is sensory analysis include visual and olfactory testing, both of which are inadequate. For example, visual examination is only possible for green coffee beans prior to roasting, very few trained experts can perform the highly subjective sensory analysis to discriminate civet coffee and years of experiences are necessary to attain this expertise.

Information flows in metabolic pathways are highly dynamic and represent the current biological states of individual cells. Hence, the metabolome has been considered as the best descriptor of physiological phenomena⁶¹. Metabolomics techniques can be powerful tools to elucidate variations in phenotypes imposed by perturbations such as gene modification, environmental factors, or physical stress. The “*black box*” process, or perturbation, during animal digestion can be translated as physical and enzymatic consequences to the coffee bean, which presents a smoother surface and color changes after digestion. Thus, metabolomics technique was chosen to select discriminant markers related to animal processing, to counteract those adulteration attempts. Metabolomics techniques have been effectively applied to distinguish the phytochemical compositions of agricultural products of different origins,⁶² varieties,^{56,62} and cultivars⁶³ for quality control and breeding.

In this chapter, gas chromatography coupled with quadrupole mass spectrometry (GC-Q/MS)-based multimarker profiling was employed to explore significant metabolites as candidates of discriminant markers for the differentiation of civet coffee and regular coffees.

A combination of gas chromatography and mass spectrometry (GC/MS) provides high sensitivity, reproducibility, and the quantitation of a large number of metabolites with a single-step extraction.^{64,65} Sample classification by means of chemometrics was performed using principal component analysis (PCA). Subsequently, orthogonal projection to latent structures combined with discriminant analysis (OPLS-DA)⁶⁶ was used to isolate statistically significant compounds as discriminant marker candidates.

OPLS-DA modeling is second derivative modification of partial least squares to latent structure, a regression analysis commonly used to decipher relationship between two data matrices, X (descriptors) and Y (responses). The X matrix, can be, for instance, chromatographic, microarray, or spectroscopic data with N samples and K wavelengths or mass-to-charge ratio and Y matrix are often represented phenotypic properties such as sensory profile, tolerance, toxicity etc. OPLS is first modification of PLS by removing variations in X that is not correlated to Y , mathematically termed as removal of systematic variation in X that is orthogonal to Y (**Fig. 2-1**)^{66,67}. The separation of orthogonal variation from predictive variation, the one that correlates with Y , facilitates easier model interpretation and less complex model.

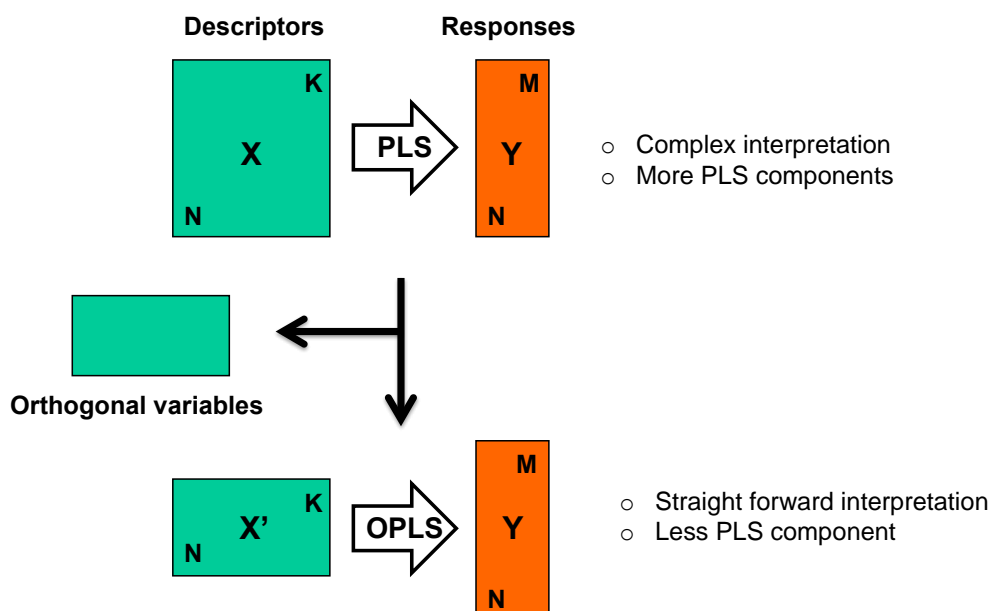


Fig. 2-1. Overview of OPLS with K , M , and N represent data matrix of descriptor, *e.g.* chromatographic profile, predictor, *e.g.* quality ranking and number of samples, respectively.

For OPLS-DA, to provide discriminant analysis, the response matrix or Y is formed a binary vector with the value 0 for one class and 1 for another class. Both OPLS and OPLS-DA model comprises two blocks of variations, predictive ($\mathbf{T}_p\mathbf{P}_p^T$) which characterizes the *between* class variation and orthogonal ($\mathbf{T}_o\mathbf{P}_o^T$), which symbolizes the *within* class variation, into this formula:

$$X = T_p P_p^T + T_o P_o^T + E$$

\mathbf{T} represents the score matrix, \mathbf{P} is the loading matrix and \mathbf{E} is a residual variable that is not correlated to both blocks of variations. To determine significant variables influenced the designed model, a visualization graph termed as S-plot can be easily built from OPLS-DA. This S alphabet-like plot combines the contribution or magnitude (covariance) and the reliability (correlation) of the metabolites as variables to model component scores. Those two vectors are determined as follows,

$$Cov(t, X_i) = \frac{t^T X_i}{N - 1}$$

$$Corr(t, X_i) = \frac{Cov(t, X_i)}{s_t s_{X_i}}$$

t = component score in OPLS-DA

i = centered variable in observation matrix \mathbf{X}

N = number of samples

s = estimated standard deviation

The applicability of the discriminant marker candidates was then further verified by evaluating their differentiating abilities against various commercial coffee products. This study presents the first report to address the selection and successful validation of discriminant markers for the authentication of civet coffee.

2.2. Experimental section

2.2.1. Coffee materials, reagents and coffee preparation

Coffee samples were divided into experimental and validation coffee groups. Experimental coffees were utilized to construct the discrimination model and to select significant compounds. Verification of the applicability of the discriminant markers was carried out using the validation coffee set. In this thesis, coffee that had been digested by the animal is referred to as civet coffee, and the other beans as regular coffee. Civet coffee and regular coffee samples of two species, *C. arabica* (Arabica) and *C. canephora* (Robusta), were used. Coffee samples were obtained from 21 sampling points in three cultivation areas in Indonesia (Java, Sumatra, and Bali). Samples were roasted in a Probat-Werke von Gimborn Maschinenfabrik GmbH model BRZ 2 (Probat, Rhein, Germany) at 205°C for 10 min to obtain a medium degree of roasting, and then were air-cooled for 5 min. These samples were acquired from Indonesian coffee and cocoa research institute in August 2011. Half of intact roasted beans were packed in vacuumed-sealed plastic and stored at 4°C. Another half were ground and stored in 50 mL sealed BD Falcon tube at -30°C with light-shielding prior to analysis (September 2011). Wet-fermentation was applied to both the civet coffee and regular coffees. For regular coffees, conventional protocols were applied after harvesting, including de-pulping, wet fermentation, and drying.

The validation set consisted of authentic civet coffee, commercial civet coffee, commercial regular coffee, fake coffee, and coffee blend. Authentic civet coffee was produced via controlled processing to ensure the quality of the beans pre- and post-digestion. The remaining samples were purchased commercially. The coffee blend was utilized to examine the feasibility of the method for differentiating mixed and pure coffees. The sample descriptions are shown in **Table 2-1 and Supplementary Table 2-1**. In **Table 2-1**, samples no. 1 – 5 and 7 – 11 were originated from same cultivation area whilst samples no.5 and 11 are benchmark for control processing (pre- and post digestion).

Methanol, chloroform, distilled water, ribitol, and pyridine were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Methoxyamine hydrochloride, quinic acid, chlorogenic acid, and alkane standard solution were purchased from Sigma Aldrich

(Milwaukee, Wisconsin, USA). *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) was purchased from GL Science, Inc. (Tokyo, Japan). Caffeic acid and sucrose were acquired from TCI chemical (Tokyo) and Kisida chemical (Osaka), respectively. The six authentic standards of the discriminant markers, their providers, and purities were as follows: citric acid (Nacalai-Tesque, Kyoto, Japan, 99.5%), malic acid (Nacalai-Tesque, 99%), pyroglutamic acid (ICN Biomedicals, Ohio, USA, 99.5%), caffeine (Sigma Aldrich, 98.5%), inositol (Wako, 99%), and glycolic acid (Sigma Aldrich, 99%).

2.2.2. Coffee beans extraction and derivatization

Coffee beans were put into a grinding mill container, cooled for 3 min on water ice cubes, and then ground with a Retsch ball mill (20 Hz, 3 min). Coffee bean powder (15 mg) was transferred into a 2 mL Eppendorf tube. In addition to pure samples, a 50:50 (wt%) blend of civet coffee and regular coffee was used. One milliliter of a single-phase extraction solvent consisting of 2.5/1/1 (v/v/v) methanol, distilled water, and chloroform, respectively, was added to extract a wide range of metabolites. A non-specific extraction procedure was applied to avoid limiting the target analysis to specific compounds and to comprehensively screen the components of civet coffee. As an internal standard, ribitol (60 μ L, diluted with deionized water to 0.2 mg/mL) was utilized. The mixture was shaken for 1 min and then centrifuged at 4°C and 16000 g for 3 min. The supernatant (900 μ L) was transferred into a 1.5 mL Eppendorf tube and diluted with 400 μ L Milli-Q water (Wako). The mixture was then vortexed and centrifuged for 3 min. A 400 μ L portion of the aqueous phase was transferred into a fresh 1.5 mL Eppendorf tube with a screw cap. The solvent was removed by vacuum centrifugation for 2 h, followed by freeze-drying overnight. All samples were analyzed in triplicate ($n = 3$).

To donor volatility for those non-volatile metabolites, derivatization utilizing oximation and trimethylsilylation were conducted. Methoxyamine hydrochloride (100 μ L, 20 mg/mL in pyridine) was added to the dried extract as the first derivatization agent. The mixture was incubated at 30°C for 90 min. After addition of the second derivatization agent, *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA, 50 μ L), the mixture was incubated at 37°C for 30 min.

Table 2-1

List of coffee samples (Experimental coffee set)

No.	Origin	Species	Type	Harvest year	Extraction period	Samples code ^a
1	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	29
2	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	31
3	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	39
4	Bali	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	33
5	Java	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	24
6	Java	<i>Coffea arabica</i>	Civet coffee	2011	Sept 2011	26
7	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	30
8	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	32
9	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	40
10	Bali	<i>Coffea arabica</i>	Regular coffee	2010	Sept 2011	34
11	Java	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	22
12	Java	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	41
13	Java	<i>Coffea arabica</i>	Regular coffee	2010	Sept 2011	42
14	Java	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	45
15	Bali	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	43
16	Bali	<i>Coffea arabica</i>	Regular coffee	2011	Sept 2011	44
17	Sumatra	<i>Coffea canephora</i>	Civet coffee	2010	Sept 2011	35
18	Sumatra	<i>Coffea canephora</i>	Regular coffee	2010	Sept 2011	36
19	Java	<i>Coffea canephora</i>	Regular coffee	2011	Sept 2011	46
20	Java	<i>Coffea canephora</i>	Regular coffee	2011	Sept 2011	47
21	Java	<i>Coffea canephora</i>	Regular coffee	2010	Sept 2011	51

^aFor multivariate analysis

2.2.3. Metabolite profiling and quantification

Metabolite profiling utilizing gas chromatography coupled to quadruple mass spectrometry (GC-Q/MS) was performed on a GCMS-QP 2010 Ultra (Shimadzu, Kyoto, Japan) equipped with a CP-SIL 8 CB low bleed column (0.25 mm x 30 m, 0.25 μ m, Varian Inc., Palo Alto, California, USA) and an AOC-20i/s (Shimadzu) as an autosampler. The mass spectrometer was tuned and calibrated prior to analysis. The derivatized sample (1 μ L) was injected in split mode, 25/1 (v/v), with an injection temperature of 230°C. The carrier gas (He) flow was 1.12 mL/min with a linear velocity of 39 cm/s. The column temperature was held at 80°C for 2 min, increased by 15°C/min to 330°C, and then held for 6 min. The transfer line and ion source temperatures were 250 and 200°C, respectively. Ions were generated by electron ionization (EI) at 0.93 kV. Spectra were recorded at 10000 u/s over the mass range 85–500 m/z . A standard alkane mixture (C₈–C₄₀) was injected at the beginning and end of the analysis for tentative identification.

To corroborate profiling data, metabolite quantification was performed solely for candidates of discriminant markers using authentic chemical standards at various concentrations. The final concentrations of the authentic standards were adjusted to 1, 10, 50, 100, 250, 500, 750, 1000, 1500, and 2000 μ M with the extraction solvent to construct a calibration curve. For extraction, the authentic standards were processed identically to the coffee bean samples. The standards were co-injected during sample analysis. Two blank solutions were prepared by adding only extraction solvent and distilled water, respectively. The limits of detection (LOD) and quantitation (LOQ) were determined via known protocols^{68,69}. The construction of the standard curve and quantitation were conducted using GC/MS Solution software (Shimadzu). No authentic standards were detected in either of the blank samples.

2.2.4. Data preprocessing and metabolite identification

Chromatographic data from GC/MS analysis were converted into ANDI files (Analytical Data Interchange Protocol, *.cdf) using the GC/MS Solution software package (Shimadzu). Peak detection, baseline correction, and peak alignment of retention times were

performed on the ANDI files using the freely available software package, MetAlign⁷⁰ with described parameters shown in **Table 2-2**. Spectra were normalized manually by adjusting the peak intensity against the ribitol internal standard.

Table 2-2. Set of MetAlign parameters for baseline correction and retention alignment.

Baseline and noise elimination parameters		
Retention begin (Scan nr)	1	
Retention end (Scan nr)	30000	
Maximum amplitude	1000000000	
Peak slope factor (x Noise)	2	
Peak threshold factor (x Noise)	4	
Peak threshold (Abs.value)	10	
Average peak width at half height (Scans)	25	
Scaling and aligning data sets		
Scaling options		No option
Initial peak search criteria	Scan Nr.	Max. Shift
Begin of 1 st region	0	20
End of 1 st region	30000	30
Begin of 2 nd region	0	0
End of 2 nd region	0	0
Tuning alignment options and criteria		Pre-align processing (iterative)
Maximum shift per 100 scans		35
Mass peak selection	1st iteration	Last iteration
Min. Factor (x noise)	7	7
Min. Nr. of masses	10	5
Select min. Nr per peak set	Group 1	Group 2
	0	0

According to the minimum reporting standards for chemical analysis from Chemical Analysis Working Group (CAWG) of Metabolomics Standards Initiative (MSI)⁷¹, level of metabolite identification are categorized into four levels.

- (i) Level 1: Identified compounds, require comparison with an authentic chemical standard with a minimum of two independent and orthogonal data (for example: retention time/retention index and mass spectrum), provide additional orthogonal data, *i.e.*, isotope labeling and if spectral matching is performed then the method and libraries should be described and made publicly available.

- (ii) Level 2, Putatively annotated compounds or tentatively identified compounds, based upon spectral similarities with public/commercial spectral libraries and without chemical reference standards.
- (iii) Level 3, putatively characterized compound classes, based on characteristic physicochemical properties of a chemical class of compounds
- (iv) Unknown compounds, metabolite identity does not match with any libraries or chemical standards.

In this thesis, identification level 1, 2 and 4 are utilized. Level 1 identification was performed based on method described in section 2.2.3. For level 2, I compared the retention indexes and unique mass spectra with in-house reference library. Retention indexes of the eluted compounds were calculated based on the standard alkane mixture. To simplify and accelerate the tentative identifications of compounds that were registered in the in-house library database, AIoutput2 (version 1.29) annotation software, developed in the authors' laboratory, was utilized⁷². The parameters were set as follows: height threshold: 1000, RT binning: 2, available index: retention index, analysis type: non-targeted, RI tolerance: 15, Match threshold: 0.75. For further comparison with the National Institute of Standards and Technology (NIST) spectral library, retention times were used instead. Metabolites without spectral similarities with these two libraries were classified as unknown compounds (level 4).

2.2.5. Multivariate data analysis

PCA and OPLS-DA were performed. OPLS-DA with an S alphabet-like plot, or S-plot, was chosen to isolate and select statistically significant and potentially biochemically interesting compounds. The variables that changed significantly are plotted at the top and bottom of the S-plot, and those that do not significantly contribute are plotted in the middle⁶⁶. A sevenfold cross validation was carried out to assess the accuracy of the discrimination model in practice. The goodness-of-fit (R^2) and predictability (Q^2) parameters were then determined. Analysis was performed with commercial software, SIMCA-P+ version 12 (Umetrics, Umeå, Sweden). Data were Pareto-scaled (centering + 1/SD) to reduce the effect of noise in the chromatograms.

2.3. Results and discussion

2.3.1. GC/MS-based metabolite profiling of civet coffee

GC-Q/MS analysis was performed on aqueous extracts of coffee beans to investigate the differences in their metabolite profiles and select discriminant markers for robust authentication. In addition, this research focused on increasing the scientific information about civet coffee. A quadrupole mass spectrometer was selected because of its availability as the most widely used mass analyzer. However, a conventional Q/MS can be operated only at a slow scan rate⁷³. With processor improvements and high-speed data processing, newly developed GC-Q/MS instruments provide increased sensitivity at high scan speeds of up to 10.000 u/s⁷⁴.

Because of their broad cultivation areas and commercial profitability, *C. arabica* and *C. canephora*, which represent 65% and 35% of the total annual coffee trade, respectively, were utilized for metabolomics analysis⁴⁶. A total of 53 out of 182 reliable peaks from 21 coffee beans were extracted using MetAlign. Moreover, 30 compounds were tentatively identified by comparison with in-house library (by retention index) and the NIST library (by retention time); ten of these were identified by co-injection with an authentic standard. Tentatively identified components consisted of organic acids, sugars, amino acids, and other compounds (**Supplementary Table 2-2**). Previously reported coffee bean constituents, including chlorogenic, quinic, succinic, citric, and malic acids; caffeine, one of the compounds supplying bitter taste in coffee; and sucrose, the most abundant simple carbohydrate, were identified⁷⁵⁻⁷⁹.

In recent research, unsupervised analysis, PCA, has been employed for data exploration and to visualize information based on sample variance^{35,80}. A PCA score plot derived from the 21 coffee beans differentiated two data groups based on their species, Arabica and Robusta (**Fig. 2-2**), and resulted in a goodness-of-fit parameter (R^2) of 0.844. Caffeine and quinic acid were significant for the Robusta coffee data sets, whereas the Arabica data set was mainly supported by various organic acids such as malic, chlorogenic, citric, and succinic acids. The data differentiation was explained by 42.9% of variance along

PC1. The results indicated that genetic diversity more strongly influenced the data separation than animal perturbation.

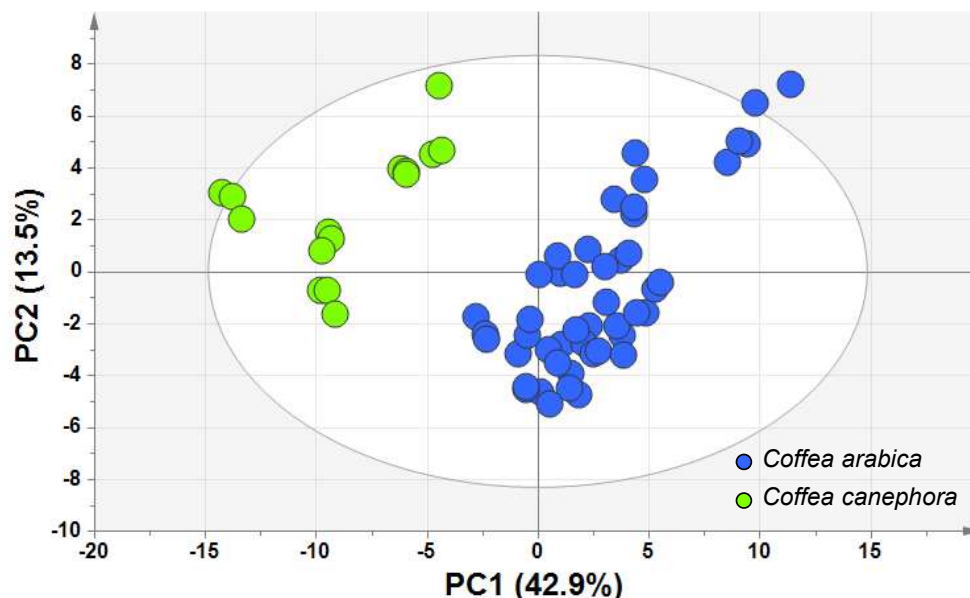


Fig. 2-2. PCA score plot of 21 experimental coffee set. Arabica (blue-filled circle) and Robusta coffee (green-filled circle) were clearly separated.

Because of the large variance among coffee species, sample differentiation based on the type of coffee, civet coffee or regular coffee, could not be observed. Additional analyses were carried out independently for each coffee species originating from the same cultivation area. The PCA score plot revealed data separation based on the type of coffee, in which civet coffee and regular coffee could be clearly separated (**Supplementary Fig. 2-1**). For the Arabica coffee data set, the separation was explained by 45.5% and 23.7% variances in PC1 and PC2, respectively. By PC2, civet coffee was closely clustered in the same region, whereas regular coffees tended to separate based on their cultivation areas. In the loading plot, malic and glycolic acids contributed highly to the civet coffee data (**Supplementary Fig. 2-1**). Thereby, coffee beans may possess similar profiles after animal digestion. Differences in cultivation areas were considered to have the least significance for data separation. In Robusta coffee, a clear separation between civet coffee and regular coffee was observed, which was explained by 79.1% variance of PC1. Significant compounds for separation, including inositol and pyroglutamic acid for civet coffee and quinic acid for regular coffee, were observed.

2.3.2. Discriminant analysis to select candidates for discriminant markers

An overview of all data samples was provided by the unsupervised analysis, PCA. However, detailed information regarding compounds contributing to the data differentiation between civet coffee and regular coffee remained unclear. Therefore, coffee bean data sets were subjected to supervised discriminant analysis (OPLS-DA). For analyses having two or more classes, OPLS-DA is the most suitable platform for isolating and selecting differentiation markers. Compounds with reliable, high contributions to the model may possess potentially biochemically interesting characteristics; thus, they can be selected as biomarker candidates⁶⁶. All OPLS-DA models exhibited R^2 and Q^2 values greater than 0.8, which would be categorized as excellent⁸¹. In addition, all models were in the range of validity after permutation tests using 200 variables. The model was considered valid after permutation for those that met the following criteria: R^2Y -intercepts and Q^2 -intercepts which did not exceed 0.3–0.4 and 0.05, respectively⁸².

Potential candidates for discriminant markers can be selected via S-plots by setting the cut-off for covariance, $p[1]$, and the correlation value, $p[corr]$, to $> |0.2|$. S-plots of the coffee data sets are shown in **Fig. 2-3B** and **2-3D**. In addition to cut-off values, candidates for discriminant markers were selected by variable importance in projection values (VIP). Large VIP values (> 1) are more relevant for model construction. The OPLS-DA score plot of Arabica coffee data sets is shown in **Fig. 2-3A**. Discrimination between civet coffee and regular coffee was obtained. The model was evaluated with R^2 and Q^2 values of 0.965 and 0.892, respectively. Interestingly, compounds that were uncorrelated with civet coffee were quinic acid, caffeine, and caffeic acid. These compounds have been reported as contributors of bitterness as well as acidity in coffee^{75–78}. In contrast, compounds that were predictive to civet coffee, *i.e.*, over the cut-off value, included citric, malic, and glycolic acids. The OPLS-DA score plot of the Robusta coffee data sets (**Fig. 2-3C**) was explained by R^2 and Q^2 values of 0.957 and 0.818, respectively. Caffeine, one of the bitter principles in coffee, was found to be significantly correlated with Robusta civet coffee data sets. Robusta coffee has been reported to contain higher amounts of caffeine than Arabica. Thus, it tends to be bitter and flavorless, whereas Arabica coffee is considered to be milder, contain more aromatic compounds, and is more appreciated by the consumer⁸³.

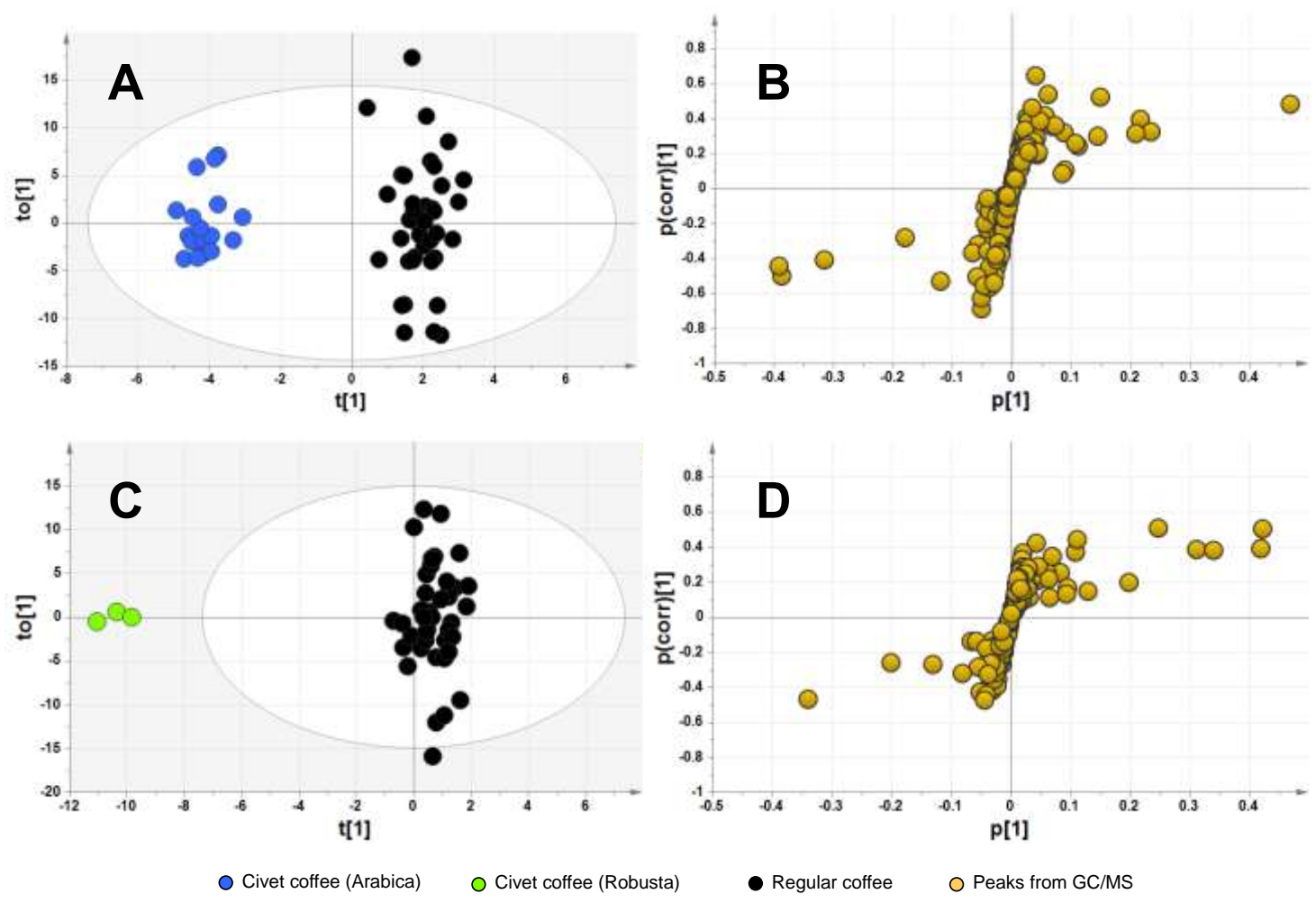


Fig. 2-3. OPLS-DA score plots (A, C) and loadings of S-plots (B, D) derived from experimental coffee set. Significant compounds were selected for p and $p[corr] > |0.2|$.

Candidates for discriminant markers for the authenticity assessment of Arabica and Robusta coffees are listed in **Table 2-3**. Discriminant markers were chosen independently for the Arabica and Robusta coffee. To confirm whether these selected markers were generated as a result of animal digestion, I investigated cause-effect relationships by quantitating the discriminant marker candidates in green and roasted coffee beans from controlled processing, pre- and post- animal digestion (Samples No. 5 and No. 11 experimental set). The results are displayed in **Supplementary Fig. 2-2**. In both the raw and roasted beans, citric acid was present in higher concentration after animal digestion, exhibiting a significant value difference ($p < 0.05$) between civet coffee and regular coffee. The concentration of caffeine was also increased after digestion, but the difference was insignificant ($p > 0.05$). As a result of roasting, the glycolic acid concentration increased dramatically ($p < 0.001$) from 0.8 to 25–28 $\mu\text{g/L}$. The production of aliphatic acids, including formic, acetic, glycolic, and lactic acids, has been reported during coffee roasting⁸⁴. Therefore, among the selected marker candidates, I confirmed citric acid as a potential marker generated by animal digestion. Citric acid, malic acid, quinic acid, and chlorogenic acid are the main acids in coffee, and acidity is generated by complex reactions involving these organic acids during roasting⁷⁶. Civet coffee has been reported to exhibit slightly higher acidity than regular coffee⁵⁷. However, the correlation between the increased levels of particular acids as result of animal digestion and the total acidity in coffee after roasting remains obscure and requires further investigation.

Table 2-3. Candidates of discriminant markers from OPLS-DA and analytical parameters for quantitation

Discriminant marker	RT (min)	VIP	RSD [%] ($n = 3$)		Linearity		LOD ($\mu\text{g/L}$)	LOQ ($\mu\text{g/L}$)
			RT	Area ^a	R^2	Range (μM)		
Glycolic acid	4.96	3.93	0.12	1.87	0.9999	1-1000	0.021	0.066
Malic acid	9.05	5.53	0.05	2.29	0.9996	1-1000	0.043	0.132
Pyroglutamic acid	9.43	1.7	0.05	3.36	0.9992	1-750	0.054	0.164
Citric acid	11.61	5.6	0.04	3.29	0.9997	1-1000	0.504	1.526
Caffeine	12.18	2.28	0.04	3.51	0.9961	100-2000	1.531	4.638
Inositol	13.45	4.47	0.03	5.09	0.9974	1-1000	0.082	0.247

^aat 100 μM

2.3.3. Validation of the applicability of discriminant markers for authenticity assessment

To verify the applicability of the selected marker candidates, I analyzed a validation coffee set that included authentic civet coffee, commercial civet coffee, commercial regular coffee, fake coffee, and coffee blend. With the exception of the authentic coffee, the remaining samples were purchased commercially. Generally, from harvest to pre-roasting, samples labeled “commercial civet coffee” and “commercial regular coffee” were processed similarly to the corresponding coffees in the experimental set. However, in some cases, different roasting parameters were applied. Fake coffee was processed to approximate the sensory profile of civet coffee⁶⁰. Commercial regular coffees were selected from different production areas.

To examine the effectiveness of the selected markers in differentiating pure and coffee blends, I mixed two commercial civet coffees, civet coffee with brand of “Golden” and “Wahana”, with a commercial regular coffee (Wahana regular) in a 50:50 (wt%) ratio. This would also compare the applicability of the discriminant markers when coffee beans from the same and different production areas were blended. Despite being selected independently, the six marker candidates were used together for method validation.

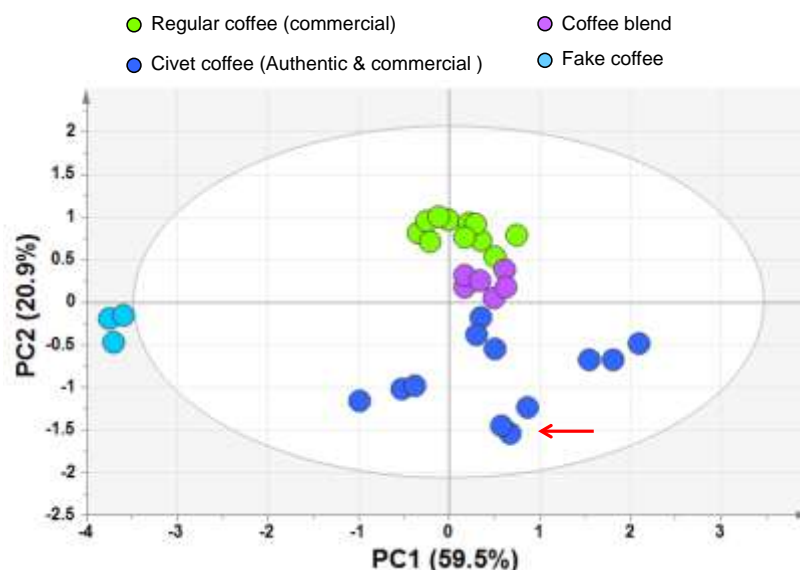


Fig. 2-4. PCA score plot of validation coffee set. Red arrow represents authentic civet coffee.

By subjecting all detected peaks to PCA, samples were populated into four clusters. The largest variance corresponded to fake coffee, as its results were clearly separated from others (**Supplementary Fig. 2-3**). Next, I projected the six marker candidates as an inclusion list into the PCA to obtain an overview of their applicability toward sample differentiation. Similarly to the previous results, separation of the four coffee groups was observed. The PCA was explained by 59.5% and 20.9% variances in PC1 and PC2, respectively (**Fig. 2-4**). Fake coffee was clustered away by PC1. Separation was likely because of attempts by the producer to obtain a profile similar to Kopi Luwak. In PC2, commercial civet coffee, coffee blend, and commercial regular coffee could be differentiated. Both authentic and commercial civet coffee were clustered within a close distribution area. Regardless of their origins and processing (roasting) parameters, commercial regular coffee data were populated in a close area, suggesting that these factors had the least significance for data separation. From the loading plot information, citric acid, malic acid, and inositol exhibited high contribution values for the civet coffee data sets. Interestingly, these three marker candidates also showed the highest VIP values for constructing the discriminant model (**Table 1**).

To display the applicability of the selected discriminant markers in the differentiation of samples in the validation set, box plots were constructed using the relative peak intensities of citric acid, malic acid, and inositol. Two different sets of coffee samples, set A and B were used for making the box plot (**Supplementary Table 2-1**). The box plots of malic acid and citric acid were able to differentiate commercial civet coffee (civet coffee “Wahana”), coffee blend, commercial regular coffee (regular coffee “Wahana”), and fake coffee. However, the inositol box plot failed to differentiate these samples. Hence, I selected a double marker that employed an inositol-pyroglutamic acid ratio (**Fig. 2-5**). Pyroglutamic acid was selected because it had the lowest contribution toward the separation of civet coffee and regular coffee (**Supplementary Fig. 2-4**). The box plot for the other commercial civet coffee (civet coffee “Golden”) and the comparisons with its coffee blend as well as fake coffee is displayed in **Supplementary Fig. 2-5**.

I confirmed the ratio of the coffee blend by quantifying the discriminant marker constituents. The analytical parameters for quantitation are shown in **Table 1**. All authentic standards exhibited good linearity (0.99 or higher) and good repeatability for at least seven

points in the applied concentration range in which analysis could be performed. To examine the quantitation validity, the LOD and LOQ for each discriminant marker were determined. The concentrations of the discriminant marker candidates in the coffee samples were determined to be higher than the LOD and LOQ of authentic standards. The concentration ratios of the selected markers, malic acid, citric acid, and the inositol-pyroglutamic acid ratio, in all the sample blends ranged from 47.76% to 53.73%. This result showed a relatively low error in terms of the ratios of the discriminant markers in sample blends compared with their actual values. Moreover, the concentration of each discriminant marker corresponded well with the box plot constructed from its respective peak intensity (**Fig. 2-4** and **Supplementary Fig. 2-5**). Hence, I confirmed the feasibility of using the proposed strategy for the robust authentication of coffee blend in a 50:50 (wt%) ratio.

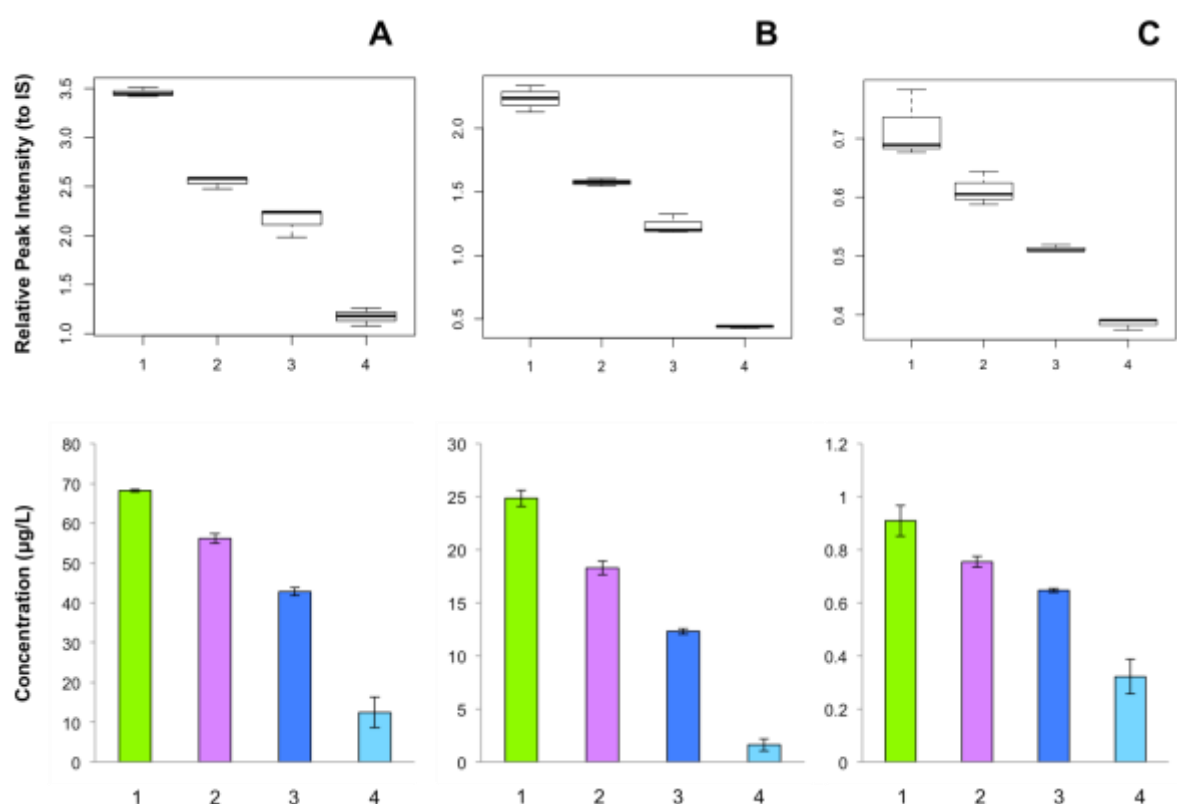


Fig. 2-5. Box plot of peak intensity and concentrations of selected discriminant markers from set A. Sample description of set A is displayed in **Supplementary Table 2-1**. (A) citric acid, (B) malic acid, and (C) inositol/pyroglutamic acid, of four coffee samples, (1) civet coffee “Wahana”, (2) coffee blend set B, (3) regular coffee “Wahana”, and (4) fake coffee.

2.4. Conclusions

In summary, this investigation represents the first attempt to address discriminant markers for the authentication of civet coffee. Sample differentiation was greatly influenced by genetic diversity (coffee species), followed by decreasing contributions from animal perturbation and cultivation area. Because of the great variation among coffee species, candidates for the discriminant markers were selected independently for each species. The selected discriminant marker candidates were verified for the authentication of commercial coffee products. The proposed markers were able to differentiate commercial civet coffee, commercial regular coffee, and fake coffee. In addition, at a certain ratio (50 wt% civet coffee content), the feasibility of employing these discriminant markers to differentiate pure and mixed coffee was acceptable. This finding highlighted the utility of metabolic profiling using GC/MS combined with multivariate analysis for the selection of discriminant markers for the authenticity assessments of valuable agricultural products. Discriminant markers are expected to perform as sole markers or in combination with sensory analysis by trained experts for the authentication of civet coffee.

Chapter 3

Application of gas chromatography-flame ionization detector (GC-FID)-based metabolite fingerprinting for authentication of Asian palm civet coffee

3.1. Introduction

In Chapter 2, selection and successful validation of discriminant markers for differentiation of civet coffee, authentic and commercial, regular coffee, fake coffee and coffee blend with 50 wt % of civet coffee content was demonstrated, presenting the portion of development of reliable method for authenticity evaluation of civet coffee. In this following two chapters, application studies will be presented to confirm the developed method for practical use.

As mentioned previously, metabolomics – the comprehensive study of metabolome – provides a snapshot of dynamics in metabolic pathways. Unlike the genome and transcriptome, metabolome has been considered as the best descriptor of an organism's phenotype^{85,86}. Dramatic changes within a cell due to various perturbations are presumed to be reflected in the metabolite profiles. Many analytical systems with various degrees of sensitivity and specificity have been developed and widely applied to metabolomics. Particularly for coffee metabolomics, several studies were reported utilizing mass spectrometry (MS)^{49,52,53,55}, Inductively coupled plasma atomic emission spectrometry (ICP-AES)⁵⁰, Fourier transform infrared (FTIR) spectroscopy⁵¹, Raman spectroscopy^{46,54}, and nuclear magnetic resonance (NMR)^{56,79}. MS has gained popularity in the past decade because of its superior sensitivity, thus it is extensively used for metabolomics studies. Coupling MS with available separation techniques such as GC and LC facilitates the selectivity of a wide range of compounds⁸⁷⁻⁹⁰. In Chapter 2, I performed metabolite profiling employing GC/MS to

pinpoint potential discrimination marker candidates for civet coffee and regular coffee. A combination of GC and MS provides straightforward analyses within a single-step extraction with good reproducibility. However, the cost for maintenance and the machine itself is relatively expensive. Therefore, it is necessary to develop a protocol that is robust, sensitive and cost-effective for quality and authenticity evaluation of civet coffee.

Gas chromatography with a universal detector such as flame ion detector (FID), has established itself as an inexpensive analytical system which can cover high-throughput analysis of carbon containing compounds⁹¹. A few applications of GC/FID in metabolomics studies were reported for quality control of transgenic rice⁹², herbal plants^{18,93}, green tea⁹¹, and pine wood tree⁹⁴. The availability of GC/FID in most small coffee factories and research institutes favors its application for routine analysis to evaluate authenticity of civet coffee. In this chapter, I attempted to develop a robust authentication technique for civet coffee using GC/FID-based metabolite fingerprinting. Metabolite fingerprinting presents rapid sample classification according to biological background rather than focusing on the small set of individual compounds thus it is a suitable approach for large size sample screening. The application of GC-FID in this study is to represent the use of cost effective instrument for rapid and reliable quality evaluation of civet coffee in industry. An alternative to employ other analytical inexpensive instrument that can cover organic analysis, a chemical class of civet coffee's marker, *e.g.* HPLC, should be applicable.

At first, GC/FID was employed for fingerprinting of metabolites extracted from the first set of coffee samples collected from different production areas. A discrimination model of the coffee's metabolite profiles, constructed by orthogonal projection to latent structures-discriminant analysis (OPLS-DA), was then compared to the previously reported GC/MS data to verify repeatability of the established protocol. Lastly, a second set of coffee samples composed of commercial coffees and coffee blends, were analyzed to confirm the validity of the method.

3.2. Experimental section

3.2.1. Coffee bean materials and chemicals

Samples were divided into experimental and validation coffee sets. The first set included twenty coffee beans that were collected from several cultivation areas in Indonesia. It consisted of civet coffee (no. 1-6, Arabica) that had been digested by civet, and undigested beans referred to as regular coffees (no. 7-20, Arabica and Robusta). All coffee samples were treated identically for post harvesting. Coffee roasting and storage were identical to the described protocol in section 2.2.1.

The second set of coffee samples included 3 civet coffees and 3 regular coffees were bought commercially and 2 additional authentic civet coffees from the Indonesian Coffee and Cocoa Research Institute. In addition, each civet coffee and regular coffee was mixed in equal proportions (50:50, wt %) to obtain representative coffee blends. A total of 17 coffee samples, 8 pure and 9 coffee blends, were then analyzed by GC/FID to verify the established protocol for coffee authentication. All coffee samples were measured in triplicates. The sample descriptions are shown in **Table 3-1 and Supplementary Table 3-1**.

To provide a verification for applicability of discriminant markers in metabolite fingerprinting strategy, chemical standards for confirmation of compound identification and their providers were as follows: citric acid and malic acid (Nacalai-Tesque), sucrose (Kisida chemical), pyroglutamic acid (ICN Biomedicals), inositol (Wako), quinic acid, chlorogenic acid and glycolic acid (Sigma Aldrich).

3.2.2. Coffee beans extraction and derivatization

Metabolite extraction and chemical derivatization were done using identical methods that have been described in detail in section 2.2.2.

Table 3-1

List of coffee samples (Experimental coffee set) for GC/FID analysis

No.	Production area	Species	Type	Harvest year	Samples ID ^a
1	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	29
2	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	31
3	Sumatra	<i>Coffea arabica</i>	Civet coffee	2011	39
4	Bali	<i>Coffea arabica</i>	Civet coffee	2011	33
5	Java	<i>Coffea arabica</i>	Civet coffee	2011	24
6	Java	<i>Coffea arabica</i>	Civet coffee	2011	26
7	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	30
8	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	32
9	Sumatra	<i>Coffea arabica</i>	Regular coffee	2011	40
10	Bali	<i>Coffea arabica</i>	Regular coffee	2010	34
11	Java	<i>Coffea arabica</i>	Regular coffee	2011	22
12	Java	<i>Coffea arabica</i>	Regular coffee	2011	41
13	Java	<i>Coffea arabica</i>	Regular coffee	2010	42
14	Java	<i>Coffea arabica</i>	Regular coffee	2011	45
15	Bali	<i>Coffea arabica</i>	Regular coffee	2011	43
16	Bali	<i>Coffea arabica</i>	Regular coffee	2011	44
17	Sumatra	<i>Coffea canephora</i>	Regular coffee	2010	36
18	Java	<i>Coffea canephora</i>	Regular coffee	2011	46
19	Java	<i>Coffea canephora</i>	Regular coffee	2011	47
20	Java	<i>Coffea canephora</i>	Regular coffee	2010	51

^aFor multivariate analysis

3.2.3. Metabolite measurement with GC/MS and GC/FID

The typical workflow for GC/MS analysis of aqueous coffee bean extract has been described in detail in Chapter 2. GC/FID was conducted on a GC-2010 (Shimadzu, Kyoto, Japan) installed with an AOC-20s autosampler and AOC-20i autoinjector. One microliter of each derivatized sample was injected in split mode, 25/1 (v/v). To establish proper comparison and validation with the reported GC/MS data, the same type of column, CP-SIL 8 CB low bleed column (0.25 mm x 30 m, 0.25 μ m, Varian Inc., Palo Alto, California, USA), and identical temperature program were applied to GC/FID analysis. The carrier gas (He) was maintained at a constant velocity of 45 cm/s. The injector and FID temperature were set at 230 and 320°C, respectively.

3.2.4. Data preprocessing and dataset construction

Raw chromatographic data of GC/FID were converted into CDF format using GCMS Solution software package (Shimadzu, Kyoto, Japan). The converted files were subjected to baseline correction, normalization and alignment of retention times using the in-house software, PiroTran ver 1.41 (GL Sciences, Tokyo, Japan), utilizing these following parameters: Removal variable of solvents (0 to 4.2 min), internal standard normalization (10.96 min, 5 point), alignment (100 point), target (1), binning interval (20), baseline correction (5 point). The retention time of internal standard ribitol was confirmed with co-injection of authentic chemical standard before being utilized as reference for normalization and retention time alignment. To reduce the run-to-run variation, the threshold for peak intensity (RSD) was set to < 20%, in each measurement replicate. To construct the data matrix, in which each row and column represent the samples and relative peak intensity at certain retention time, respectively, the outcome data were imported into Pirouette ver 4.0 (Infometrix, Inc, Woodinville, Washington, USA). The data matrix was then subjected to multivariate analysis.

3.2.5. Multivariate data analysis

Multivariate analysis was carried out using SIMCA-P+ ver. 13 (Umetrics, Umeå, Sweden) to reduce dimensionality of the huge MS data and extract biological interpretation. PCA and OPLS-DA were used to decipher the relationships between two data matrices, X (predicted variables), and Y (observed variables)⁶⁶. Here, the chromatographic GC/FID data were used as X and for Y , the binary vector of 0 and 1 was assigned for civet coffee and regular coffee, respectively. The data were Pareto scaled prior to analysis without transformation.

3.3. Results and discussion

3.3.1. Chromatographic data of GC/FID

Representative GC/FID and GC/MS chromatograms of aqueous coffee extracts are shown in **Fig. 3-1**. I compared the chromatogram obtained from GC/FID with the one from GC/MS analysis using the same coffee extract and column type. The chromatographic data of GC/FID and GC/MS gave similar metabolite patterns, which contained the peaks from diverse metabolites, *i.e.*, glycolic acid (peak no. 1), malic acid (peak no. 2), pyroglutamic acid (peak no. 3), citric acid, (peak no. 4) quinic acid (peak no. 5), inositol (peak no. 6), sucrose (peak no. 7) and chlorogenic acid (peak no. 8). A total of 678 peaks were obtained from GC/FID, compared to 182 peaks from GC/MS analysis.

For metabolite fingerprinting, it is not necessary to determine the individual information of every peak¹. Nonetheless to confirm the overall quality of GC/FID analysis, peak confirmation of the GC/FID chromatogram was performed by comparing to the identified peaks in the GC/MS data and co-injection of authentic chemical standards. Whilst most of the detected peaks that represented key coffee metabolites were identical between GC/FID and GC/MS, I also observed a shift in their retention times, such as in glycolic acid (5.02 and 4.96 min), malic acid (9.11 and 9.05 min), and citric acid (11.68 and 11.61 min), respectively. Since metabolomics data are often subject to unwanted variations⁹⁵, the retention time shift reported here, albeit not severe, may be due to experimental variation between analytical instruments.

As shown in **Fig. 3-1**, the overall chromatographic profiles between GC/FID and GC/MS were similar. However, it is noticeable that GC/FID analysis provided higher relative peak intensity than GC/MS for almost all detected peaks. The higher relative peak intensity often implies higher sensitivity, which has been seen from the signal-to noise (S/N) ratio, as GC/FID analysis has been described to generate higher sensitivity compared to the mass detector which frequently operated in a full-scan mode for gathering entire profiles of biological samples^{1,18,91}. Measurement of total ions over mass range resulted in the limitation of sensitivity for the mass detector. The efficient reduction of relative intensity for detected peaks within the range of 4.2 and 6 min was also observable for GC/FID analysis. The peaks were confirmed by comparison with the NIST library and identified as siloxane, common

peak contaminants from injector and vial septa (**Fig. 3-1A**). The result was explicable since FID primarily responds to a wide variety of carbon-containing organic compounds whereas a mass detector relies on the recognition of the entire ionized and fragmented molecules. The results suggested the practicability of using GC/FID for metabolite fingerprinting of coffee beans as it provided higher sensitivity over a similar range of detected compounds than GC/MS analysis.

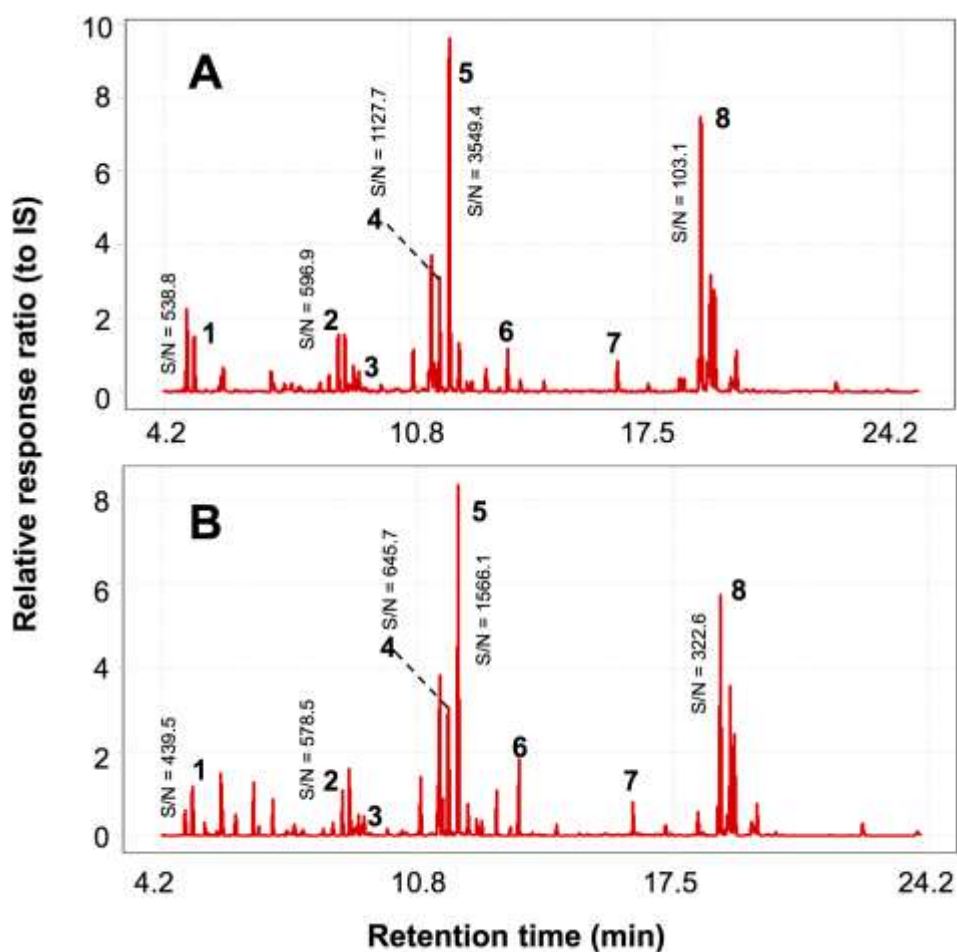


Fig. 3-1. Gas chromatograms of representative coffee bean extracts obtained from (A) GC/FID; and (B) GC/MS analysis. Both analyses used same column, CP-SIL 8 CB low bleed. Peak tentative identification: (1) glycolic acid, (2) malic acid, (3) pyroglutamic acid, (4) citric acid, (5) quinic acid, (6) inositol, (7) sucrose, and (8) chlorogenic acid. S/N, signal-to-noise ratio.

3.3.2. Multivariate data analysis of coffee bean extracts by GC/FID-based metabolite fingerprinting

Metabolite fingerprinting of coffee extracts for the development of rapid assessment method was done using GC coupled with a universal detector, FID. Metabolite identification has been reported as tedious work and a major challenge in the metabolomics workflow^{1,96-98}. However, determination of the individual level and identity for each metabolite is not a key requirement for metabolite fingerprinting strategy therefore, metabolite fingerprinting is suitable for quality screening of large number of coffee samples.

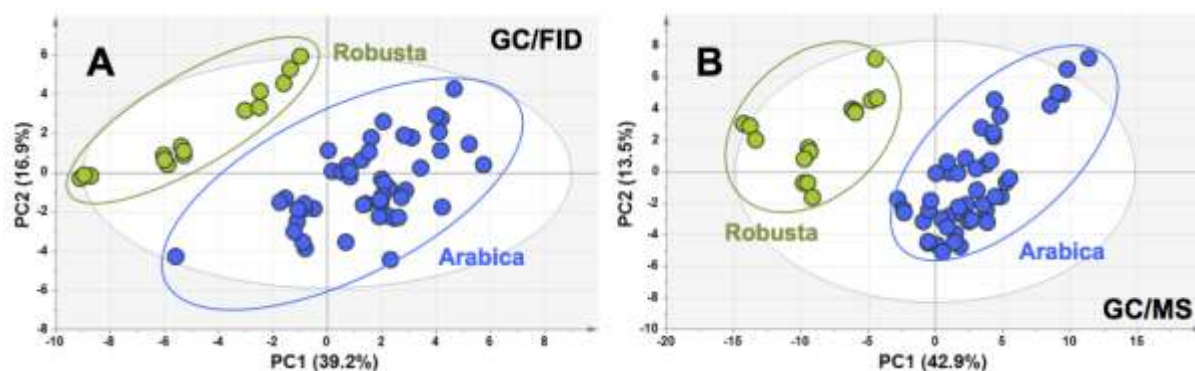


Fig. 3-2. PCA score plot derived from (A) GC/FID and (B) GC/MS analyses. Green and blue filled-circles represent Robusta and Arabica coffee, respectively.

A comparison of the multivariate analyses obtained from GC/MS analysis with that of GC/FID in order to evaluate the performance quality of the latter platform was carried out. The quality of PCA and OPLS-DA models constructed from GC/MS data has been described in Chapter 2. PCA was performed as unsupervised and non-biased method to reduce the dimensionality of multivariate data and visualize the differences in the metabolite fingerprints of coffee extracts. The score plot of PCA derived from GC/FID analysis, where the first two components (PC1 and PC2) accounted for 39.2% and 16.9% of the total variance, revealed distinct separation between samples due to their genetic trait, Arabica and Robusta (**Fig. 3-2**). Arabica and Robusta were analyzed in this study to represent the two major coffee species traded annually⁴⁶. The Arabica coffee data set, comprised of both civet and regular coffee, clearly separated from Robusta coffee. Results indicate that genetic variability between species gave greater impact for data separation in comparison to perturbation during animal

digestion and cultivation area (along PC1-axis). The PCA obtained from GC/FID had better quality in terms of goodness-of-fit ($R^2X = 0.88$) compared to the GC/MS data ($R^2X = 0.84$) (Fig. 3-2).

OPLS-DA modeling to indicate the role of animal perturbation in coffee sample separation by setting a binary vector with the value 0 for the civet coffee class and 1 for the regular coffee class were made. A discrimination model was then built from the total data set, in which the chromatographic data were used as predicted variable (X) and the binary vector as the observed variable (Y). The OPLS-DA score plot of the total data set of GC/FID indicated the apparent clustering between samples on the basis of animal perturbation (Fig. 3-3). Civet coffee and regular coffee samples were clearly separated in the OPLS 1 (Fig. 3-3). Separation of Arabica and Robusta coffee remained observable in the right cluster occupied by regular coffee samples in OPLS 2. In GC/FID, the OPLS-DA model was built with an R^2Y value of 0.996 and a Q^2 value of 0.78 (Fig. 3-3A). The correlation coefficient (R^2Y), describes how a model fit a set of predicted data set related to class separation. The model derived from GC/FID data is less robust in terms of model predictability (Q^2) compared to the model from GC/MS data ($R^2Y = 0.965$ and $Q^2 = 0.892$, Fig. 3-3B). The lower Q^2 value in GC/FID may signify an overfitted model which resulted from the use of all confirmed peaks from data processing to construct the discrimination model in GC/FID. The use of irrelevant components increases the risk of overfitting and eventually create poor predictive precision⁹⁹⁻¹⁰¹. However, the Q^2 value ≥ 0.5 was widely considered as good and acceptable for a model derived from biological samples^{18,19,81}.

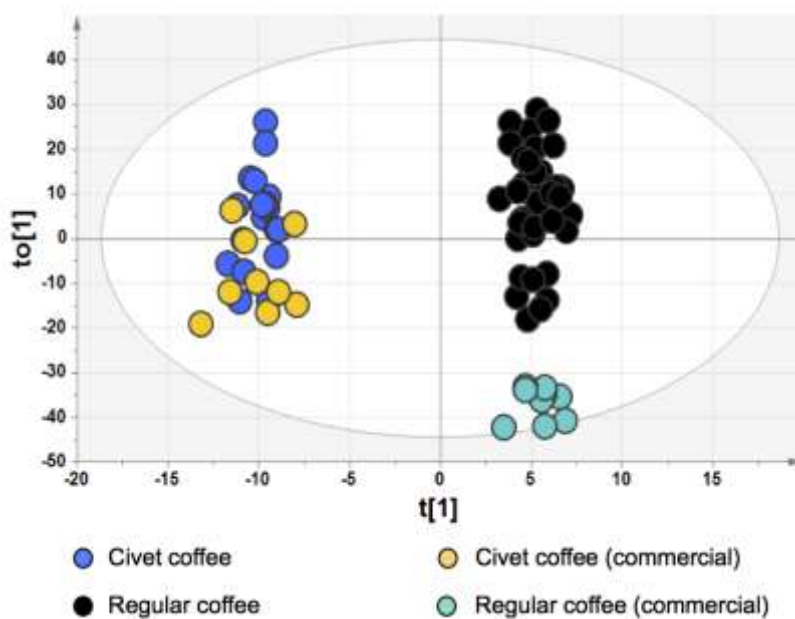


Fig. 3-4. OPLS-DA score plot of validation test. Civet coffee samples (commercial) (orange filled-circle) and commercial regular coffee (tosca-filled circle) were projected into the same class with civet and regular coffee samples from experimental set, respectively.

Permutation tests were performed in the PLS-DA model to confirm the quality of OPLS-DA model created from GC/FID. According to Setoyama *et al*, if the OPLS-DA model is overfitted, the R^2Y and Q^2 values would not virtually change after permutation⁴⁹. Both parameters were in the range of the requirement for a reliable model; R^2Y -intercepts fluctuated between 0.3–0.4 and Q^2 -intercept was below 0.05, respectively (**Supplementary Fig. 3-1**)^{82,102}. These denoted that there was a change in the values of the two parameters. Furthermore, external validation was then performed to estimate the accuracy of the performance of the discrimination model in practice.

I prepared and analyzed the GC/FID data of three commercial samples of civet and commercial regular coffee in separate days. The samples were then projected onto the constructed OPLS-DA model as external validation. **Fig. 3-4** showed that commercial both fit into the discriminant model; if those samples were classified into the same class with prior data ($R^2 = 0.982$ and $Q^2 = 0.741$). Additionally, the six samples from each class were left randomly (test set) and the OPLS-DA was performed three times for the remaining samples only (training set). The R^2Y and Q^2 values after cross validation were simultaneously

calculated and the values obtained were in the range of 0.907–0.957 and 0.616–0.667, respectively. Although a drop in R^2Y and Q^2 values was seen, the quality of the model was still acceptable. To visualize how accurately external validation will perform to differentiate samples, I set the cut off of the prediction at 0.5. As a result, almost all test samples were correctly classified except for one civet coffee sample (sample ID 24_3) (**Fig. 3-5**). Taken together, the practicability of the GC/FID coupled to metabolite fingerprinting strategy for rapid discrimination and prediction of new samples with statistical significance, was confirmed. The results also suggested that the combined techniques could effectively minimize variability, *i.e.*, error from day-to-day measurements.

Statistically significant variables contributing to the differentiation of civet coffee and regular coffee were selected from the S-plot of the OPLS-DA. S-plot combines both covariance (contribution or magnitude) and correlation (reliability between the variables (metabolites) with the model class designation⁶⁶. Consequently, on the basis of their contribution and reliability, the variables that changed significantly are plotted at the top and bottom of the S-plot, and those that do not significantly contribute are plotted in the middle. Three variables were highlighted in the S-plot of GC/FID, variable 5.02_, 9.11_, and 11.68_ (labeled after their retention time).

To confirm the reproducibility of the metabolite fingerprinting strategy, selection of variables from GC/FID was done for those with identical retention time to that of the biomarker from the reported GC/MS analysis and later confirmed with co-injection of the chemical standard for giving a valid identification (**Fig. 3-3C – 3-3D**). In GC/FID, the S-plot shows that variable 11.68_ had the highest contribution p , which means a high correlation $p(\text{corr})$, for class separation, followed by variable 5.02_ (**Fig. 3-3C**). Variable 9.11_ was plotted in the middle region with ambiguous significance level. The contribution of those variables in the model projection could also be explained using variable important in the projection (VIP). The average of 95% confidence interval VIP is equal to 1.0⁹¹; therefore large VIP values (> 1) are often considered relevant for explaining the OPLS-DA model.

Table 3-2 displays the comparison between the three variables extracted from the GC/FID data with corresponding biomarkers from GC/MS analysis. The VIP score of each

variable represents high contribution to the model. Lower VIP score for variable 9.11_ corroborated with the S-plot result. It is implied that passage through civet's digestive tract may enhance the level of particular organic acids⁵⁷. However, to understand the underlying biological meaning of these biomarkers as a result of animal digestion and its correlation to the sensory profile, further investigation is needed.

Table 3-2. Comparison of significant variables extracted from GC/FID and GC/MS.

No.	Variable ID ^a	GC/MS		GC/FID		
		RT (min)	VIP ^c	Variable ID ^b	RT (min)	VIP ^c
1	Glycolic acid	4.96	3.93	5.02_	5.02	4.44
2	Malic acid	9.05	5.53	9.11_	9.11	1.48
3	Citric acid	11.61	5.6	11.68_	11.68	8.39

^aconfirmed with co-injection with chemical standard and comparison of mass fragment spectra with NIST database

^bselected on the basis of RT comparison with GC/MS data and co-injection of chemical standard

^cVariable Importance in the Projection, extracted from OPLS-DA

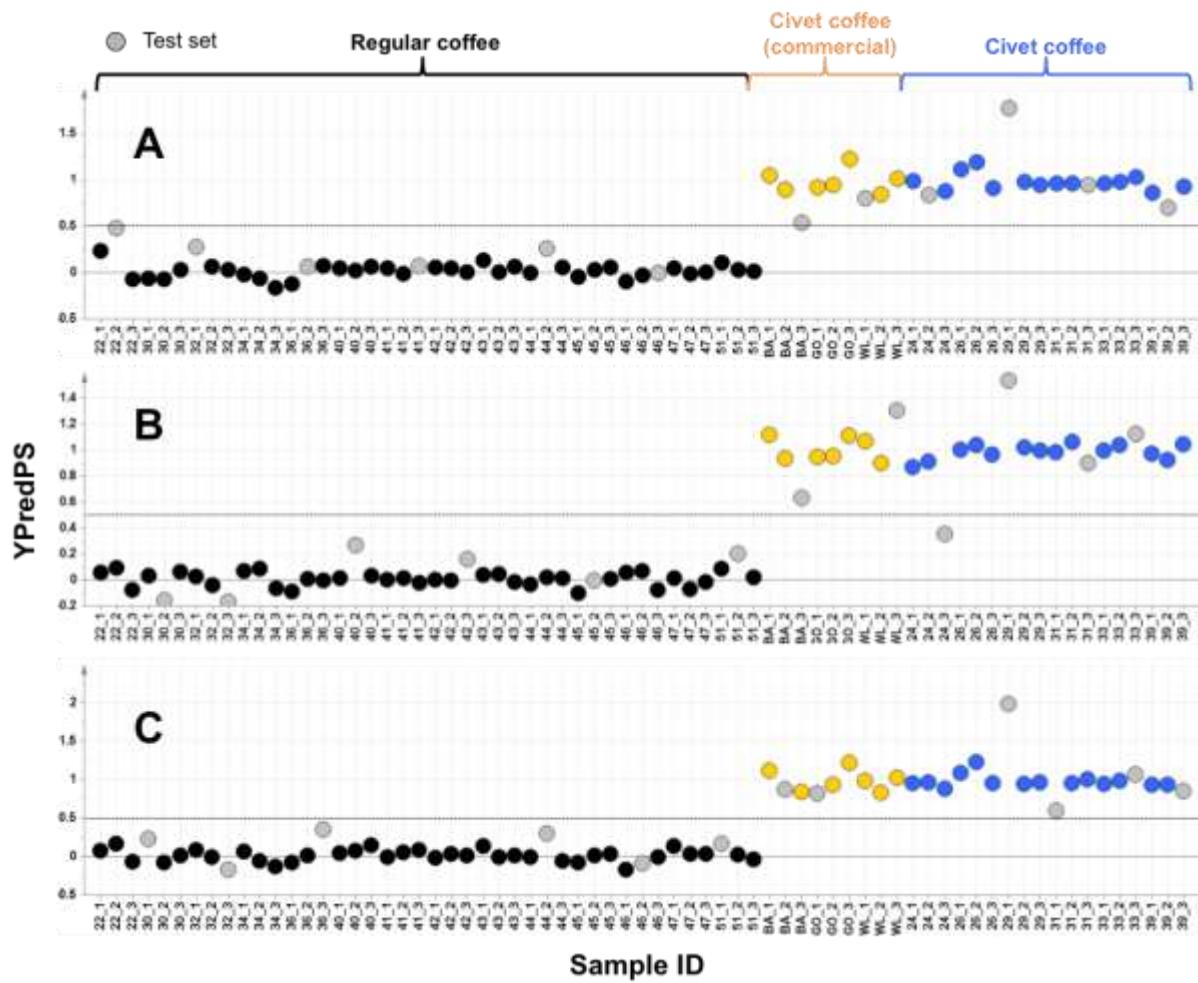


Fig. 3-5. Validation test of OPLS-DA model derived from GC/FID. The six samples from each class were left randomly (test set, grey filled-circle) and the OPLS-DA was performed three times for only remaining samples (training set, colored filled-circle) (A, B, C). The cutoff of the prediction (dashed line) was set at 0.5.

3.3.3. Validation of metabolite fingerprinting strategy to confirm authenticity of commercial coffee products

A set of commercial samples from the coffee market has been analyzed to provide scientific evidence of the GC/FID application in the coffee industry. Since processing commercial samples is based on the customers' preference, the roasting temperature may vary from experimental coffee. Commercially available regular and civet coffee were selected from different production areas. To set the validation threshold, I acquired two authentic civet coffees from different production years as benchmark samples. Furthermore, a total of 9 coffee blends were prepared from the combination of each commercial sample with mixing ratio of 50:50 (wt %). These four differentiation parameters, occurrence of perturbation, production area, roasting parameter and mixing ratio, would present comprehensive coverage for validation.

PCA modeling was applied to the validation coffee data set. As shown in **Fig. 3-6**, commercial coffee samples were distinguished from one another on account of perturbation and mixing ratio by 47.5% variance in PC1 ($R^2X = 0.838$). Both civet coffees, commercial and authentic, were populated in a wide margin at the right region of the intercept (0,0). Four civet coffee samples (2 commercial and 2 authentic coffee) from neighboring production areas were clustered within close range in PC2. It is suggested that the production area may play a significant role for separation of civet coffee. One commercial civet coffee (No. 5) was located outside the 95% confidence interval of the Hotelling's T^2 , thus marking the possibility of an outlier. Coffee blends were populated within the area around the intercept. This is probably linked to the equal mixing ratio between civet and regular coffee. Interestingly, regular coffee samples were populated in a close area, only partially separated owing to production area (foreign and local). According to the loading scatter plot (**Supplementary Fig. 3-2**), some significant variables, including variables 11.68_, 9.11_ and 5.02_, were captured and highlighted as responsible for the differentiation of commercial samples in the PCA score plot.

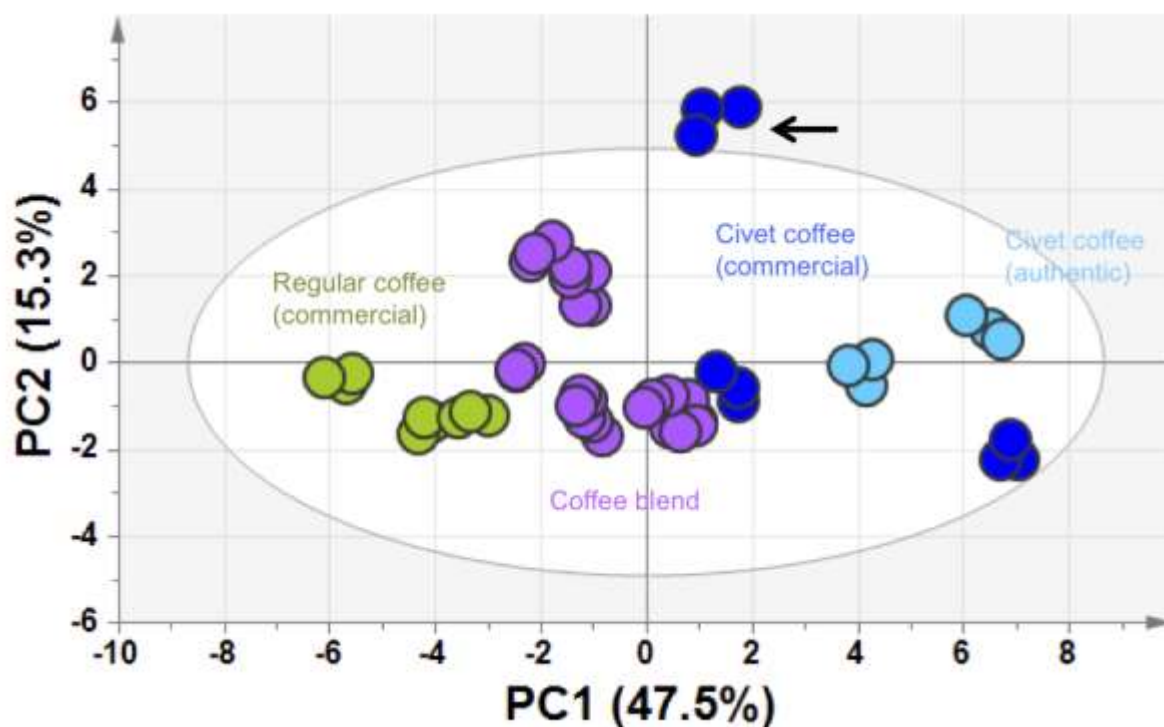


Fig. 3-6. PCA score plot derived from GC/FID chromatograms of 17 coffee samples in validation set (n = 3). Arrow indicates a possibility of counterfeit.

Fig. 3-7 summarizes the box plot construction to visualize performance of significant variables captured by PCA and OPLS-DA for sample classification. Relative peak intensities of significant variables were used to create a whisker-box plot. Generally, the box plot was divided into three groups on the account of relative intensity level; civet coffee, coffee blends, regular coffee. An overlap box plot was spotted only between civet coffee and coffee blend clusters due to equal mixing ratio, but not for civet coffee and regular coffee. The box plot clustering was clearly observed for variables 9.11_ and 11.68_, corresponding to malic and citric acid, respectively, in GC/MS analysis. Although represented with high VIP and correlation, such result could not be reproduced for variable 5.02_ (**Fig. 3-8**). Variable 5.02_ corresponds to glycolic acid in GC/MS analysis, a biomarker candidate that has been shown to have poor predictive performance for authentication (described in Chapter 2). The box plot pattern of variable 5.02_ conflicted with the other two variables, indicating a dependency to regular coffee. As shown in **Fig. 3-7A – 3-7B**, the commercial civet coffee No. 5 gave very low relative intensities for variable 9.11_ and 11.68_, therefore suffered from overlap with the

other clusters. This outlier can also be found in PCA score plot in **Fig. 3-6**, suggesting a possible counterfeit or a difference in coffee species used for analysis, *i.e.*, Robusta, since the rest of the civet coffee samples were Arabica coffee and display relevant margin in their relative intensities (regardless of roasting temperature and production area). If sample No. 5 is indeed civet coffee (Arabica), its relative intensity for those significant variables should be much higher. These results verified the feasibility of employing the significant variables obtained from GC/FID for practical use (for authentication). Collectively, the results demonstrate in principle that GC/FID coupled with metabolite fingerprinting is a good complementary and cost effective analysis platform for quality assessment of civet coffee.

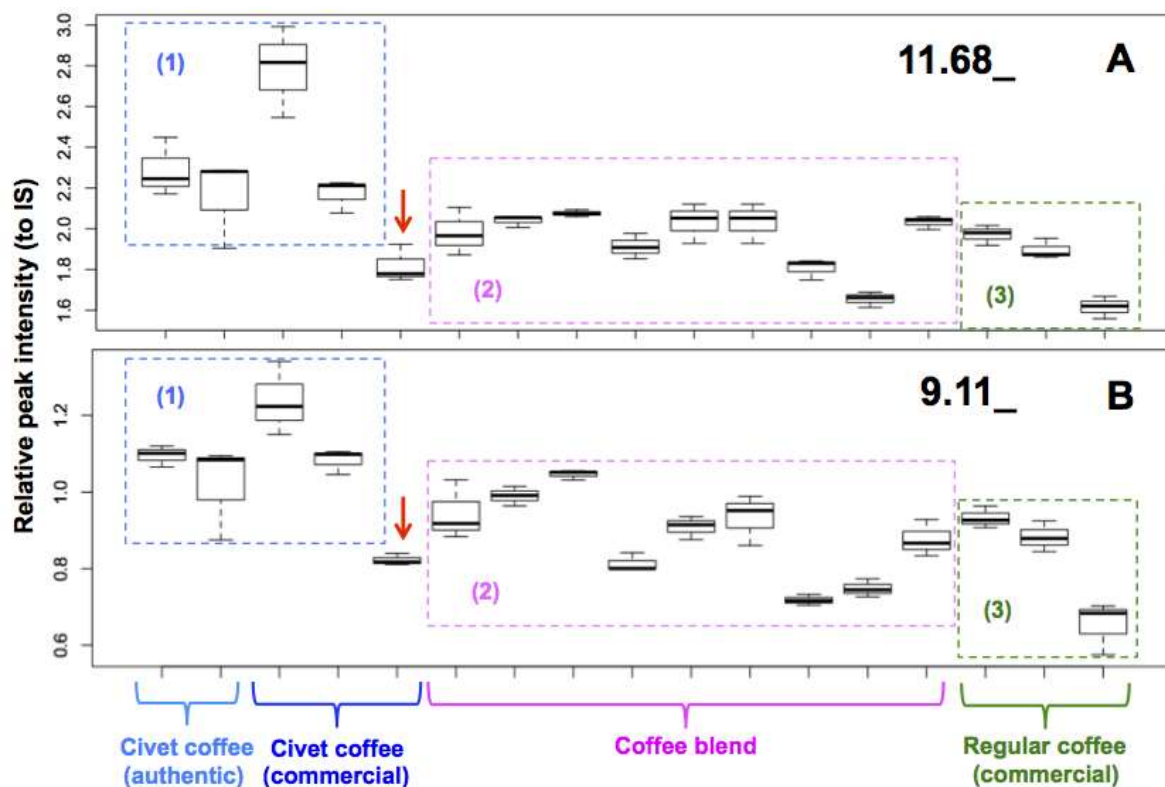


Fig. 3-7. Box plot of significantly different variables between samples in validation set from result of OPLS-DA, (A) 11.68_ and (B) 9.11_.

Both variables successfully differentiated group of commercial samples from one another, except one commercial civet coffee, sample No. 5 in **Supplementary Table 3-1** (red arrow) that was spotted as potential outlier in **Fig. 3-6**.

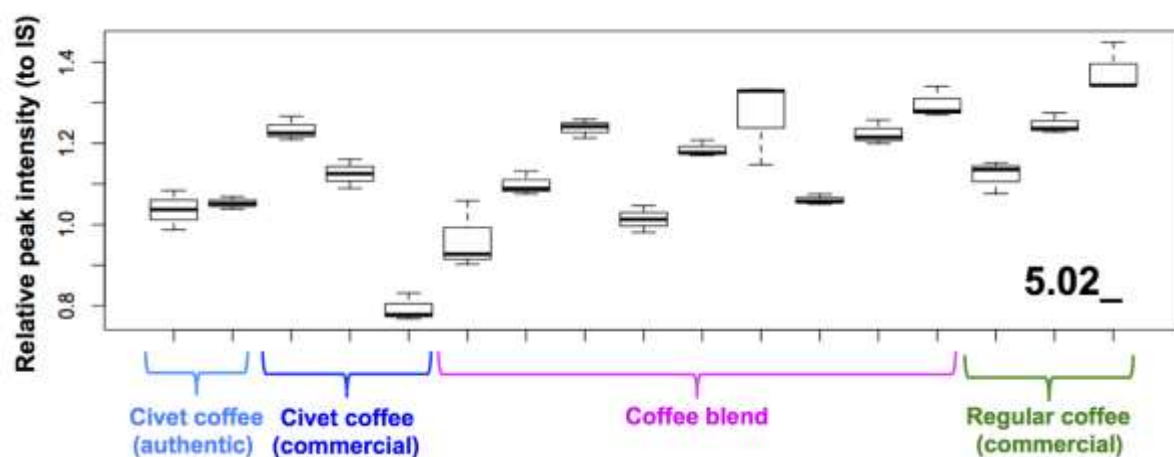


Fig. 3-8. Box plot of variable 5.02_.

3.4. Conclusions

In previous chapter, a metabolite profiling strategy for civet coffee quality assessment through GC/MS was demonstrated. In this method, automated procedures for tentative identification of unknown peaks by matching the information derived from mass spectral fragmentation patterns with chemical databases are necessary. However, the unambiguous and exhaustive metabolite identification in a biological system has been well documented as challenging^{1,96-98}. In contrast, an alternative strategy using GC/FID analysis paired with metabolite fingerprinting provided rapid classification of coffee samples without prior metabolite details. The significant compounds contributing to civet coffee quality assessment, such as malic and citric acid, showed better sensitivity in FID compared to an MS detector. By employing multivariate data analysis such as OPLS-DA, construction of a good prediction model to confirm authenticity of commercial samples was accomplished. This chapter presented that metabolite fingerprinting through GC/FID could effectively reproduce coffee quality prediction from the previous technique. With elimination of the tedious identification steps, the GC/FID system offered high-speed analysis for coffee quality assessment. This advantage can be beneficial to manufacturers for quality control, especially for authentication of commercial coffee and other agricultural products in industrial scale.

Chapter 4

Quantification of Asian palm civet coffee fraction in a coffee blend via metabolomics: a proof of concept

4.1. Introduction

Considering that illegal mixture of cheap coffee into civet coffee is a growing concern among consumers and government, I emphasized a proof of concept of metabolomics application to develop prediction model for determination of degree of coffee adulteration by quantifying mixing ratio of coffee blend.

Among many varieties of commercially marketed coffee, few are highly valued as “premium” or “gourmet” coffees due to their superior flavor or rarity. Of these premium coffees, Asian palm civet coffee has developed a reputation as one of the world’s priciest and rarest coffee due to its exotic production process. The transient fermentation inside the civet’s gut hypothetically adds a distinct flavor to the coffee beans. As a result, its rarity as well as the coffee’s exotic and unique production process ultimately accounts for its high selling price.

An important concern related to the price gap between civet and regular coffees is the growing attempt of fraud involving mixture of cheaper coffee into premium civet coffee. Blending between two or more coffees is deliberated as illegal when manufacturer does not emphasize the action during product’s labeling. Mixing of expensive coffee due to its valued trait with cheap beans has been considered as one of the most common incidents regarding coffee adulteration. It is therefore essential to develop robust methods to determine the ratio of civet coffee in blends. Particularly for agricultural products such as coffee and tea, quality evaluation has been conventionally assessed on the basis of human sensory perception^{103–105}. However, this method tended to be highly subjective with only up to 20% precision⁴⁶.

Moreover, hiring highly trained specialists to assess product quality is unavoidably expensive. Visual examination is only possible for green coffee beans, as civet coffee appears to be reddish and darker in color compared to non-animal treated coffee⁵⁷. When coffee is roasted, a color-based discrimination might lead to false results. Therefore, an alternative, instrument- and non-human-based measurement method to estimate the quality of coffee similar to the previously developed protocols for various food and agricultural products should be established^{6,34,106,107}.

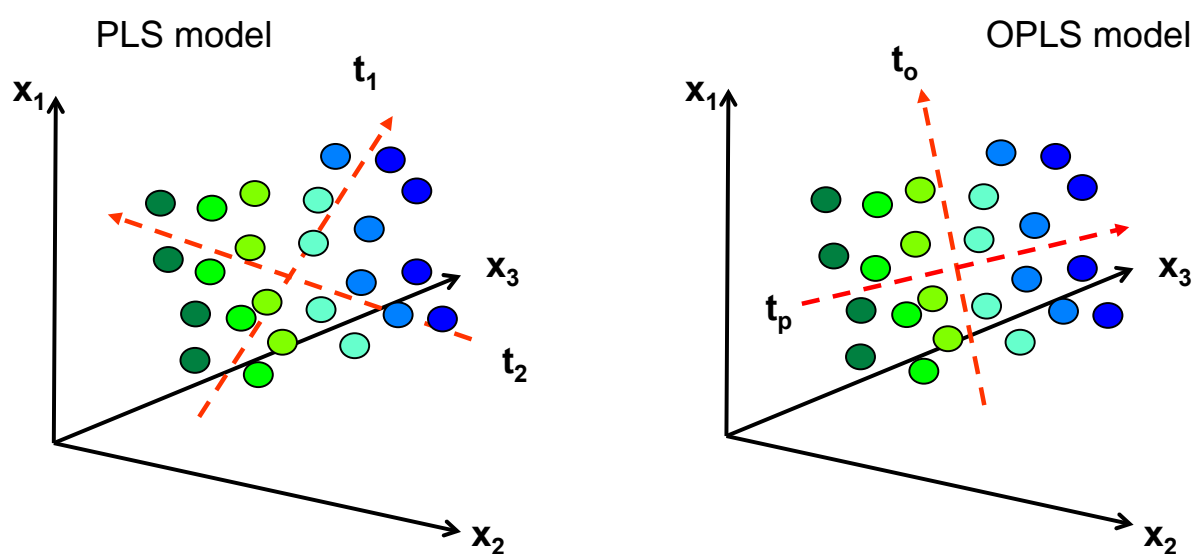


Fig. 4-1. The difference between PLS and OPLS, the model rotation allows (order) separation between different colored circles (between class) by the predictive component t_p and same colored circles (within class) by the orthogonal component t_o .

An effective and reliable method to differentiate civet coffee from regular coffee has been described previously in Chapter 1 based on the identification and quantification of biochemical markers through metabolomics. Furthermore, metabolomics analysis to detect and quantify coffee adulteration, *i.e.*, Arabica-Robusta fractions in coffee blends, have been stated employing a wide range of analytical instruments^{46,55,108,109}. In this chapter, evaluation of the use of gas chromatography – mass spectrometry (GC/MS) combined with orthogonal projection to latent structures (OPLS) regression analysis as a reliable method to quantify the degree of civet coffee adulteration on the basis of the mixing ratio, as well as to characterize the major compounds responsible for samples distinction, was conducted. A set of coffee

blends of civet coffee and cheap coffee with known-ratio was used to construct a prediction model and subsequently evaluated to predict the known–unknown samples.

As described in Chapter 2, OPLS splits the variations in X into two directions, one that is linearly related and another that is orthogonal to Y ^{66,110}. Moreover, the model is rotated so that variation that correlated with class separation is found in the predictive component, t_p , and variation not related to class separation, or uncorrelated variation, is located in orthogonal components, t_o , providing maximal connection to Y and allows easier, straightforward model interpretation (**Fig. 4-1**). Additional information about steps to construct OPLS and OPLS-DA can be found in **Supplementary Principle 1**.

4.2. Experimental section

4.2.1. Coffee bean materials

Two sets of coffee blends from certified (authentic) and commercial coffee were prepared to construct the prediction model. The certified samples of civet coffee, regular coffee and non-animal treated coffee were provided by the Indonesian Coffee and Cocoa Research Institute (ICCRI). Civet coffee and regular coffee from Sidikalang, Indonesia, which have been well known for its scarcity and unique aroma¹¹¹, were acquired commercially. The samples were ground and mixed in different mixing proportions to obtain representative blends. Ultimately, samples with 1 g of pure civet coffee, pure regular coffee, and blends of 10, 20, 30, 40, 50, 60, 70, 80, and 90% civet coffee were made. These groups of samples were referred to as training or experimental set.

For validation, two sets of coffee blends were made from authentic and commercial civet coffees with unknown regular coffees (acquired commercially), with 25 and 75% mixing ratio. The validation set was analyzed in a separate week to investigate the influence of expected variance resulting from day-to-day measurements. This sample set was then projected to the model to verify its performance for prediction of mixing ratio of unknown samples. The list of samples is described in **Supplementary Table 4-1**.

4.2.2. Sample preparation, chromatographic and quantitative analyses with GC/MS

Coffee bean samples were subjected to metabolite extraction using mixture of polar and non-polar solvent and chemical derivatization utilizing same procedures in Chapter 2 section 2.2.2.

Analysis of coffee bean extract was performed on a GCMS-QP 2010 Ultra (Shimadzu, Kyoto, Japan) fitted with a 30 m × 0.25 mm i.d. fused silica capillary column coated with 0.25 µm low bleed, CP SIL 8 CB column (Varian Inc., Palo Alto, CA, USA). The instrument was installed with an AOC-20i/s autosampler (Shimadzu). The injection, transfer line and ion source temperature were 230, 250 and 200°C, respectively. The helium gas flow rate through column was 1.12 mL/min with a linear velocity of 39 cm/s. The temperature programs of GC were as follows: held at 80°C for 2 min isothermally and then increased at 15°C/min to 330°C and held for 6 min. Ions were generated by a 0.93 kV and were recorded over the mass range m/z 85 – 500. Prior to analysis, tuning and calibration of mass spectrometry were carried out to confirm its stability. One microliter of sample was injected in split mode (25:1, v/v). The coffee samples and blank (only extraction solvent) were analyzed in a randomized order. To calculate retention index for tentative identification, standard alkane mixtures were injected at the beginning and end of the analysis. Concentration of selected significant metabolites in coffee blends was calculated using the linear calibration curve between concentration of chemical standards and peak area. The concentration of chemical standards was adjusted to 1, 10, 50, 100, 250, 500, 750 and 1000 µM with the extraction solvent. Construction of calibration curve was performed with co-injection of respective chemical standards at various concentrations and comparing their retention time and mass fragment with databases. Details of quantitative analysis were explained in Chapter 2 section 2.2.3.

4.2.3. Data handling and statistical analysis

To provide compatibility between mass spectral data over various data processing software packages, raw chromatographic data were converted into ANDI files (Analytical Data Interchange Protocol, *.cdf) using the GC-MS Solution software from Shimadzu. The converted total ion chromatograms were imported to freely available software package,

MetAlign⁷⁰, for peak detection and multiple alignments of the retention times, using optimized parameters described in section 2.2.4.

Tentative identification by comparing the mass spectral fragmentation pattern and retention indices with in-house spectral database was carried out. Retention indices were calculated based on retention times of standard alkane mixture. The identification was accelerated with the use of the annotation software, AIoutput2⁷². This software package was merged with the in-house mass spectral database, which was developed from authentic chemical standards. For verification of tentatively identified compounds and unknown peaks, the database provided by NIST was also used.

The result outcome (a 2-D data matrix of comma-separated values (.csv) files in which each row and column denotes for sample identity and relative peak intensity at certain retention time, respectively) was normalized by manually adjusting the peak intensity to the ribitol internal standard. The RSD threshold (< 20%) was applied for peak intensity across four measurement replicates to filter out measurement-to-measurement variation. Subsequently, the data matrix was then subjected to SIMCA-P version 13 (Umetrics, Umea, Sweden) for multivariate analysis. Principal component analysis (PCA) was initially performed in a non-biased, unsupervised way, to visualize the differences in the metabolite profiles corresponding to each mixing ratio. Orthogonal projection to latent structures, OPLS, a modified PLS regression technique, was then chosen to create a prediction model between two sets of variables namely: 1) responses or prediction (*X*) corresponding to the metabolite profiles from GC/MS data and 2) observation (*Y*), corresponding to the mixing ratio of coffee blends.

4.3. Results and discussion

4.3.1. Chromatographic profiles of coffee bean extract

Due to its remarkable sensitivity and resourcefulness to rapidly elucidate chemical structures, the technique using a combination of gas chromatography and mass spectrometry (GC/MS) remains to be one of the most widely used techniques for plant metabolomics^{34,112–114}. Thus, this technique was used on the coffee bean extract samples to construct blend ratio-prediction model for authentication. **Fig. 4-2** illustrated the total ion chromatogram of the pure and blended samples of coffee bean extract. Both pure and blended extracts revealed similar chromatographic patterns over 24 minutes of analysis. However, substantial and characteristic differences in the relative intensity among three extracts can be observed visually. Relative intensity of peaks in the blended sample represented the mixing ratio of 50:50 between two pure samples (**Fig. 4-2B**). Therefore, chromatographic profiles of coffee beans play an essential role for sample discrimination.

Mass spectrometry-based analysis produces enormous data, featuring essential information from large number of data points and experimental noise. A data processing software was utilized to extract information out of the complex matrix as well as to transform it into a more comparable form. The software package MetAlign, with an optimized parameter for coffee metabolomics, was chosen for processing of the GC chromatographic data since it provides options for baseline subtraction, peak distinction from noises and retention time alignment over multiple samples. To generate comparable analysis for both experimental and validation data sets, 72 unique coffee peaks were extracted out of a convoluted data corroborated with blank sample, with 24 peaks were tentatively identified from spectral similarity with at least two compound spectra databases and 11 peaks were identified after confirmed by comparison with corresponding chemical standards (**Supplementary Table 2-2**).

Diverse chemical constituents, comprised of a large number of low molecular weight acids including quinic acid, chlorogenic acid (the major acid in coffee⁷⁶), sugars, amino acids, health benefits or risks-related compounds such as trigonelline and caffeine^{115–118}, were tentatively identified from the chromatogram of the coffee extract. Then, authentic chemical standards were co-injected to give a valid identification of important metabolites.

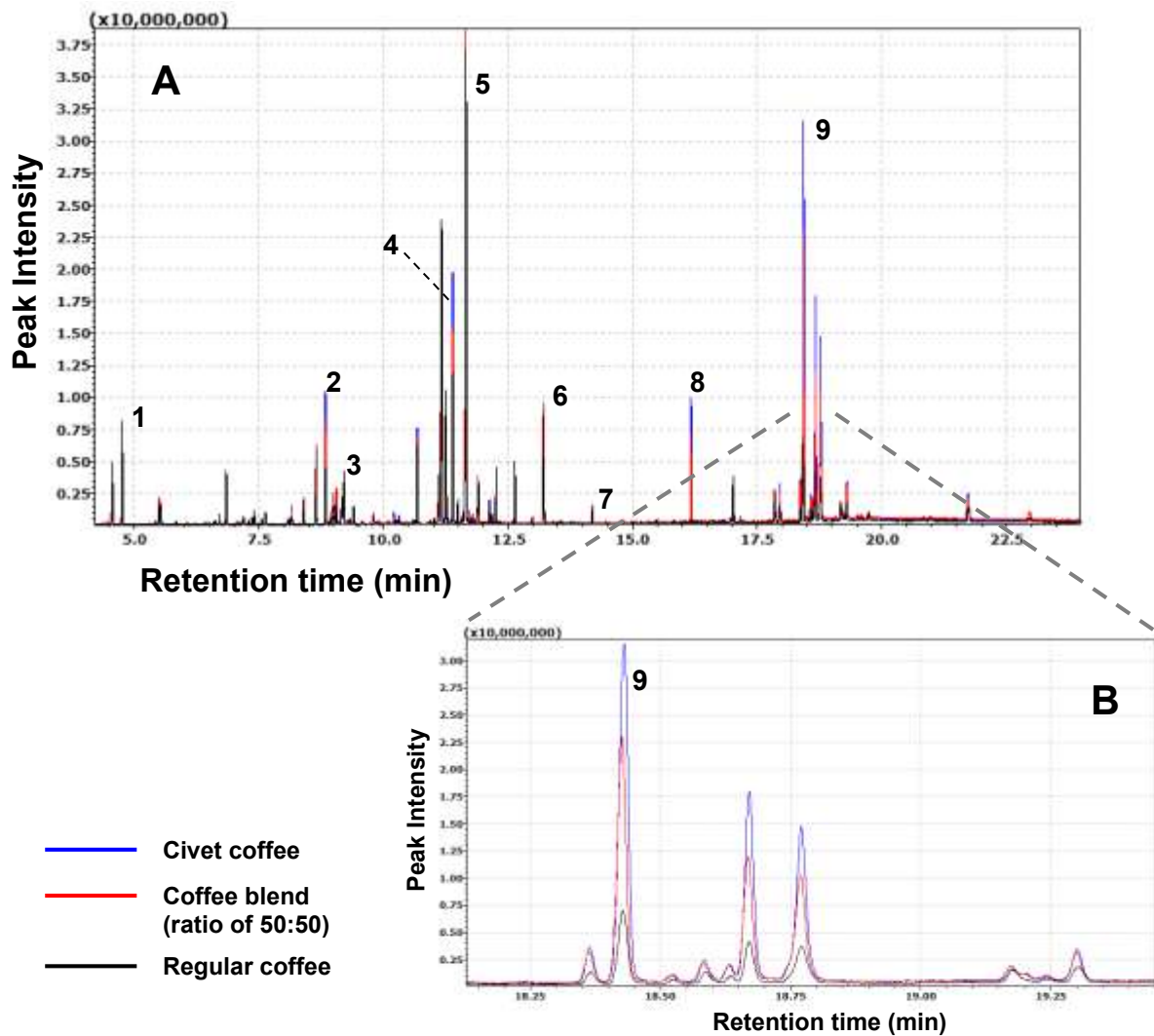


Fig. 4-2. Total ion chromatograms of the pure and coffee blend samples with 50:50 ratio (A). Inset (B) showed substantial differences in the peak intensity among pure and coffee blends. Representative peak annotations: (1) glycolic acid, (2) malic acid, (3) pyroglutamic acid, (4) citric acid, (5) quinic acid, (6) inositol, (7) caffeic acid, (8) sucrose, (9) chlorogenic acid.

4.3.1. Metabolite profiles of coffee bean extract

4.3.1.1. Chromatographic profile of coffee blends

Unsupervised PCA was employed to emphasize the overview of all data trends and their relationship with different mixing ratio. Total data set was preprocessed by autoscaling or unit variance (UV): $\tilde{x}_{ij} = (x_{ij} - \bar{x}_i)/s_i$, where \tilde{x}_{ij} is the value of the variable i relative to the sample j , \bar{x}_i is the mean of variable i , and s_i is the standard deviation of the variable i ^{107,119}.

Fig. 4-3A – 4-3B depicted the PCA score plots of authentic and commercial coffee sets. A well-defined sample separation in conformity with low- and high-mixing percentage was achieved, explaining 65.5 and 53.2% variance of PC1, respectively. The blends containing more than 50% civet coffee contents occupied the left region of PC1, samples with 50% ratio were grouped around the intercept (0,0) while the lower mixing percentage blends were clustered in the right region. The score plots of authentic and commercial coffee had R^2X (goodness-of-fit) values of 0.769 and 0.753, respectively. For both data sets, almost all data were plotted inside 95% hotelling- T^2 (except for one outlier) indicating the reliability of metabolite measurement among sample replicates. The results suggested that metabolite profiles play a significant role in sample differentiation through a variety of mixing ratio.

Additionally, metabolites that contribute to the separation of low- and high-mixing ratio coffee blends were extracted from corresponding loading plots (**Supplementary Fig. 4-1**). Several metabolites, such as chlorogenic acid, citric acid, sucrose, quinic acid, and caffeic acid showed strong significance on the clear separation of the two major groups of coffee blends along PC1. Particularly, the first three metabolites strongly contributed to the high-mixing ratio coffee blends, whilst quinic acid and caffeic acid gave a low impact.

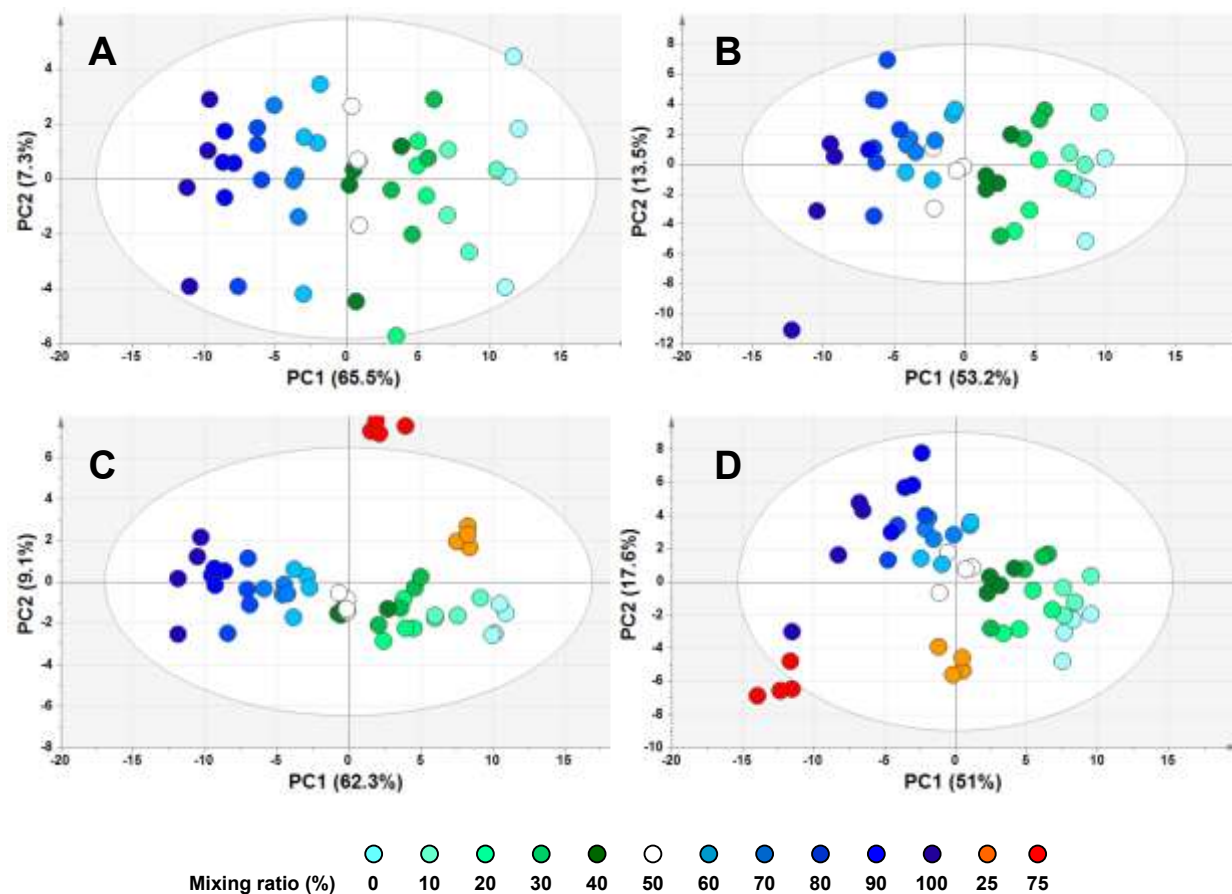


Fig. 4-3. PCA displayed well-defined samples separation according to low- and high-mixing ratio in (A) authentic and (B) commercial coffee sets. High-mixing ratio samples were clustered along the negative axis of PC1, while the lower mixing ratio blends were grouped at the positive axis. External validation sets (coffee with 25 and 75% of civet coffee content) were unevenly distributed due to day-to-day variation (C, D).

Next, incorporation of the external validation data into the experimental data set to obtain the total data set was conducted. Prior to multivariate analysis, the total data was normalized to the ribitol internal standard and scaled to unit variance. The PCA score plots are shown in **Fig. 4-3C – 4-3D**. It is obvious that in the external validation set, coffee blends with 25 and 75% civet coffee contents were segregated from experimental data along PC2; The PC1 axis represented blend percentage. This variation could arise from many causes such as signal-time drift due to instrumental fluctuation, uncontrollable variation in coffee processing and other factors. Ultimately, such unwanted variation can hinder the interpretation of biologically important phenomena from data set^{95,120,121}.

4.3.1.2. Subset-wise scaling to minimize run-to-run variation

In order to extract biological information, the unwanted variation mentioned above must be adequately dealt in an appropriate way. Many efforts have been made over the years for managing data preprocessing and normalization of multivariate data in metabolomics^{95,121,122}. In addition to unit variance, data preprocessing with Pareto; $\tilde{x}_{ij} = (x_{ij} - \bar{x}_i)/\sqrt{s_i}$, Centering; $\tilde{x}_{ij} = x_{ij} - \bar{x}_i$, and combination with log transformation have been done, but none was effective in removing such variation.

Thus, preprocessing of each matrix subset before data integration was performed using subset-wise scaling. The concept is adopted from blockwise scaling¹²³, which allows each matrix subset to be thought as a unit and to be given the appropriate variance. Matrix subset from each data set, experimental and validation, were normalized to the internal standard and UV scaled separately before integration (**Fig. 4-4**). According to Eriksson *et al*, Pareto, UV and no scaling may be applied as the basis scaling method in this particular data pretreatment¹²³.

The method effectively removed biological and experimental variabilities (day-to-day measurement, instrumental drift, etc.) in the experimental (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% civet coffee) and validation data set (25 and 75% civet coffee) as shown in **Fig. 4-5**. The given variance was considered smaller than the value obtained if the total subset was scaled thoroughly. Validation subsets fitted onto the former data set; the separation can be seen along the PC1 axis in accordance to the percentage of coffee

blends. No significant changes were found in the data structure of important metabolites displayed in loading plot (**Supplementary Fig. 4-2**). A better data structure and good data fitness were constructed when using unit variance as the basis scaling method. Unit variance gives all metabolites an equal contribution of influencing the data analysis regardless of their discrepancies, *i.e.*, two independent data sets, and therefore it has been proven to minimize between-batches variation¹²².

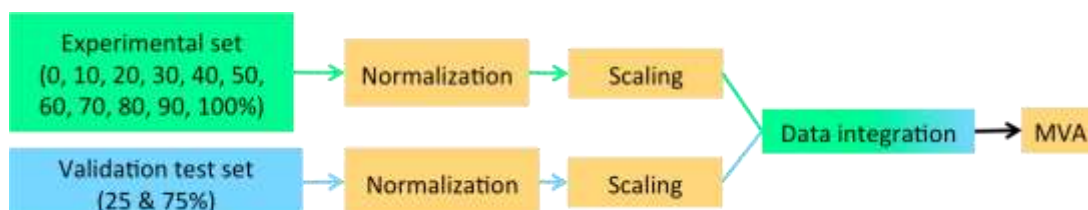


Fig. 4-4. Scheme of subset-wise scaling

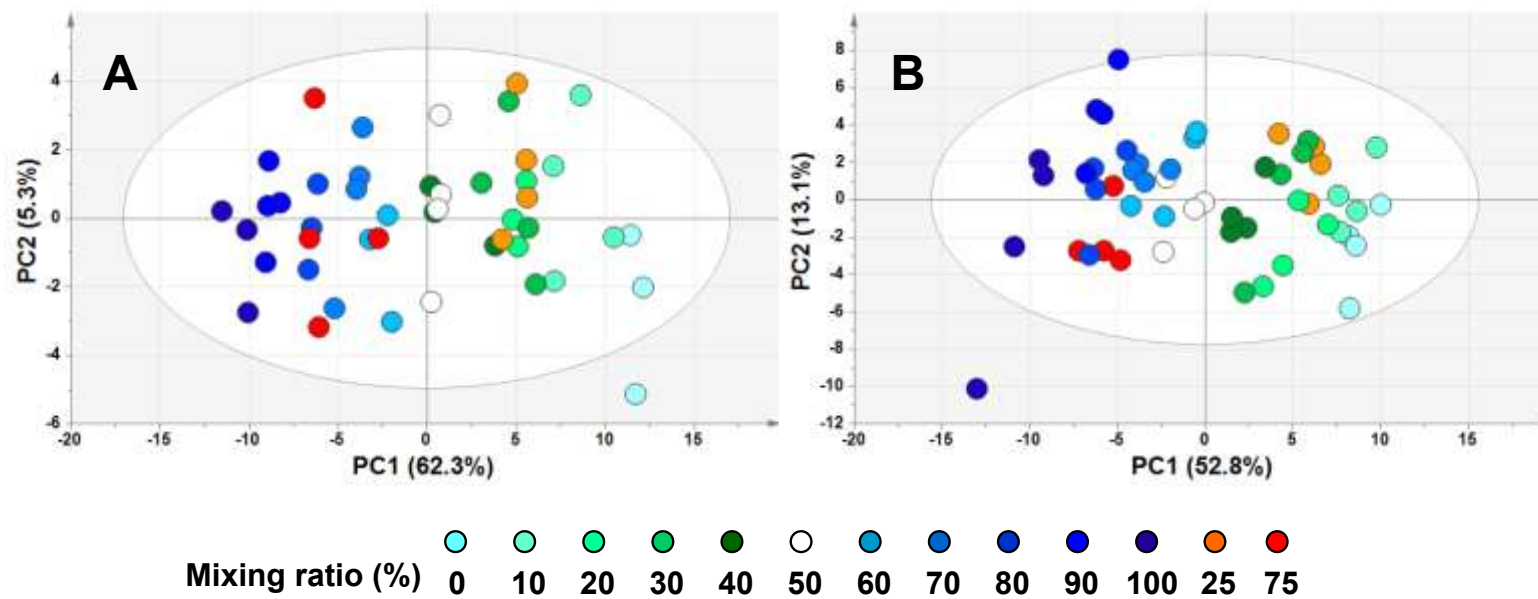


Fig. 4-5. PCA score plots of authentic (A) and commercial (B) coffee sets after pretreatment with subset-wise scaling. The data pretreatment technique effectively removed experimental variability among experimental and validation coffee sets.

4.3.2. Mixing ratio-prediction model with OPLA

4.3.2.1. Construction of OPLS model

In PCA results, it was obvious that coffee blends were clustered into two groups on the basis of low- and high-mixing ratio. Nonetheless, overlapping data points were also observed among samples in each group. Therefore, it is important to create a quality-prediction model by means of OPLS regression to determine whether the ratio of each coffee blend can be associated with their metabolite profiles. In contrast to PCA, a priori information is a prerequisite in OPLS to redirect the analysis towards the proposed objectives^{66,67}. In this study, priori information was the mixing ratio of coffee blends, represented as *Y* variable, whereas *X* was calculated from the relative peak intensity of each metabolite.

The entire dataset from each coffee blend was divided into two parts: eleven training set samples (coffee blends with the following civet coffee contents: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100%) and test set from validation set, coffee blends with 25% and 75% civet coffee contents. The test set was then excluded for model validation. The OPLS regression was built with the remaining data of training set and then validated with projection of the left out test set onto the constructed model. To compare the consequences of different scaling methods for minimizing data variability, both conventional and subset-wise scaling were applied.

Initially, matrix subset of experimental and external validation coffees were integrated to obtain the total data set. The total data set was then scaled to unit variance. The model resulted in R^2Y and Q^2 values of 0.975 and 0.974 for authentic coffee and 0.987 and 0.982 for commercial coffee set. The model predictability to estimate the mixing ratio of new coffee blend samples was tested by projecting the test set into prediction model (**Fig. 4-6A – 4-6D**). Both OPLS models showed overfitting and poor predictive ability, implied by the large validation error (RMSEP = 31.8 and 25.9 for authentic and commercial coffee, respectively) over the model estimations on the basis of training set (RMSEE = 5.1 and 3.77). The prediction model suffered from variation in the samples' properties, processing and inter-day measurement.

In the past, uncorrelated metabolite data of X variables interfering in the prediction of mixing ratio (Y variables) often cause imprecise predictions. A spectral filtering technique, orthogonal signal correction (OSC) was used to enhance the quality of PLS prediction that is associated with this interference^{34,63,124}. Nevertheless, OSC filter is prone to overfitting and may give results that are too optimistic⁶⁷. Its application are then progressively replaced concomitantly with the development of a generic preprocessing technique, OPLS/O2PLS⁶⁷ that is more robust to simplify the complexity of systematic variation from X . By employing the OPLS technique, the quality of prediction model was still poor. Therefore in this study, the interference of uncorrelated variables is not a relevant factor affecting the robustness of the prediction model. Other possible causes are experimental and biological variations. Subset-wise scaling was then applied to preprocess data subset for prediction of mixing ratio of authentic and commercial coffee sets. Data pretreatment methods can correct the aspects, *i.e.*, samples' variability, that distort the predictability of the model thus hampering the biological interpretation¹¹⁹. By using subset wise-scaling technique, the training set fit the OPLS model resulting in a linear regression model (**Fig. 4-6E – 4-6F**). Interestingly, whilst the RMSEP value significantly improved from 31.8 to 5.81 and 25.9 to 5.52 for former and latter data sets, the R^2Y and Q^2 values were unaffected. The unchanged R^2Y , Q^2 and the decrease in RMSEP value suggested that the scaling method improved the model estimation of training set by removing unwanted variations while maintaining the fitness and predictability power of the model. The details of OPLS features are described in **Table 4-1**. I used this prediction model to estimate the composition of the external validation samples (%), and the values obtained were 27.4 ± 3 and 72.6 ± 8.1 for civet coffee content in authentic coffee set while that from commercial coffee were 26.5 ± 7 and 73.5 ± 5.2 , respectively. **Table 4-2** summarizes the observed and predicted value of the mixing ratio of all coffee samples. The prediction models presented rather high error particularly for lower-mixing ratio samples but the predicted values remained relevant considering the variability of samples' properties. Generally, prediction model built from commercial coffee exhibited better precision than those from the authentic coffee set.

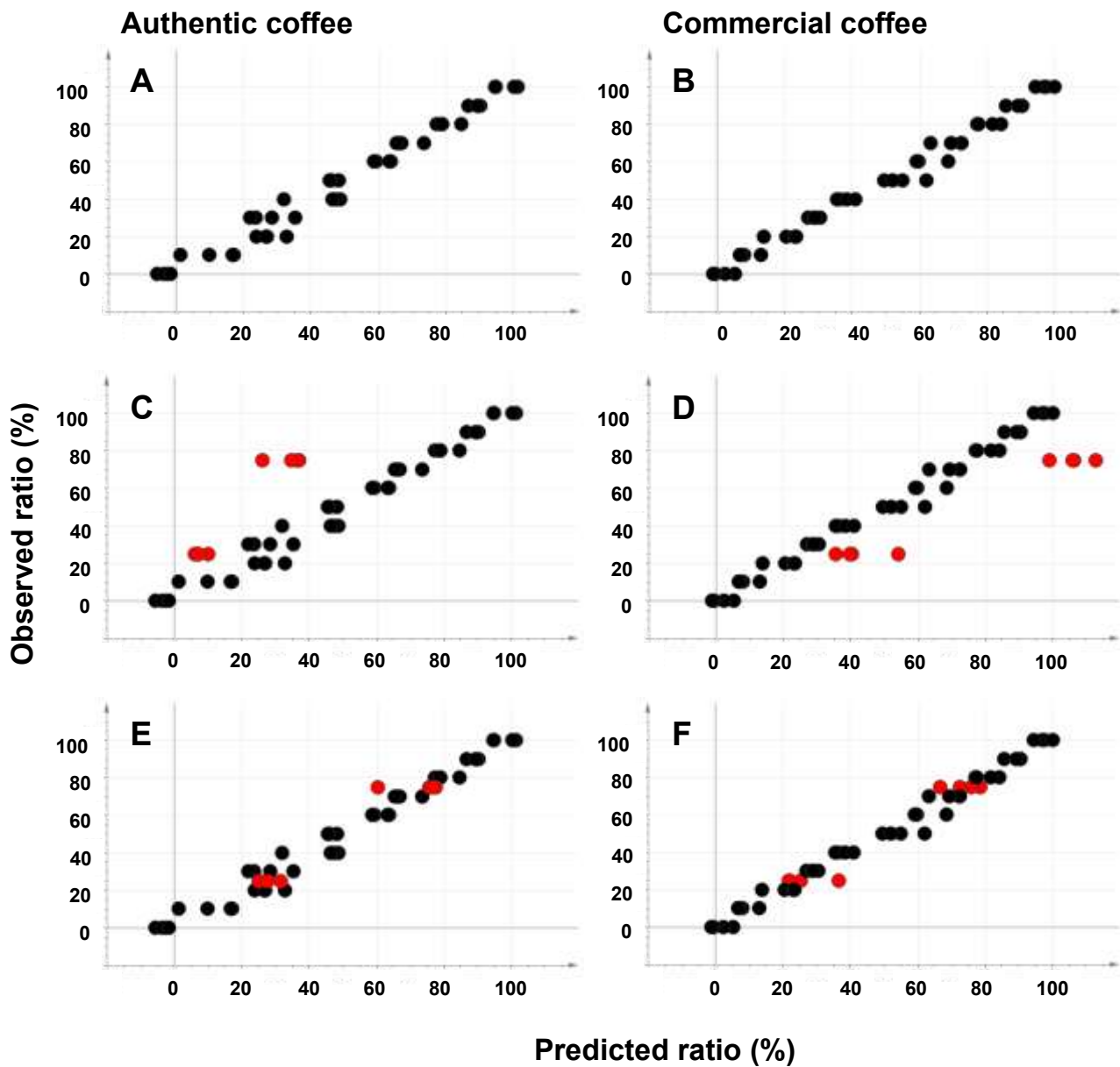


Fig. 4-6. Prediction models were built using 11 sets of experimental coffee blend as training set (A, B) and verified with projection of validation coffee set as testing set, coffee blend with 25 and 75% civet coffee (red) (C, D). OPLS models showed overfitting and poor predictive ability. By employing subset-wise scaling technique, the testing set fit the OPLS model (E, F) for both coffee sets.

Conventionally, variables that are highly relevant for explaining the mixing ratio of coffees can be extracted from VIP (variable importance in the projection) values. VIP values greater than 1 are considered to be the most relevant. However, due to pretreatment using unit variance, all metabolites become equally important as shown by comparable VIP values in **Fig. 4-7**. In addition to the VIP value, there are many ways to interpret relevance of OPLS model, such as PLS weights and regression coefficients. Here, I have focused on the regression coefficients to examine the model variables. Whereas VIP is a weighted sum of squares of the OPLS weights (strengths), regression coefficients represent the strength as well as direction of correlation between *X* and *Y* variables¹²⁵. In this study, positive regression coefficients implied high correlation to civet coffee properties. Taken from VIP and regression coefficients values, organic acids and sugars showed high association with civet coffee. Malic acid and citric acid, the selected discriminant markers of civet coffee, were among the relevant metabolites in both authentic and commercial coffee sets.

Table 4-1. Features of OPLS prediction models

Coffee set	Scaling	OPLS features			
		R^2Y	Q^2	RMSEE	RMSEP
Authentic	Conventional	0.975	0.974	5.11	31.8
	Subset-wise	0.975	0.974	5.11	5.81
Commercial	Conventional	0.987	0.982	3.77	25.9
	Subset-wise	0.987	0.982	3.77	5.52

Table 4-2. The observed and predicted mixing ratio from OPLS models

Observed ratio (%)	Predicted ratio (%) ^a	
	Authentic coffee	Commercial coffee
0	-3.2±1.7	1.5±3
10	11.4±7.4	8.7±2.9
20	27.6±3.7	20.3±4.5
25	27.4±3	26.5±7
30	27.5±5.9	28.9±1.5
40	43.7±7.8	37.9±2.4
50	46.5±1.2	54.7±5.4
60	61.4±2.6	61.9±4.5
70	68.1±3.7	69.5±4.3
75	72.6±8.1	73.5±5.2
80	80±3.2	80.2±3.4
90	89.1±1.6	89±2.3
100	97.9±1.7	97.4±2.3

^amean±SD, n=4

4.3.3. Quantification of important metabolites

To further confirm the OPLS prediction, quantification of important metabolites for prediction of coffee blends was carried out. The calibration curves for all authentic standards displayed remarkable linearity (0.99 or more) and repeatability for the applied concentration range (**Supplementary Table 4-2**). The concentrations of important metabolites were determined to be higher than the LOD (limit of detection) and LOQ (limit of quantitation), therefore demonstrating a valid quantification. The important metabolites exhibited concentrations that were proportionally elevated with the increase of civet coffee contents (**Fig. 4-8**). However, only citric acid showed a reasonable concentration level of external validation coffee set (**Supplementary Fig. 4-3**). It was revealed in Chapter 2 and 3 that citric acid showed the strongest correlation to civet coffee among other candidates. I have repeated the quantification using a new calibration set to explore the involvement of instrumental drift in which the result obtained was consistent with the above finding (**Supplementary Fig. 4-4**). Particularly for high-mixing ratio, the value of citric acid concentration reached a plateau. At certain concentration, the coffee signals may overreach the dynamic range of the instrument, become saturated and subsequently generate the plateau effect. The results suggested the necessity to perform validation, *i.e.*, quantification, after multivariate analysis. Data pretreatment indisputably gives influence to each constituent in the matrix subset thus their signal intensities may not represent the actual values. Picking the best data pretreatment technique is an important ‘key’ step in metabolomics¹²⁶. A particular data pretreatment technique is not always the most suitable for analysis and it may depend on the data structure.

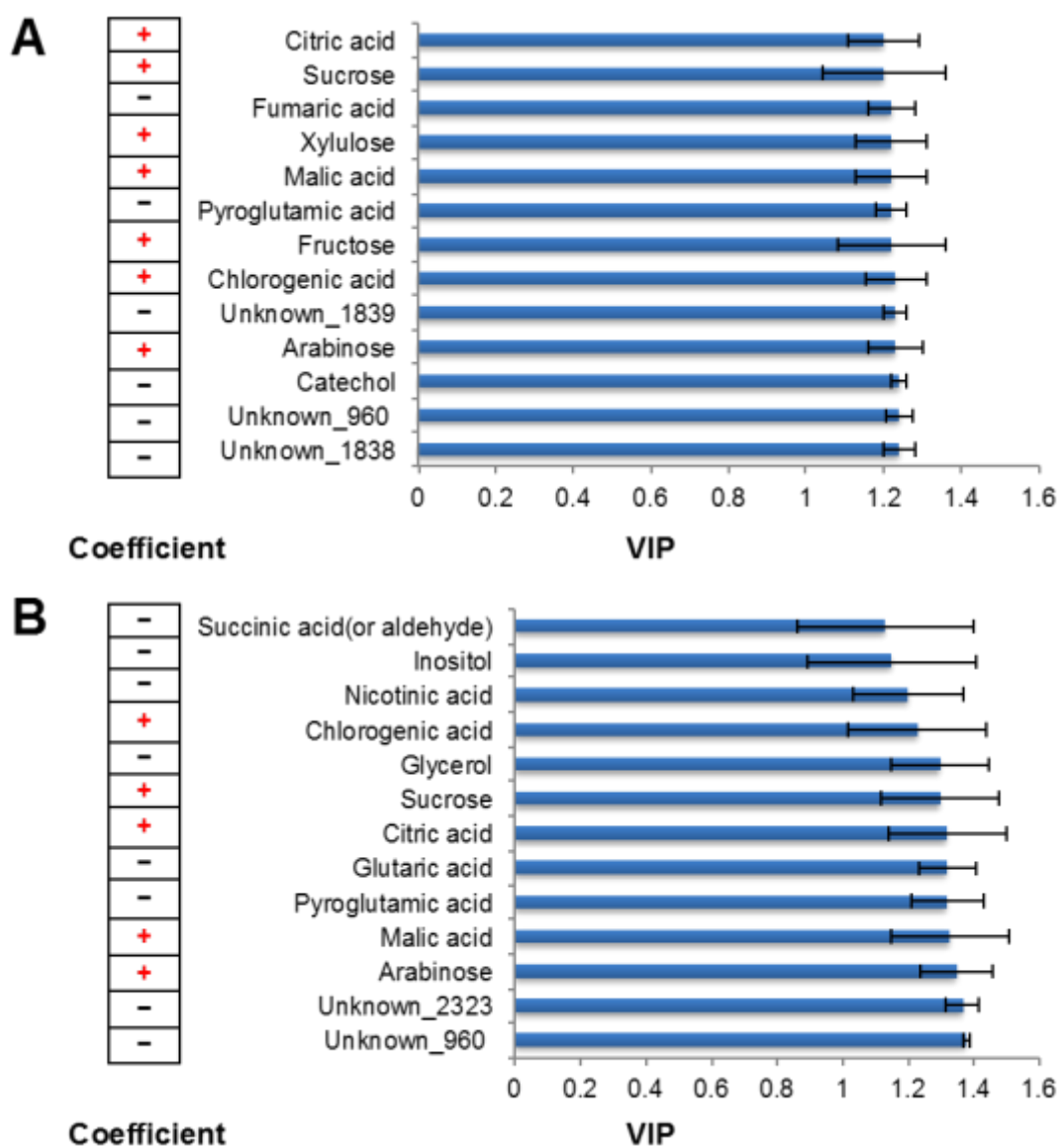


Fig. 4-7. VIP (variable importance in the projection) and coefficient values extracted from the best OPLS prediction model. (A) authentic, (B) commercial coffee.

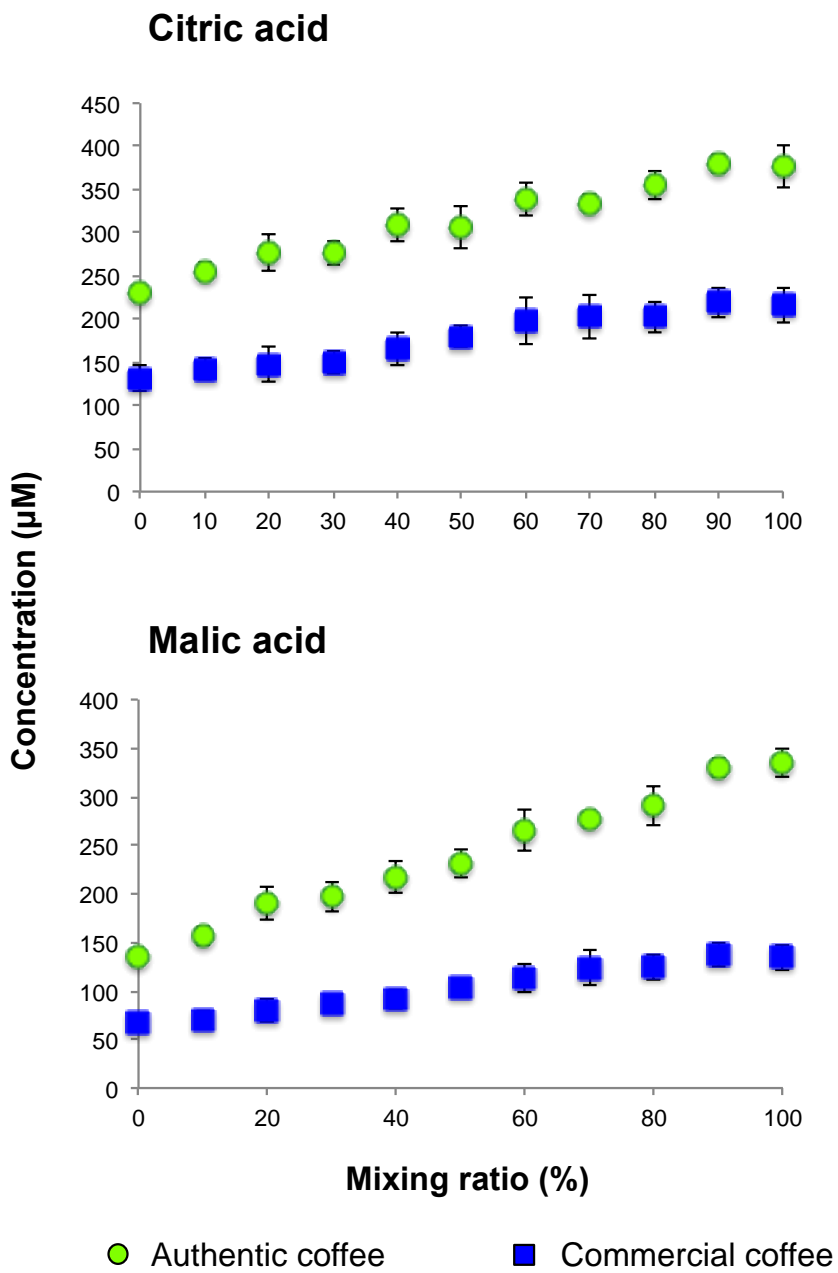


Fig. 4-8. Concentrations of citric acid and malic acid were evenly corresponded to the mixing ratio across various coffee blend samples.

4.4. Conclusions

The work discussed in this chapter was conducted to develop a prediction method and evaluate its accuracy for civet coffee quality assessment. The OPLS model was built using a limited number and variety of coffee samples with known blend ratio and may not address the influence of variability of coffee species, origins, and processing. However, the prediction model constructed from metabolite profiles of different coffee blends combinations, authentic and commercial coffees, resulted in a linear regression, indicating a proof-of-concept of the reliability of the GC/MS-based metabolomics approach to feasibly quantify the degree of coffee adulteration. Additionally, the result demonstrated robust application of data pretreatment in multivariate prediction technique for accurate estimation of coffee fraction in blends and its relevance to counteract adulteration of specialty coffee.

Chapter 5 Conclusions and perspectives

The capacity of metabolomics approach to serve as descriptor of phenotype, that is sensitively affected by perturbations or stresses, favor its practicability for detection of changes in the food and crop components resulted from fraudulent adulteration. Screening of specific marker in association with fraud and comparative component analysis to classify authentic and fraud products are among the strategies for detection of adulteration assisted by metabolomics. In this thesis, the use of metabolomics technology was demonstrated to select discriminant markers and its applicability for rapid authentication screening of civet coffee.

Civet coffee is among human foods produced by its passage through the digestive tract of an animal. Despite of its natural perturbation, civet coffee is also subjected to various human perturbations. At first, screening and selection of discriminant marker that associated with digestion of civet were demonstrated through metabolite profiling employing common instrument, GC/MS. The increase of acidity principles and reduced level of bitterness- and astringent-related compounds, such as caffeine and quinic acid, were observed in the metabolite profile of civet coffee, consistently in agreement with previous report concerning chemical changes occurred to coffee beans after animal digestion. I then tested the applicability of the discriminant marker candidates to verify authenticity of commercial coffee products. Citric acid, malic acid and ratio of inositol/pyroglutamic acid exhibited potential application in the differentiation authentic, fake civet coffee, regular coffee and coffee blend samples with 5 wt % civet coffee content.

Further application studies were presented in the second part. Firstly, attempt to develop rapid and reliable protocol for authenticity screening in industry was carried out. To accomplish, I applied metabolite fingerprinting utilizing GC coupled with universal and cost-effective detector, FID (flame ionization detector). In contrary to metabolite profiling, authenticity evaluation via metabolite fingerprinting can be done without metabolite details. GC/FID analysis provided higher sensitivity for almost all detected peaks and could successfully reproduce quality of prediction from GC/MS for discrimination of various commercial coffee products.

In the second part of application study, I developed a proof-of-concept of metabolomics technology for determining degree of coffee adulteration on the basis of mixing ratio from 10 to 90%. Through OPLS regression analysis, I made prediction model to quantify mixing ratio of coffee blends of civet and regular coffee. The model presented accurate ratio estimation of known-unknown samples according to citric and malic acid. To the best of my knowledge, this is the first study to address the screening, identification and successful validation of discriminant markers for the authentication and differentiation of civet coffee, regular coffee, coffee blends and fake coffee, that uncovered by other previous techniques, and for routine application in industry. Identification of discriminant markers allows method standardization, *i.e.*, determination of the threshold level of discriminant markers, for future practical purposes.

Presently, the quality of several valuable crop products such as coffee and tea have traditionally assessed on the basis of human sensory perception (sight, smell, taste and touch). However, this method tended to be highly subjective with only up to 20% precision. Due to the need of long-termed training, it is inevitably expensive to hire highly trained assessor. Moreover, trained specialist for civet coffee evaluation is still a few (personal communication with ICCRI). Because civet coffee is a crop product used for human consumption, option to do sensory evaluation cannot be fully omitted. Sensory evaluation is useful to broaden knowledge concerning discriminant markers and its relation with aroma and flavor. Yet, the selected discriminant markers chosen from this study are expected to perform as sole markers or in combination with sensory test.

Practical application of this study in industry can be performed by combining knowledge of discriminant markers and sensory profile. At first, rapid test using GC/FID to measure level of the desired markers is expedient to significantly reduce effort and time of analysis for screening of large number of coffee samples. Subsequently, assessment of the short list coffee samples which shown characteristics of civet coffee, *i.e.*, certain level of markers, by sensory specialists will improve the result's confidence regarding authenticity of the samples. Since the taste of food is likely defined from complex interactions of many compounds²⁰⁻²¹, the use of sensory evaluation is also beneficial to counteract illicit attempt of

putting chemical substances in order to enhance the level of discriminant markers in the coffee samples.

However, due to the skill specificity, coffee sensory specialist may not be available in each coffee factories or research institutes. In this instance, the selected discriminant markers can be used solely. To prevent unlawful mixing with chemical substances, analysis in this platform is highly recommended to include green beans as a control, in addition to roasted coffees that are being subject of evaluation. After roasting of green beans, both control and tested coffees are analyzed with rapid GC/FID test. Authentication is then confirmed by discriminant analysis employing statistical model or by merely quantifying the level of the markers among those samples. Furthermore, because the discriminant markers chosen in this study are mainly organic acids, the established authenticity evaluation is projected to be widely applicable for other conventional analytical techniques to measure organic acids, *i.e.*, HPLC.

References

- (1) Fiehn, O. Metabolomics--the Link between Genotypes and Phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155–171.
- (2) Putri, S. P. & Fukusaki, E. Metabolomics In A Nutshell. In *Mass Spectrometry-Based Metabolomics*. **2014**. Florida, USA: CRC Press.
- (3) Fiehn, O.; Robertson, D.; Griffin, J.; vab der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L. W.; Goodacre, R.; Hardy, N. W.; Taylor, C.; et al. The Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 175–178.
- (4) Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E. Metabolomic Analysis in Food Science: A Review. *Trends Food Sci. Technol.* **2009**, *20*, 557–566.
- (5) Defernez, M.; Gunning, Y. M.; Parr, A. J.; Shepherd, L. V. T.; Davies, H. V.; Colquhoun, I. J. NMR and HPLC-UV Profiling of Potatoes with Genetic Modifications to Metabolic Pathways. *J. Agric. Food Chem.* **2004**, *52*, 6075–6085.
- (6) Pongsuwan, W.; Bamba, T.; Harada, K.; Yonetani, T.; Kobayashi, A.; Fukusaki, E. High-Throughput Technique for Comprehensive Analysis of Japanese Green Tea Quality Assessment Using Ultra-Performance Liquid Chromatography with Time-of-Flight. *J. Agric. Food Chem.* **2008**, *56*, 10705–10708.
- (7) Beckmann, M.; Enot, D. P.; Overy, D. P.; Draper, J. Representation, Comparison, and Interpretation of Metabolome Fingerprint Data for Total Composition Analysis and Quality Trait Investigation in Potato Cultivars. *J. Agric. Food Chem.* **2007**, *55*, 3444–3451.
- (8) Levandi, T.; Leon, C.; Kaljurand, M.; Garcia-Cañas, V.; Cifuentes, A. Capillary Electrophoresis Time-of-Flight Mass Spectrometry for Comparative Metabolomics of Transgenic versus Conventional Maize. *Anal. Chem.* **2008**, *80*, 6329–6335.

- (9) Son, H. S.; Ki, M. K.; Van Den Berg, F.; Hwang, G. S.; Park, W. M.; Lee, C. H.; Hong, Y. S. ¹H Nuclear Magnetic Resonance-Based Metabolomic Characterization of Wines by Grape Varieties and Production Areas. *J. Agric. Food Chem.* **2008**, *56*, 8007–8016.
- (10) Dixon, R. A.; Strack, D. Phytochemistry Meets Genome Analysis, and Beyond..... *Phytochemistry* **2003**, *62*, 815–816.
- (11) Putri, S. P.; Yamamoto, S.; Tsugawa, H.; Fukusaki, E. Current Metabolomics: Technological Advances. *J. Biosci. Bioeng.* **2013**, *116*, 9–16.
- (12) Blekherman, G.; Laubenbacher, R.; Cortes, D. F.; Mendes, P.; Torti, F. M.; Akman, S.; Torti, S. V.; Shulaev, V. Bioinformatics Tools for Cancer Metabolomics. *Metabolomics* **2011**, *7*, 329–343.
- (13) Ogura, T.; Bamba, T.; Fukusaki, E. Development of a Practical Metabolite Identification Technique for Non-Targeted Metabolomics. *J. Chromatogr. A* **2013**, *1301*, 73–79.
- (14) Nakayama, Y.; Tamada, Y.; Tsugawa, H.; Bamba, T.; Fukusaki, E. Novel Strategy for Non-Targeted Isotope-Assisted Metabolomics by Means of Metabolic Turnover and Multivariate Analysis. *Metabolites* **2014**, *4*, 722–739.
- (15) Nakayama, Y.; Putri, S. P.; Bamba, T.; Fukusaki, E. Metabolic Distance Estimation Based on Principle Component Analysis of Metabolic Turnover. *J. Biosci. Bioeng.* **2014**, *118*, 350–355.
- (16) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Gui, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35*, 521–526.

- (17) Braaksma, M.; Bijlsma, S.; Coulier, L.; Punt, P. J.; Van Der Werf, M. J. Metabolomics as a Tool for Target Identification in Strain Improvement: The Influence of Phenotype Definition. *Microbiology* **2011**, *157*, 147–159.
- (18) Kobayashi, S.; Putri, S. P.; Yamamoto, Y.; Donghyo, K.; Bamba, T.; Fukusaki, E. Gas Chromatography-Mass Spectrometry Based Metabolic Profiling for the Identification of Discrimination Markers of *Angelicae Radix* and Its Application to Gas Chromatography-Flame Ionization Detector System. *J. Biosci. Bioeng.* **2012**, *114*, 232–236.
- (19) Kobayashi, S.; Nagasawa, S.; Yamamoto, Y.; Donghyo, K.; Bamba, T.; Fukusaki, E. Metabolic Profiling and Identification of the Genetic Varieties and Agricultural Origin of *Cnidium Officinale* and *Ligusticum Chuanxiong*. *J. Biosci. Bioeng.* **2012**, *114*, 86–91.
- (20) Wishart, D. S. Metabolomics: Applications to Food Science and Nutrition Research. *Trends Food Sci. Technol.* **2008**, *19*, 482–493.
- (21) Ehrman, T. M.; Barlow, D. J.; Hylands, P. J. Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities. *J. Chem. Inf. Model.* **2007**, *47*, 254–263.
- (22) Choi, H. K.; Yoon, J. H.; Kim, Y. S.; Kwon, D. Y. Metabolomic Profiling of Cheonggukjang during Fermentation by ¹H NMR Spectrometry and Principal Components Analysis. *Process Biochem.* **2007**, *42*, 263–266.
- (23) Cajka, T.; Riddellova, K.; Tomaniova, M.; Hajslova, J. Ambient Mass Spectrometry Employing a DART Ion Source for Metabolomic Fingerprinting/profiling: A Powerful Tool for Beer Origin Recognition. *Metabolomics* **2011**, *7*, 500–508.
- (24) Ko, B. K.; Ahn H. J.; Van Den Berg, F.; Lee, C. H.; Hong, Y. S. Metabolomic Insight into Soy Sauce through ¹H NMR Spectroscopy. *J. Agric. Food Chem.* **2009**, *57*, 6862–6870.

- (25) Gibney, Michael J., Walsh, Marianne, Brennan, Lorraine, Roche, Helen M., German, Bruce, van Ommen, B. Metabolomics in Human Nutrition: Opportunities and Challenges 1–3. *Am. J. Clin. Nutr.* **2005**, *82*, 497–503.
- (26) Pereira, G. E.; Gaudillere, J. P.; Pieri, P.; Hilbert, G.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D. Microclimate Influence on Mineral and Metabolic Profiles of Grape Berries. *J. Agric. Food Chem.* **2006**, *54*, 6765–6775.
- (27) Moco, S.; Bino, R. J.; Vorst, O.; Verhoeven, H. A.; de Groot, J.; van Beek, T. A.; Vervoort, J.; de Vos, C. H. R. A Liquid Chromatography-Mass Spectrometry-Based Metabolome Database for Tomato. *Plant Physiol.* **2006**, *141*, 1205–1218.
- (28) Acharjee, A.; Kloosterman, B.; de Vos, R. C.; Werij, J. S.; Bachem, C. W.; Visser, R. G.; Maliepaard, C. Data Integration and Network Reconstruction with -omics Data Using Random Forest Regression in Potato. *Anal. Chim. Acta* **2011**, *705*, 56–63.
- (29) Thissen, U.; Coulier, L.; Overkamp, K. M.; Jetten, J.; van der Werff, B. J. C.; van de Ven, T.; van der Werf, M. J. A Proper Metabolomics Strategy Supports Efficient Food Quality Improvement: A Case Study on Tomato Sensory Properties. *Food Qual. Prefer.* **2011**, *22*, 499–506.
- (30) Suzuki, M.; Kusano, M.; Takahashi, H.; Nakamura, Y.; Hayashi, N.; Kobayashi, M.; Ichikawa, T.; Matsui, M.; Hirochika, H.; Saito, K. Rice-Arabidopsis FOX Line Screening with FT-NIR-Based Fingerprinting for GC-TOF/MS-Based Metabolite Profiling. *Metabolomics* **2010**, *6*, 137–145.
- (31) Kende, A.; Portwood, D.; Senior, A.; Earll, M.; Bolygo, E.; Seymour, M. Target List Building for Volatile Metabolite Profiling of Fruit. *J. Chromatogr. A* **2010**, *1217*, 6718–6723.
- (32) Vaclavik, L.; Lacina, O.; Hajslova, J.; Zweigenbaum, J. The Use of High Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry Coupled to Advanced Data Mining and Chemometric Tools for Discrimination and Classification of Red Wines according to Their Variety. *Anal. Chim. Acta* **2011**, *685*, 45–51.

- (33) Intelmann, D.; Haseleu, G.; Dunkel, A.; Lagemann, A.; Stephan, A.; Hofmann, T. Comprehensive Sensomics Analysis of Hop-Derived Bitter Compounds during Storage of Beer. *J. Agric. Food Chem.* **2011**, *59*, 1939–1953.
- (34) Pongsuwan, W.; Bamba, T.; Yonetani, T.; Kobayashi, A.; Fukusaki, E. Quality Prediction of Japanese Green Tea Using Pyrolyzer Coupled GC/MS Based Metabolic Fingerprinting. *J. Agric. Food Chem.* **2008**, *56*, 744–750.
- (35) Ruosi, M. R.; Cordero, C.; Cagliero, C.; Rubiolo, P.; Bicchi, C.; Sgorbini, B.; Liberto, E. A Further Tool to Monitor the Coffee Roasting Process: Aroma Composition and Chemical Indices. *J. Agric. Food Chem.* **2012**, *60*, 11283–11291.
- (36) Ikeda, T.; Kanaya, S.; Yonetani, T.; Kobayashi, A.; Fukusaki, E. Prediction of Japanese Green Tea Ranking by Fourier Transform near-Infrared Reflectance Spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 9908–9912.
- (37) Rodrigues, J. A.; Barros, A. S.; Carvalho, B.; Brandão, T.; Gil, A. M. Probing Beer Aging Chemistry by Nuclear Magnetic Resonance and Multivariate Analysis. *Anal. Chim. Acta* **2011**, *702*, 178–187.
- (38) Hu, F.; Furihata, K.; Kato, Y.; Tanokura, M. Nondestructive Quantification of Organic Compounds in Whole Milk without Pretreatment by Two-Dimensional NMR Spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 4307–4311.
- (39) Tiziani, S.; Schwartz, S. J.; Vodovotz, Y. Profiling of Carotenoids in Tomato Juice by One- and Two-Dimensional NMR. *J. Agric. Food Chem.* **2006**, *54*, 6094–6100.
- (40) Nord, L. I.; Vaag, P.; Duus, J. Quantification of Organic and Amino Acids in Beer by ¹H NMR Spectroscopy. *Anal. Chem.* **2004**, *76*, 4790–4798.
- (41) Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; Van Der Hoof, J. J. J.; Wishart, D. S. The Food Metabolome: A Window over Dietary Exposure. *Am. J. Clin. Nutr.* **2014**, *99*, 1286–1308.

- (42) Zeisel, S. H. Nutrigenomics and Metabolomics Will Change Clinical Nutrition and Public Health Practice: Insights from Studies on Dietary Requirements for Choline. *Am. J. Clin. Nutr.* **2007**, *86*, 542–548.
- (43) Seeram, N. P.; Henning, S. M.; Zhang, Y.; Suchard, M.; Li, Z.; Heber, D. Pomegranate Juice Ellagitannin Metabolites Are Present in Human Plasma and Some Persist in Urine for up to 48 Hours. *J. Nutr.* **2006**, *136*, 2481–2485.
- (44) John Dennis, M. Recent Developments in Food Authentication. *Analyst* **1998**, *123*, 151–156.
- (45) Cordella, C.; Moussa, I.; Martel, A-C.; Sbirrazzuoli, N.; Louisette, L-C. Recent Developments in Food Characterisation and Adulteration Detection: Technique-Oriented Perspective. *J. Agric. Food Chem.* **2002**, *50*, 1751–1764.
- (46) Wermelinger, T.; D'Ambrosio, L.; Klopprogge, B.; Yeretian, C. Quantification of the Robusta Fraction in a Coffee Blend via Raman Spectroscopy: Proof of Principle. *J. Agric. Food Chem.* **2011**, *59*, 9074–9079.
- (47) ICO Report. Provisional exports by exporting countries to all destinations.
International Coffee Organization. **2015**. <http://www.ico.org/prices/m3.htm>
- (48) Briandet, R.; Kemsley, E. K.; Wilson, R. H. Approaches to Adulteration Detection in Instant Coffees Using Infrared Spectroscopy and Chemometrics. *J. Sci. Food Agric.* **1996**, *71*, 359–366.
- (49) Setoyama, D.; Iwasa, K.; Seta, H.; Shimizu, H.; Fujimura, Y.; Miura, D.; Wariishi, H.; Nagai, C.; Nakahara, K. High-Throughput Metabolic Profiling of Diverse Green Coffea Arabica Beans Identified Tryptophan as a Universal Discrimination Factor for Immature Beans. *PLoS One* **2013**, *8*, e70098.
- (50) Anderson, K. A.; Smith, B. W. Chemical Profiling To Differentiate Geographic Growing Origins of Coffee. *J. Agric. Food Chem.* **2002**, *50*, 2068–2075.

- (51) Briandet, R.; Kemsley, E. K.; Wilson, R. H. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *J. Agric. Food Chem.* **1996**, *44*, 170–174.
- (52) Mancha Agresti, P. D. C.; Franca, A. S.; Oliveira, L. S.; Augusti, R. Discrimination between Defective and Non-Defective Brazilian Coffee Beans by Their Volatile Profile. *Food Chem.* **2008**, *106*, 787–796.
- (53) Mendonça, J. C. F.; Franca, A. S.; Oliveira, L. S.; Nunes, M. Chemical Characterisation of Non-Defective and Defective Green Arabica and Robusta Coffees by Electrospray Ionization-Mass Spectrometry (ESI-MS). *Food Chem.* **2008**, *111*, 490–497.
- (54) El-Abassy, R. M.; Donfack, P.; Materny, A. Discrimination between Arabica and Robusta Green Coffee Using Visible Micro Raman Spectroscopy and Chemometric Analysis. *Food Chem.* **2011**, *126*, 1443–1448.
- (55) Garrett, R.; Vaz, B. G.; Hovell, A. M. C.; Eberlin, M. N.; Rezende, C. M. Arabica and Robusta Coffees: Identification of Major Polar Compounds and Quantification of Blends by Direct-Infusion Electrospray Ionization – Mass Spectrometry. *J. Agric. Food Chem.* **2012**, *60*, 4255–4258.
- (56) Wei, F.; Furihata, K.; Koda, M.; Hu, F.; Kato, R.; Miyakawa, T.; Tanokura, M. C NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin. *J. Agric. Food Chem.* **2012**, *60*, 10118–19125.
- (57) Marcone, M. F. Composition and Properties of Indonesian Palm Civet Coffee (Kopi Luwak) and Ethiopian Civet Coffee. *Food Res. Int.* **2004**, *37*, 901–912.
- (58) Grassman Jr., L. I. Movement and fruit selection of two Paradoxurinae species in a dry evergreen forest in Southern Thailand. *Small Carnivore Conservation.* **1998**, *19*, 25-29.
- (59) Ongo, E.; Falasconi, M.; Sberveglieri, G.; Antonelli, A.; Montevecchi, G.; Sberveglieri, V.; Concina, I.; Sevilla III, F. Chemometric Discrimination of Philippine Civet Coffee

- Using Electronic Nose and Gas Chromatography Mass Spectrometry. *Procedia Engineering*; 2012; Vol. 47, pp. 977–980.
- (60) Luwak White Koffie. **2015**. <http://kopiluwak.id/luwak-white-koffie>.
- (61) Blow, N. Metabolomics: Biochemistry's New Look. *Nature* **2008**, *455*, 697–700.
- (62) Choi, M. Y.; Choi, W.; Park, J. H.; Lim, J.; Kwon, S. W. Determination of Coffee Origins by Integrated Metabolomic Approach of Combining Multiple Analytical Data. *Food Chem.* **2010**, *121*, 1260–1268.
- (63) Fujimura, Y.; Kurihara, K.; Ida, M.; Kosaka, R.; Miura, D.; Wariishi, H.; Maeda-Yamamoto, M.; Nesumi, A.; Saito, T.; Kanda, T.; et al. Metabolomics-Driven Nutraceutical Evaluation of Diverse Green Tea Cultivars. *PLoS One* **2011**, *6*, e23426.
- (64) Pongsuwan, W.; Fukusaki, E.; Bamba, T.; Yonetani, T.; Yamahara, T.; Kobayashi, A. Prediction of Japanese Green Tea Ranking by Gas Chromatography/mass Spectrometry-Based Hydrophilic Metabolite Fingerprinting. *J. Agric. Food Chem.* **2007**, *55*, 231–236.
- (65) Jonsson, P.; Johansson, E. S.; Wuolikainen, A.; Lindberg, J.; Schuppe-Koistinen, I.; Kusano, M.; Sjöström, M.; Trygg, J.; Moritz, T.; Antti, H. Predictive Metabolite Profiling Applying Hierarchical Multivariate Curve Resolution to GC-MS Data - A Potential Tool for Multi-Parametric Diagnosis. *J. Proteome Res.* **2006**, *5*, 1407–1414.
- (66) Wiklund, S.; Johansson, E.; Sjö, L.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J. Visualization of GC / TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models. *Anal. Chem.* **2008**, *80*, 115–122.
- (67) Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, *16*, 119–128.
- (68) Guan, H.; Brewer, W. E.; Morgan, S. L. New Approach to Multiresidue Pesticide Determination in Foods with High Fat Content Using Disposable Pipette Extraction

- (DPX) and Gas Chromatography-Mass Spectrometry (GC-MS). *J. Agric. Food Chem.* **2009**, *57*, 10531–10538.
- (69) The European Agency for the Evaluation of Medicinal Products, Human Medicines Evaluation Unit, Validation of Analytical Procedures: Methodology, ICH Topic Q 2B, November **1996**.
- (70) Plant Research International. MetAlign. **2015**.
<http://www.wageningenur.nl/nl/Expertises-Dienstverlening/Onderzoeksinstituten/rikilt/show/MetAlign.htm>.
- (71) Sumner, L. W.; Amberg, A.; Barret, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fieh, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriot, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed Minimum Reporting Standards for Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. **2013**, *3*, 211–221.
- (72) Tsugawa, H.; Tsujimoto, Y.; Arita, M.; Bamba, T.; Fukusaki, E. GC/MS Based Metabolomics: Development of a Data Mining System for Metabolite Identification by Using Soft Independent Modeling of Class Analogy (SIMCA). *BMC Bioinformatics* **2011**, *12*, 131.
- (73) Dallüge, J.; Vreuls, R. J. J.; Van Iperen, D. J.; Van Rijn, M.; Brinkman, U. A. T. Resistively Heated Gas Chromatography Coupled to Quadrupole Mass Spectrometry. *J. Sep. Sci.* **2002**, *25*, 608–614.
- (74) Shimadzu, Inc. GCMS-QP2010 Ultra. **2015**.
<http://www.shimadzu.com/an/gcms/qpultra.html>.
- (75) Moon, J. K.; Hyui Yoo, S. U. N.; Shibamoto, T. Role of Roasting Conditions in the Level of Chlorogenic Acid Content in Coffee Beans: Correlation with Coffee Acidity. *J. Agric. Food Chem.* **2009**, *57*, 5365–5369.

- (76) Rodrigues, C. I.; Marta, L.; Maia, R.; Miranda, M.; Ribeirinho, M.; Máguas, C. Application of Solid-Phase Extraction to Brewed Coffee Caffeine and Organic Acid Determination by UV/HPLC. *J. Food Compos. Anal.* **2007**, *20*, 440–448.
- (77) Blumberg, S.; Frank, O.; Hofmann, T. Quantitative Studies on the Influence of the Bean Roasting Parameters and Hot Water Percolation on the Concentrations of Bitter Compounds in Coffee Brew. *J. Agric. Food Chem.* **2010**, *58*, 3720–3728.
- (78) Perrone, D.; Donangelo, C. M.; Farah, A. Fast Simultaneous Analysis of Caffeine, Trigonelline, Nicotinic Acid and Sucrose in Coffee by Liquid Chromatography-Mass Spectrometry. *Food Chem.* **2008**, *110*, 1030–1035.
- (79) Wei, F.; Furihata, K.; Koda, M.; Hu, F.; Miyakawa, T.; Tanokura, M. Roasting Process of Coffee Beans as Studied by Nuclear Magnetic. *J. Agric. Food Chem.* **2012**, *60*, 1005–1012.
- (80) Liberto, E.; Ruosi, M. R.; Cordero, C.; Rubiolo, P.; Bicchi, C.; Sgorbini, B. Non-Separative Headspace Solid Phase Microextraction-Mass Spectrometry Profile as a Marker to Monitor Coffee Roasting Degree. *J. Agric. Food Chem.* **2013**, *61*, 1652–1660.
- (81) Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. Experimental Design and Optimization. *Chemom. Intell. Lab. Syst.* **1998**, *42*, 3–40.
- (82) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361–1375.
- (83) Wiggins, J. N. Botany, Genetics and Genomics of Coffee. In *Coffee: Growing, Processing, Sustainable Production a Guidebook for Growers, Processors, Traders and Researchers*, second revised edition. **2009**. Germany: Wiley VCH.

- (84) Ginz, M.; Balzer, H. H.; Bradbury, a. G. W.; Maier, H. G. Formation of Aliphatic Acids by Carbohydrate Degradation during Roasting of Coffee. *Eur. Food Res. Technol.* **2000**, *211*, 404–410.
- (85) Ibáñez, A. J.; Fagerer, S. R.; Schmidt, A. M.; Urban, P. L.; Jefimovs, K.; Geiger, P.; Dechant, R.; Heinemann, M.; Zenobi, R. Mass Spectrometry-Based Metabolomics of Single Yeast Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8790–8794.
- (86) Zenobi, R. Single-Cell Metabolomics: Analytical and Biological Perspectives. *Science* **2013**, *342*, 1243259.
- (87) Shiga, K.; Yamamoto, S.; Nakajima, A.; Kodama, Y.; Imamura, M.; Sato, T.; Uchida, R.; Obata, A.; Bamba, T.; Fukusaki, E. Metabolic Profiling Approach To Explore Compounds Related to the Umami Intensity of Soy Sauce. *J. Agric. Food Chem.* **2014**, *62*, 7317–7322.
- (88) Yamamoto, S.; Shiga, K.; Kodama, Y.; Imamura, M.; Uchida, R.; Obata, A.; Bamba, T.; Fukusaki, E. Analysis of the Correlation between Dipeptides and Taste Differences among Soy Sauces by Using Metabolomics-Based Component Profiling. *J. Biosci. Bioeng.* **2014**, *118*, 56–63.
- (89) Mimura, N.; Isogai, A.; Iwashita, K.; Bamba, T.; Fukusaki, E. Gas Chromatography/mass Spectrometry Based Component Profiling and Quality Prediction for Japanese Sake. *J. Biosci. Bioeng.* **2014**, *118*, 406–414.
- (90) Taguchi, K.; Fukusaki, E.; Bamba, T. Simultaneous Analysis for Water- and Fat-Soluble Vitamins by a Novel Single Chromatography Technique Unifying Supercritical Fluid Chromatography and Liquid Chromatography. *J. Chromatogr. A* **2014**, *1362*, 270–277.
- (91) Jumtee, K.; Bamba, T.; Fukusaki, E. Fast GC-FID Based Metabolic Fingerprinting of Japanese Green Tea Leaf for Its Quality Ranking Prediction. *J. Sep. Sci.* **2009**, *32*, 2296–2304.

- (92) Zhou, J.; Ma, G.; Yuan, K.; Lu, X.; Zhu, Z.; Wu, Y.; Xu, G.. Metabolic Profiling of Transgenic Rice with *cryIAc* and *scK* Genes: an Evaluation of Unintended Effects at Metabolic Level by Using GC-FID and GC-MS. *J. Chromatogr. B* **2009**, *877*, 725–732.
- (93) Ma, C.; Wang, H.; Lu, X.; Xu, G.; Liu, B. Metabolic Fingerprinting Investigation of *Artemisia Annua* L. in Different Stages of Development by Gas Chromatography and Gas Chromatography-Mass Spectrometry. *J. Chromatogr. A* **2008**, *1186*, 412–419.
- (94) Yeh, T.; Morris, C. R.; Goldfarb, B.; Chang, H.; Kadla, J. F. Utilization of Polar Metabolite Profiling in the Comparison of Juvenile Wood and Compression Wood in Loblolly Pine (*Pinus taeda*). *Tree Physiol.* **2006**, *26*, 1497–1503.
- (95) De Livera, A. M.; Dias, D. A; De Souza, D.; Rupasinghe, T.; Pyke, J.; Tull, D.; Roessner, U.; McConville, M.; Speed, T. P. Normalizing and Integrating Metabolomics Data. *Anal. Chem.* **2012**, *84*, 10768–10776.
- (96) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Metabolite Projection Analysis for Fast Identification of Metabolites in Metabonomics. Application in an Amiodarone Study. *Anal. Chem.* **2006**, *78*, 3551–3561.
- (97) Moco, S.; Vervoort, J.; Bino, R. J.; De Vos, R. C. H.; Bino, R. Metabolomics Technologies and Metabolite Identification. *TrAC Trends Anal. Chem.* **2007**, *26*, 855–866.
- (98) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. a; Jones, D. P.; Pulendran, B. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* **2013**, *9*, e1003123.
- (99) Wang, Z.; Zhang, M.; Harrington, P. D. B. Comparison of Three Algorithms for the Baseline Correction of Hyphenated Data Objects. *Anal. Chem.* **2014**, *86*, 9050–9057.
- (100) Eriksson, L.; Trygg, J.; Wold, S. A Chemometrics Toolbox Based on Projections and Latent Variables. *J. Chemom.* **2014**, *28*, 332–346.
- (101) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

- (102) Campo, E.; Ferreira, V.; Escudero, A.; Marqués, J. C.; Cacho, J. Quantitative Gas Chromatography–olfactometry and Chemical Quantitative Study of the Aroma of Four Madeira Wines. *Anal. Chim. Acta* **2006**, *563*, 180–187.
- (103) Le Gall, G.; Colquhoun, I. J.; Defernez, M. Metabolite Profiling Using ¹H NMR Spectroscopy for Quality Assessment of Green Tea, *Camellia Sinensis* (L.). *J. Agric. Food Chem.* **2004**, *52*, 692–700.
- (104) Wei, F.; Furihata, K.; Miyakawa, T.; Tanokura, M. A Pilot Study of NMR-Based Sensory Prediction of Roasted Coffee Bean Extracts. *Food Chem.* **2014**, *152*, 363–369.
- (105) Nebesny, E.; Budryn, G. Evaluation of Sensory Attributes of Coffee Brews from Robusta Coffee Roasted under Different Conditions. *Eur. Food Res. Technol.* **2006**, *224*, 159–165.
- (106) Monfreda, M.; Gobbi, L.; Grippa, A. Blends of Olive Oil and Seeds Oils: Characterisation and Olive Oil Quantification Using Fatty Acids Composition and Chemometric Tools. Part II. *Food Chem.* **2014**, *145*, 584–592.
- (107) Monfreda, M.; Gobbi, L.; Grippa, A. Blends of Olive Oil and Sunflower Oil: Characterisation and Olive Oil Quantification Using Fatty Acid Composition and Chemometric Tools. *Food Chem.* **2012**, *134*, 2283–2290.
- (108) Pizarro, C.; Esteban-Díez, I.; González-Sáiz, J. M. Mixture Resolution according to the Percentage of Robusta Variety in Order to Detect Adulteration in Roasted Coffee by near Infrared Spectroscopy. *Anal. Chim. Acta* **2007**, *585*, 266–276.
- (109) Wang, J.; Jun, S.; Bittenbender, H. C.; Gautz, L.; Li, Q. X. Fourier Transform Infrared Spectroscopy for Kona Coffee Authentication. *J. Food Sci.* **2009**, *74*, C385–C391.
- (110) Kvalheim, O. M.; Karstang, T. V. Interpretation of Latent-Variable Regression Models. *Chemom. Intell. Lab. Syst.* **1989**, *7*, 39–51.

- (111) Cheong, M. W.; Tong, K. H.; Ong, J. J. M.; Liu, S. Q.; Curran, P.; Yu, B. Volatile Composition and Antioxidant Capacity of Arabica Coffee. *Food Res. Int.* **2013**, *51*, 388–396.
- (112) Ernst, M.; Silva, D. B.; Silva, R. R.; Vêncio, R. Z. N.; Lopes, N. P. Mass Spectrometry in Plant Metabolomics Strategies: From Analytical Platforms to Data Acquisition and Processing. *Nat. Prod. Rep.* **2014**, *31*, 784–806.
- (113) Bjarnholt, N.; Li, B.; D'Alvise, J.; Janfelt, C. Mass Spectrometry Imaging of Plant Metabolites--Principles and Possibilities. *Nat. Prod. Rep.* **2014**, *31*, 818–837.
- (114) Dickschat, J. S. Capturing Volatile Natural Products by Mass Spectrometry. *Nat. Prod. Rep.* **2014**, *31*, 838–861.
- (115) Riedel, A.; Hochkogler, C. M.; Lang, R.; Bytof, G.; Lantz, I.; Hofmann, T.; Somoza, V. N-Methylpyridinium, a Degradation Product of Trigonelline upon Coffee Roasting, Stimulates Respiratory Activity and Promotes Glucose Utilization in HepG2 Cells. *Food Funct.* **2014**, *5*, 454–462.
- (116) Folwarczna, J.; Zych, M.; Nowińska, B.; Pytlik, M.; Janas, A. Unfavorable Effect of Trigonelline, an Alkaloid Present in Coffee and Fenugreek, on Bone Mechanical Properties in Estrogen-Deficient Rats. *Mol. Nutr. Food Res.* **2014**, *58*, 1457–1464.
- (117) Ludwig, I. A.; Clifford, M. N.; Lean, M. E. J.; Ashihara, H.; Crozier, A. Coffee: Biochemistry and Potential Impact on Health. *Food Funct.* **2014**, *5*, 1695–1717.
- (118) Borota, D.; Murray, E.; Keceli, G.; Chang, A.; Watabe, J. M.; Ly, M.; Toscano, J. P.; Yassa, M. A. Post-Study Caffeine Administration Enhances Memory Consolidation in Humans. *Nat. Neurosci.* **2014**, *17*, 201–203.
- (119) Van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. *BMC Genomics.* **2006**, *7*, 142.

- (120) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets. *Anal. Chem.* **2006**, *78*, 2262–2267.
- (121) Steuer, R.; Morgenthal, K.; Weckwerth, W.; Selbig, J. A gentle guide to the analysis of metabolomics data. *Methods in Mol Biol.* **2007**, *358*, 105-126.
- (122) Pinto, R. C.; Gerber, L.; Eliasson, M.; Sundberg, B.; Trygg, J. Strategy for Minimizing between-Study Variation of Large-Scale Phenotypic Experiments Using Multivariate Analysis. *Anal. Chem.* **2012**, *84*, 8675–8681.
- (123) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. Scaling. In *Multi and megavariate data analysis, part 2, advanced applications and method extension (second revision edition)*. 2006. Umea, Sweden: Umetrics AB.
- (124) Eriksson, L.; Trygg, J.; Johansson, E.; Bro, R.; Wold, S. Orthogonal Signal Correction, Wavelet Analysis, and Multivariate Calibration of Complicated Process Fluorescence Data. *Anal. Chim. Acta* **2000**, *420*, 181–195.
- (125) Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some Recent Developments in PLS Modeling. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 131–150.
- (126) Gromski, P. S.; Xu, Y.; Hollywood, K. A.; Turner, M. L.; Goodacre, R. The Influence of Scaling Metabolomics Data on Model Classification Accuracy. *Metabolomics* **2015**, *11*, 684–695.

Appendices

Supplementary Table 2-1. List of coffee samples (Validation coffee set) (Chapter 2)

No.	Brand ^a	Origin	Species	Type	Production year	Samples extraction period
1	Andungsari	East Java, Indonesia	<i>Arabica</i>	(Authentic) Civet coffee	2012	Feb 2013
2	Golden	West Java, Indonesia	<i>Arabica</i>	(Commercial) Civet coffee	2012	Feb 2013
3	Wahana	Sumatra, Indonesia	<i>Arabica</i>	(Commercial) Civet coffee	2012	Feb 2013
4	Bali	Bali, Indonesia	<i>Unknown</i>	(Commercial) Civet coffee	2011	Feb 2013
5	Wahana	Sumatra, Indonesia	<i>Arabica</i>	(Commercial) regular coffee	2012	Feb 2013
6	Kona coffee	Hawaii, USA	<i>Arabica</i>	(Commercial) regular coffee	2012	Feb 2013
7	Cerrado Chapado coffee	Brazil	<i>Arabica</i>	(Commercial) regular coffee	2012	Feb 2013
8	Aceh Special coffee	Sumatra, Indonesia	<i>Arabica</i>	(Commercial) regular coffee	2012	Feb 2013
9	Confidential brand	Java, Indonesia	<i>Arabica</i>	Fake coffee	2012	Feb 2013
10	Coffee blend A	Blend between “Wahana (civet coffee)” and “Wahana (regular coffee)” with ratio of 50:50 (wt %)				
11	Coffee blend B	Blend between “Golden” and “Wahana (regular coffee)” with ratio of 50:50 (wt %)				

^aProvider

- | | | | |
|-----|-----------------------------------|-----|---------------------------------|
| (1) | ICCRI, Indonesia | (6) | Hiro coffee, Japan |
| (2) | CV. Kopi Luwak, Indonesia | (7) | Hiro coffee, Japan |
| (3) | Wahana-Mandheling Kopi, Indonesia | (8) | Hiro coffee, Japan |
| (4) | Hema-Wiwi Bali, Indonesia | (9) | PT. Java Prima Abadi, Indonesia |
| (5) | Wahana-Mandheling Kopi, Indonesia | | |

For Box-plot

Set A: Wahana (civet coffee), Coffee blend A, Wahana (regular coffee), and fake coffee

Set B: Golden, Coffee blend B, Wahana (regular coffee), and fake coffee

Supplementary Table 2-2. Tentative and confident metabolite identification

Chapter 2					Chapter 4				
Metabolite	RT (min)	RI	Nominal mass	Identification method ^a	Metabolite	RT (min)	RI	Nominal mass	Identification method ^a
Hydroxypyridine	4.61	1036.86	152	A, B	2-Hydroxypyridine	4.35	1032.86	152	A, B
Pyruvate	4.69	1044.72	174	A	Pyruvate	4.45	1040.95	174	A
Lactic acid	4.79	1053.43	117	A, B	Lactic acid	4.55	1049.79	147	A, B
Glycolic acid	4.99	1070.33	147	A, B, C	Glycolic acid	4.75	1067.31	147	A, B, C
n-Butylamine	5.38	1101.33	174	A	ND				
Hydroxybutyric acid	5.56	1118.95	131	A, B	Hydroxybutyric acid	5.31	1117.71	131	A, B
Oxalate	5.63	1125.33	147	A	ND				
Glycerol	7.093	1262.68	147	A	Glycerol	6.83	1262.32	147	A, B
Pyrophosphate	7.095	1262.52	299	A, B	Pyrophosphate	6.87	1266.11	299	A, B
Nicotinic acid	7.45	1295.98	180	A, B	Nicotinic acid	7.18	1295.88	180	A, B
Succinic acid	7.55	1306.23	147	A	Succinic acid	7.28	1306.92	147	A
Catechol	7.62	1313.76	254	A, B	Catechol	7.34	1313.65	254	A, B
Glyceric acid	7.65	1317.77	147	A, B	Glyceric acid	7.39	1318.44	147	A, B
Fumaric acid	7.89	1343.41	245	A, B	Fumaric acid	7.63	1343.44	245	A, B
ND					Methylmaleic acid	8.08	1392.19	147	A, B
Glutaric acid	8.32	1387.93	147	A	Glutaric acid	8.14	1398.05	147	A, B
Malic acid	9.09	1473.82	147	A, B, C	Malic acid	8.83	1476.1	147	A, B, C
Trigonelline	9.12	1477.46	210	B	Trigonelline	9.14	1512.65	210	B
Aspartic acid	9.38	1505.19	232	A, B	ND				
Pyroglutamic acid	9.47	1517.45	156	A, B, C	Pyroglutamic acid	9.19	1519.75	156	A, B, C
Pyrogallol	9.6	1532.66	239	A	Pyrogallol	9.32	1534.75	239	A
ND					Benzoic acid	9.55	1562.73	267	A
Arabinose	10.47	1640.22	103	A, B	Arabinose	10.19	1643.91	103	A, B
Xylulose	10.56	1653.98	205	A, B	Xylulose	10.3	1657.46	205	A, B
Ribose	10.57	1654.84	103	A, B	Ribose	10.31	1658.32	103	A, B
ND					1,6-Anhydroglucose	10.59	1694.88	204	A, B

Supplementary Table 2-2 (Continued)

Chapter 2					Chapter 4				
Metabolite	RT (min)	RI	Nominal mass	Identification method^a	Metabolite	RT (min)	RI	Nominal mass	Identification method^a
Arabitol	10.94	1700.51	217	A, B	Arabitol	10.63	1688.61	217	A, B
Ribitol	10.95	1700.75	217	A, B, C	Ribitol	10.67	1705.05	217	A, B, C
Shikimic acid	11.57	1785.81	204	A	Shikimic acid	11.29	1788.72	147	A
Citric acid	11.67	1798.12	147	A, B, C	Citric acid	11.38	1802.25	147	A, B, C
Quinic acid	11.93	1836.38	345	A, B, C	Quinic acid	11.65	1840.33	345	A, B, C
Fructose	11.97	1843.07	103	A, B	Fructose	11.71	1849.9	103	A, B
Sorbose	12.07	1855.89	217	A, B	Sorbose	11.78	1859.72	217	A, B
Glucose	12.15	1867.45	147	A, B	Glucose	11.86	1870.97	147	A, B
Caffeine	12.19	1873.08	194	A, B, C	Caffeine	11.89	1875.53	194	A, B, C
Sorbitol	12.45	1911.54	147	A, B	Sorbitol	12.17	1915.74	147	A, B
Inositol	13.51	2071.76	217	A, B, C	Inositol	13.21	2075.39	217	A, B, C
Caffeic acid	13.85	2127.54	219	A, B, C	Caffeic acid	13.55	2131.29	219	A, B, C
Tryptophan	14.36	2211.41	202	A, B	ND				
Octadecanoate	14.49	2234.82	117	A, B	Octadecanoic acid	14.19	2237.86	117	A, B
Uridine	15.79	2466.49	217	A, B	Uridine	15.45	2465.55	217	A, B
Sucrose	16.49	2598.29	217	A, B	Sucrose	16.17	2603.78	217	A, B, C
Melibiose	17.33	2770.21	204	A, B	Melibiose	17.01	2775.94	204	A, B
Chlorogenic acid	18.76	3083.23	345	A, B	Chlorogenic acid	18.42	3089.79	345	A, B, C

^a(a = in-house mass spectral library, b = NIST library, c = authentic standard). Comparison with only mass spectral libraries termed as tentative identification (A or B or A, B). Confident identification was done by comparison with libraries and authentic standard (A, B, C) ND = not detected or not matched with any libraries.

Supplementary Table 3-1. List of coffee samples (Validation coffee set) (Chapter 3)

No.	Brand ^a	Production area	Classification	Production year	Samples ID ^b
1	Andungsari	Java, Indonesia	Civet coffee (authentic)	2012	Au_1
2	Andungsari	Java, Indonesia	Civet coffee (authentic)	2013	Au_2
3	Golden	Java, Indonesia	Civet coffee (commercial)	2012	GO
4	Wahana	Sumatra, Indonesia	Civet coffee (commercial)	2012	WL
5	Bali	Bali, Indonesia	Civet coffee (commercial)	2011	BA
6	Wahana	Sumatra, Indonesia	Regular coffee (commercial)	2012	WR
7	Kona	Hawaii, USA	Regular coffee (commercial)	2012	KO
8	Cerrado Chapado	Brazil, Brazil	Regular coffee (commercial)	2012	CE
9	Coffee blends (9 samples)	Combination of 3 civet coffee (commercial) & 3 regular coffee (commercial)			

^aProvider

- 1, 2 ICCRI, Indonesia
- 3 CV. Kopi Luwak, Indonesia
- 4, 5 Wahana-Mandheling, Indonesia
- 6 Hema-Wiwi Bali, Indonesia
- 7, 8 Hiro coffee, Japan

^bFor multivariate analysis

Supplementary Table 4-1. List of coffee samples for construction and verification of OPLS prediction model (Chapter 4).

Coffee sets	No.	Brand	Production area	Classification	Production year
Authentic coffee set ^a	1	Andungsari	Java, Indonesia	Civet coffee	2012
	2	Andungsari	Java, Indonesia	Regular coffee	2013
Commercial coffee set	3	Kopi Luwak Sidikalang	Sumatra, Indonesia	Civet coffee	2012
	4	Wahana Sidikalang	Sumatra, Indonesia	Regular coffee	2012
External validation set I (For authentic coffee set)	Mixing between civet coffee No. 1 and unknown commercial regular coffee A ^b				
External validation II (For commercial coffee set)	Mixing between civet coffee No. 3 and unknown commercial regular coffee B ^b				

^aCoffee samples were roasted in Probat-Werke von Gimborn Maschinenfabrik GmbH model BRZ 2 (Probat, Rhein, Germany) at 205°C for 10 min and followed by immediate air-cooling for 5 min. Roasted coffee beans were kept in sealed Falcon tubes at -30°C until they were used (experiment was done in early to mid 2013)

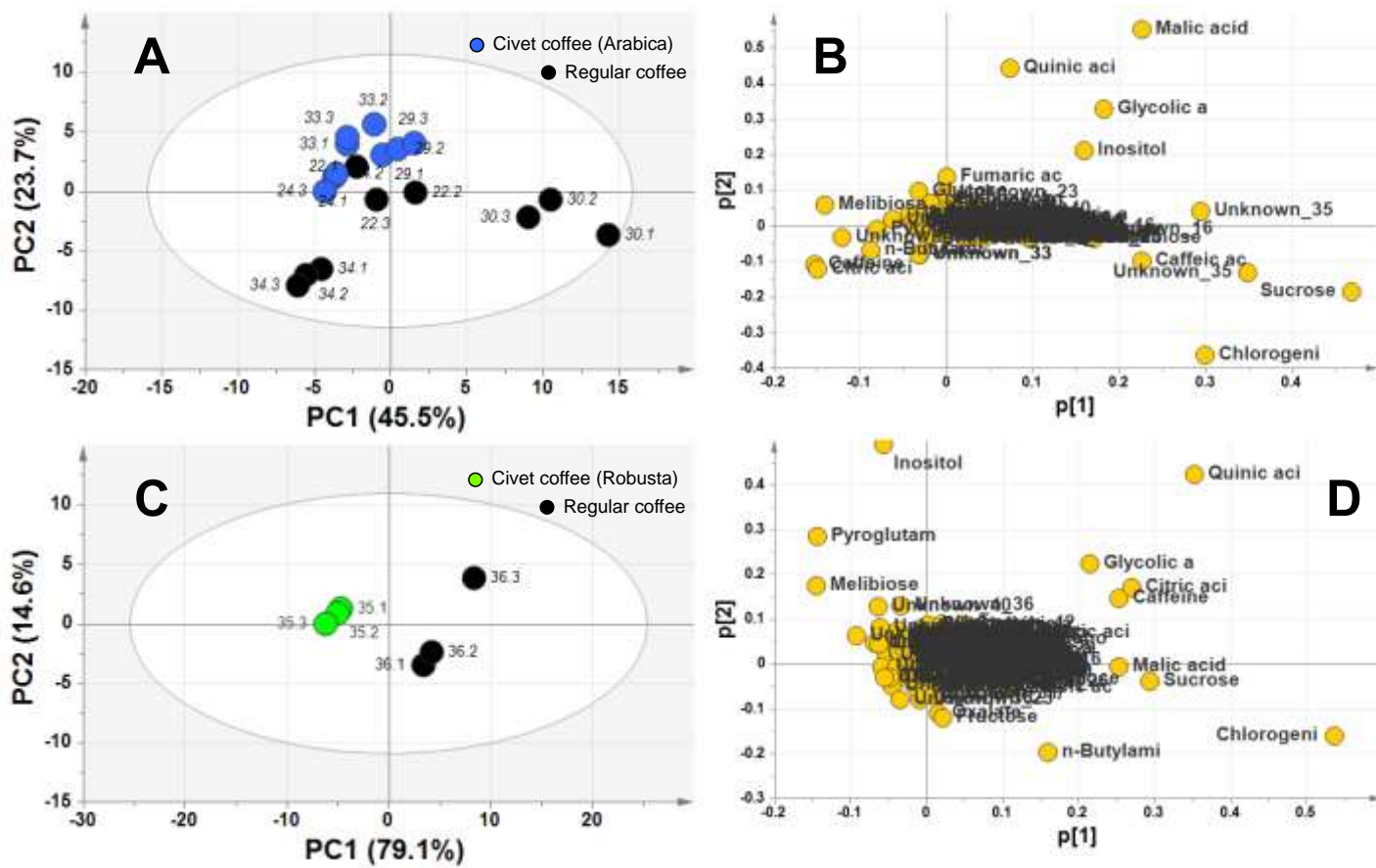
^bWith mixing percentage, 25 and 75% of civet coffee contents. Regular coffee A and B were acquired commercially from local market.

Supplementary Table 4-2. Analytical parameters for quantification

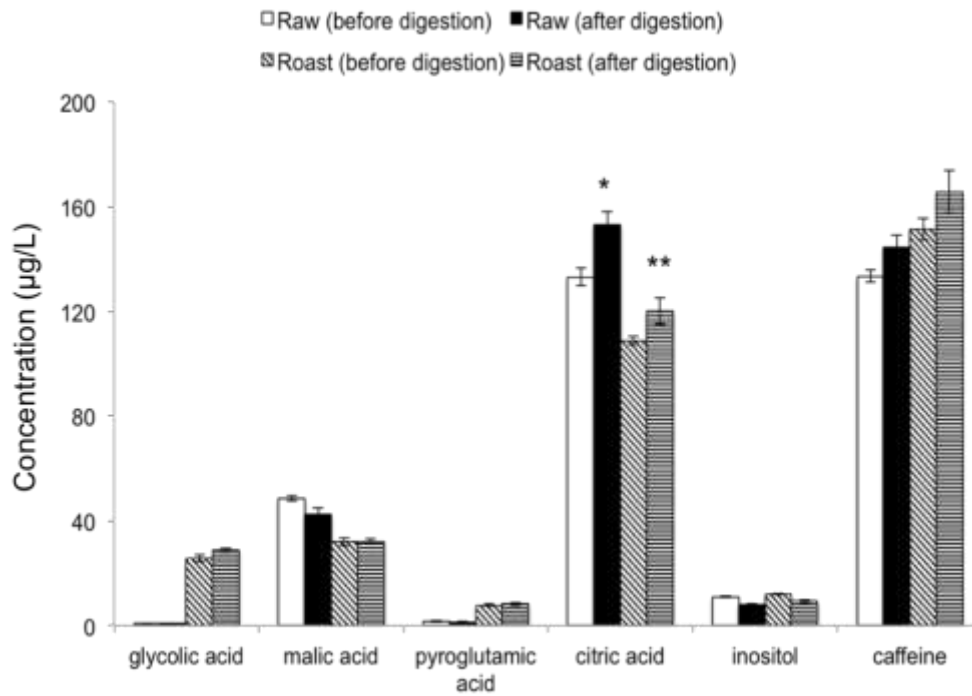
Significant metabolites	RT (min)	RSD [%] (<i>n</i> = 4)		<i>R</i>²	Linearity Range (μM)	LOD (μM)	LOQ (μM)
		RT	Area^a				
Malic acid	8.82	0.06	7.74	0.998	1 – 1000	0.737	2.234
Citric acid	11.38	0.04	4.35	0.994	1 – 1000	1.087	3.295
Chlorogenic acid	18.42	0.03	9.9	0.998	1 – 1000	0.429	1.299

^aAt concentration 100μM

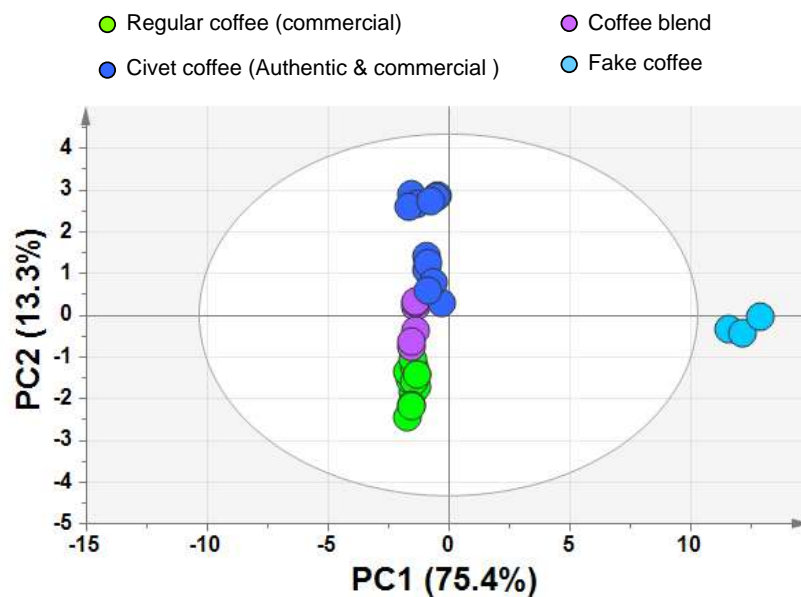
Supplementary Fig. 2-1. PCA score and loading of civet and regular coffee from same cultivation area, Arabica (A, B) and Robusta (C, D).



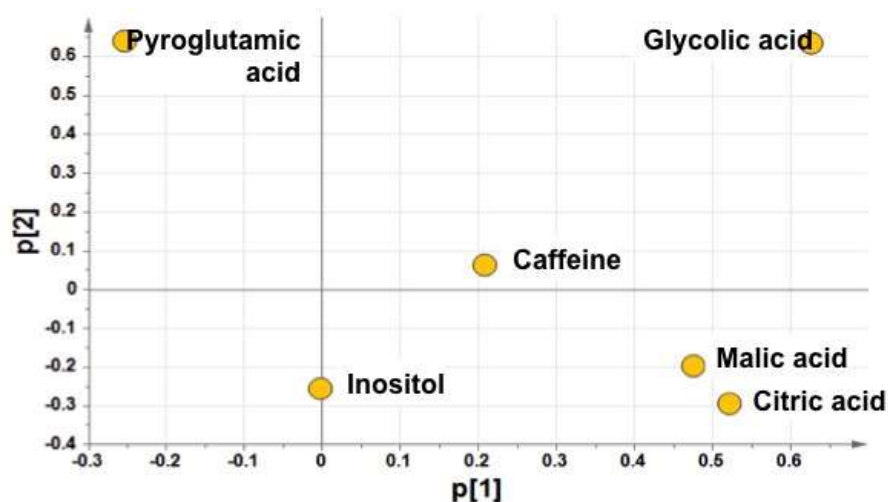
Supplementary Fig. 2-2. Concentrations levels of six marker candidates from controlled processing (samples no.5 and 11 experimental set). *, $p < 0.001$; **, $p < 0.05$.



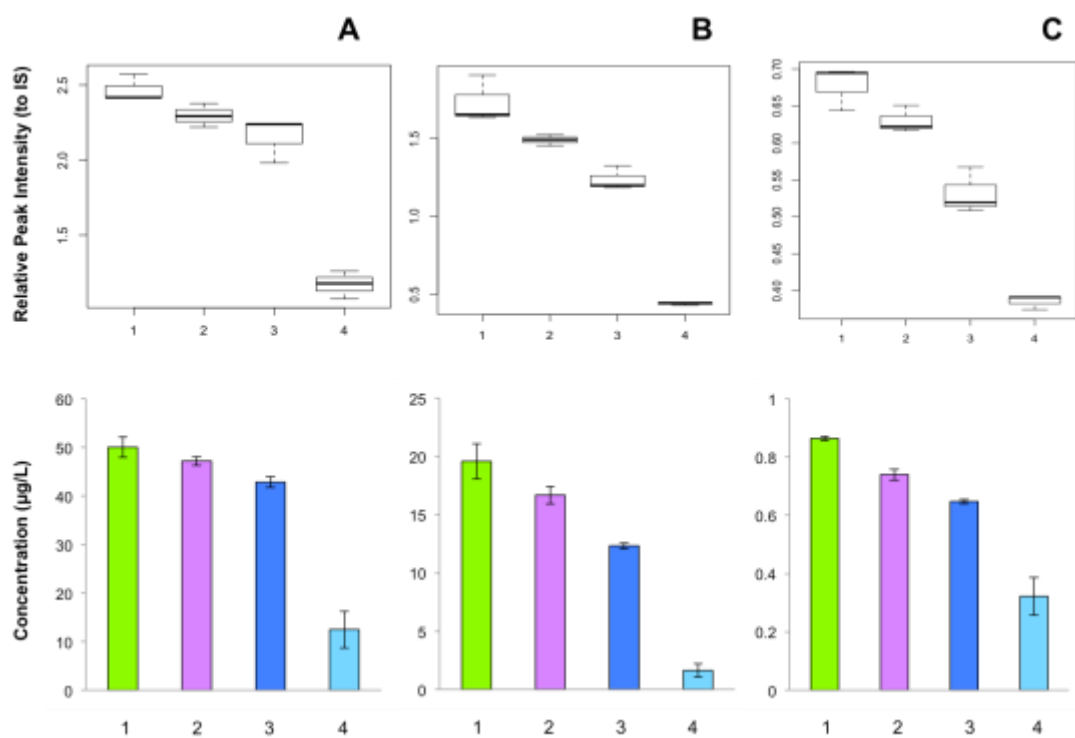
Supplementary Fig. 2-3. PCA score plot of validation coffee set using all detected compounds.



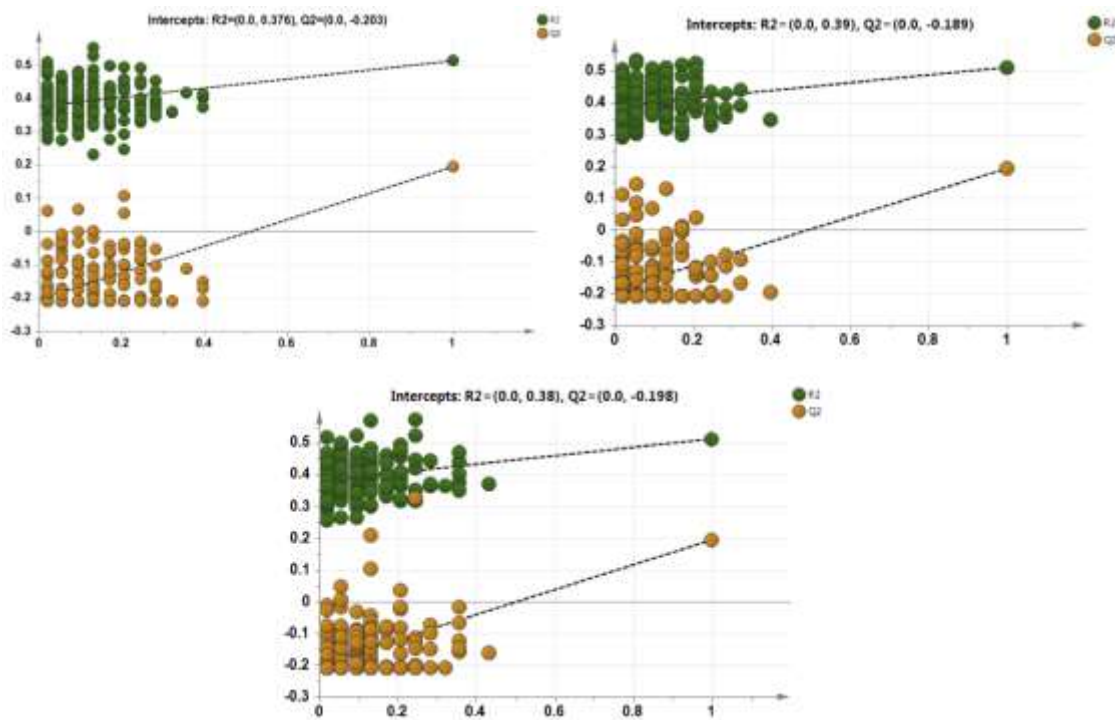
Supplementary Fig. 2-4. Loading plot of PCA derived from validation coffee set in Fig. 2-3.



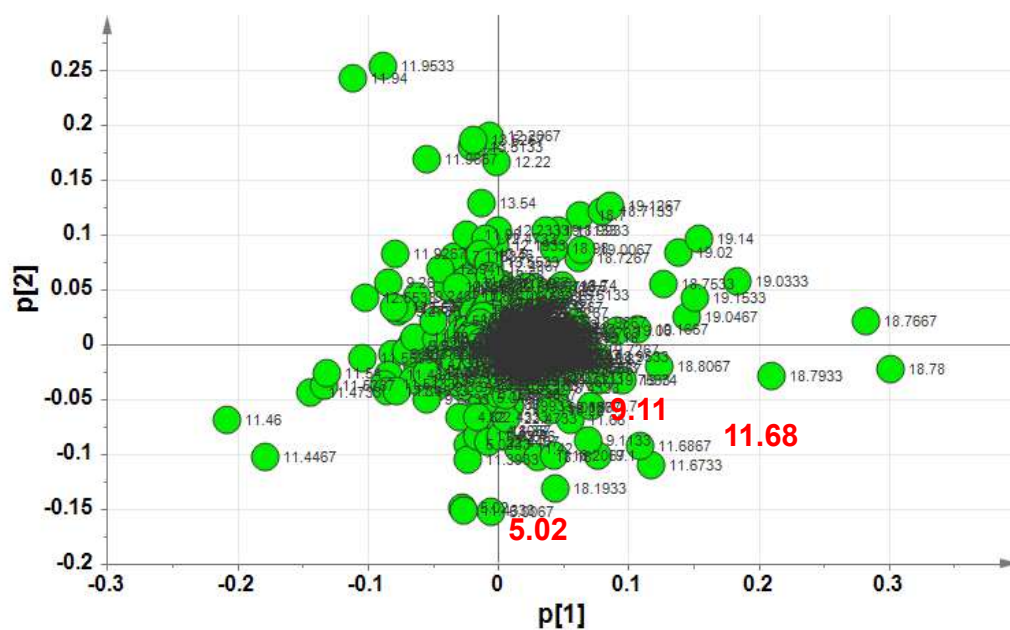
Supplementary Fig. 2-5. Box plot of peak intensity and concentrations of selected discriminant markers from set B. Sample description of set B is displayed in **Supplementary Table 2-1**. (A) citric acid, (B) malic acid, and (C) inositol/pyroglutamic acid, of four coffee samples, (1) civet coffee “Golden”, (2) coffee blend set B, (3) regular coffee “Wahana”, and (4) fake coffee.



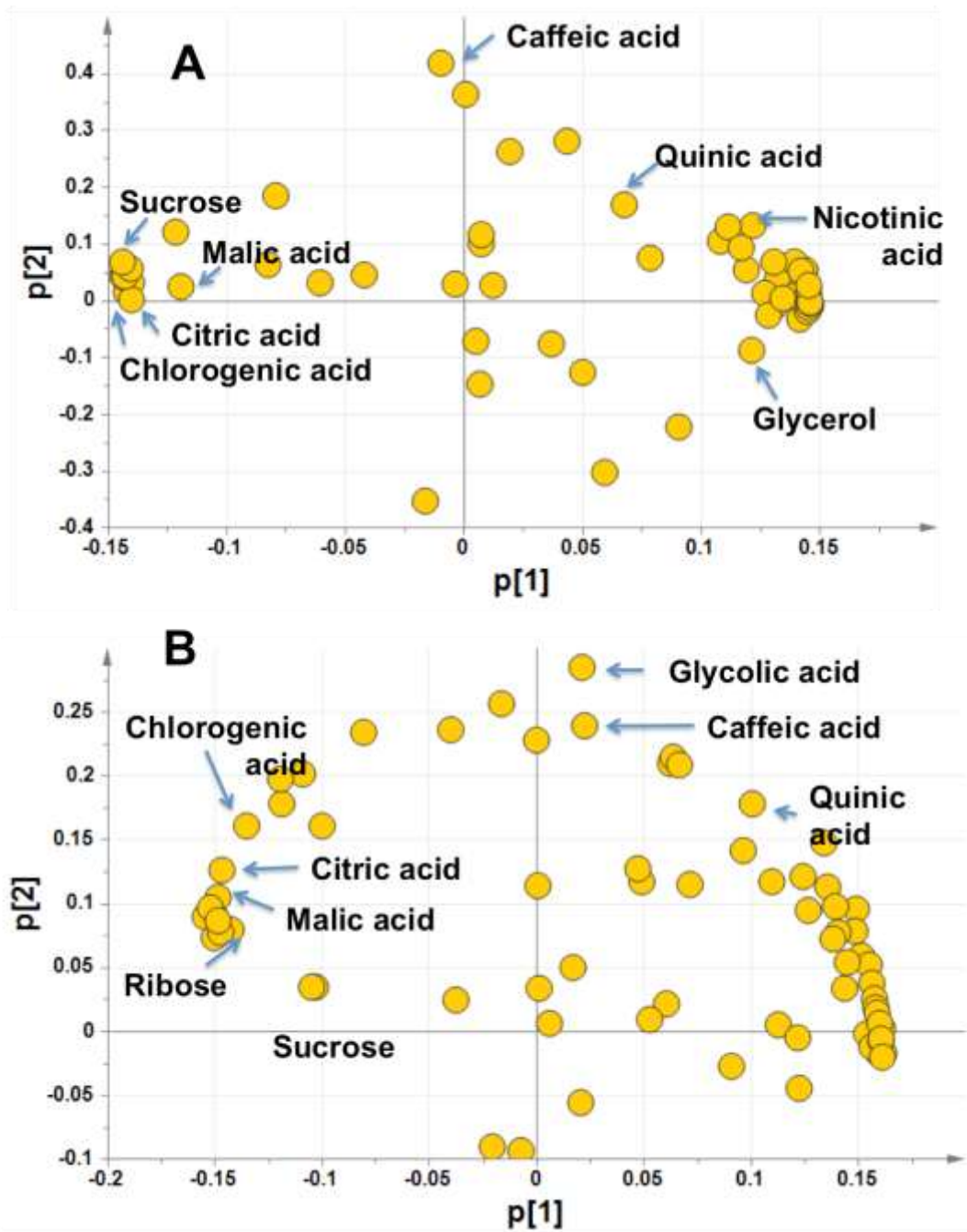
Supplementary Fig. 3-1. Permutation test of OPLS-DA models with 200 random permutation variables.



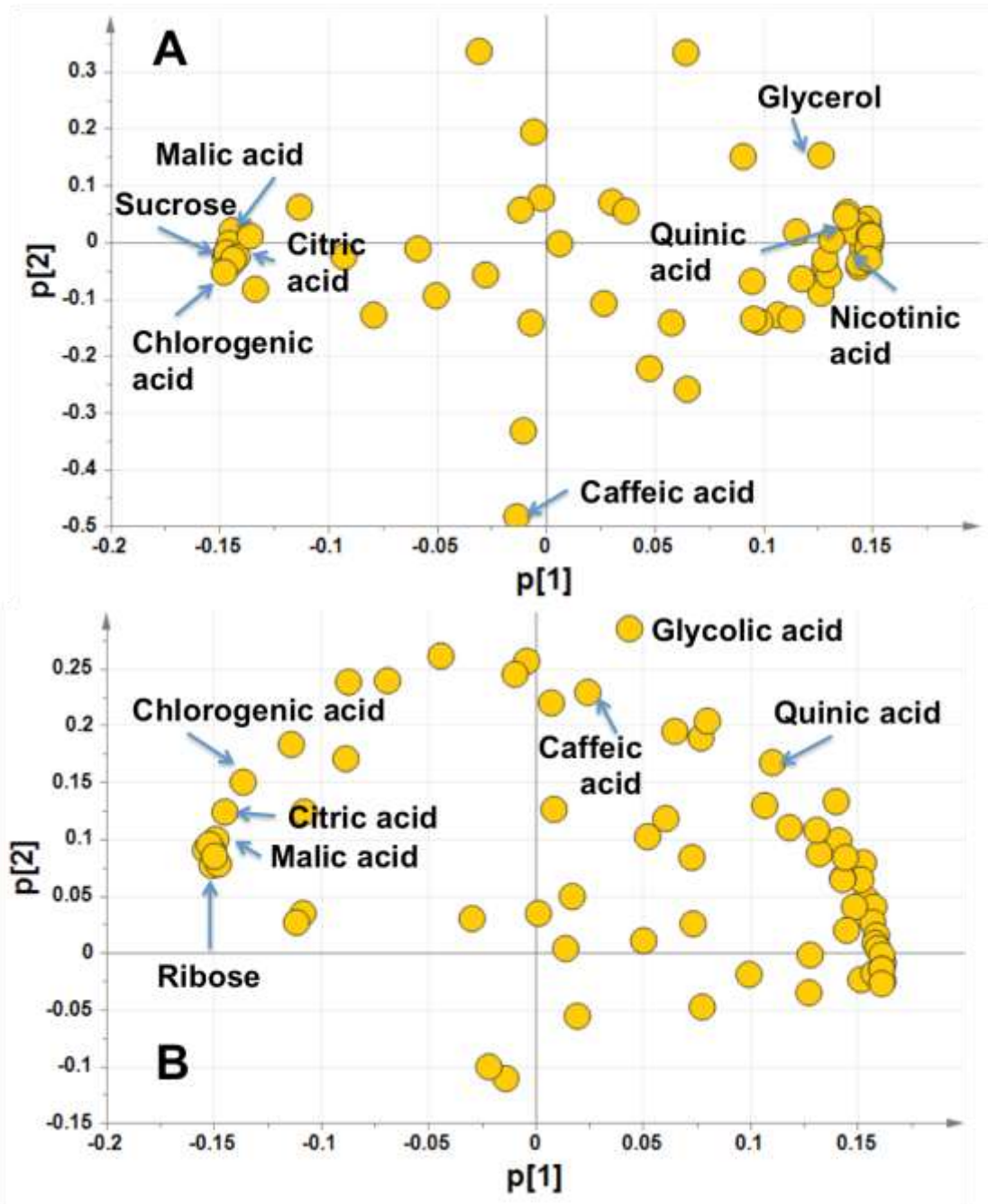
Supplementary Fig. 3-2. PCA loading plot of validation coffee set shown in Fig. 3-6. Red letters indicate variables important for constructing OPLS-DA model.



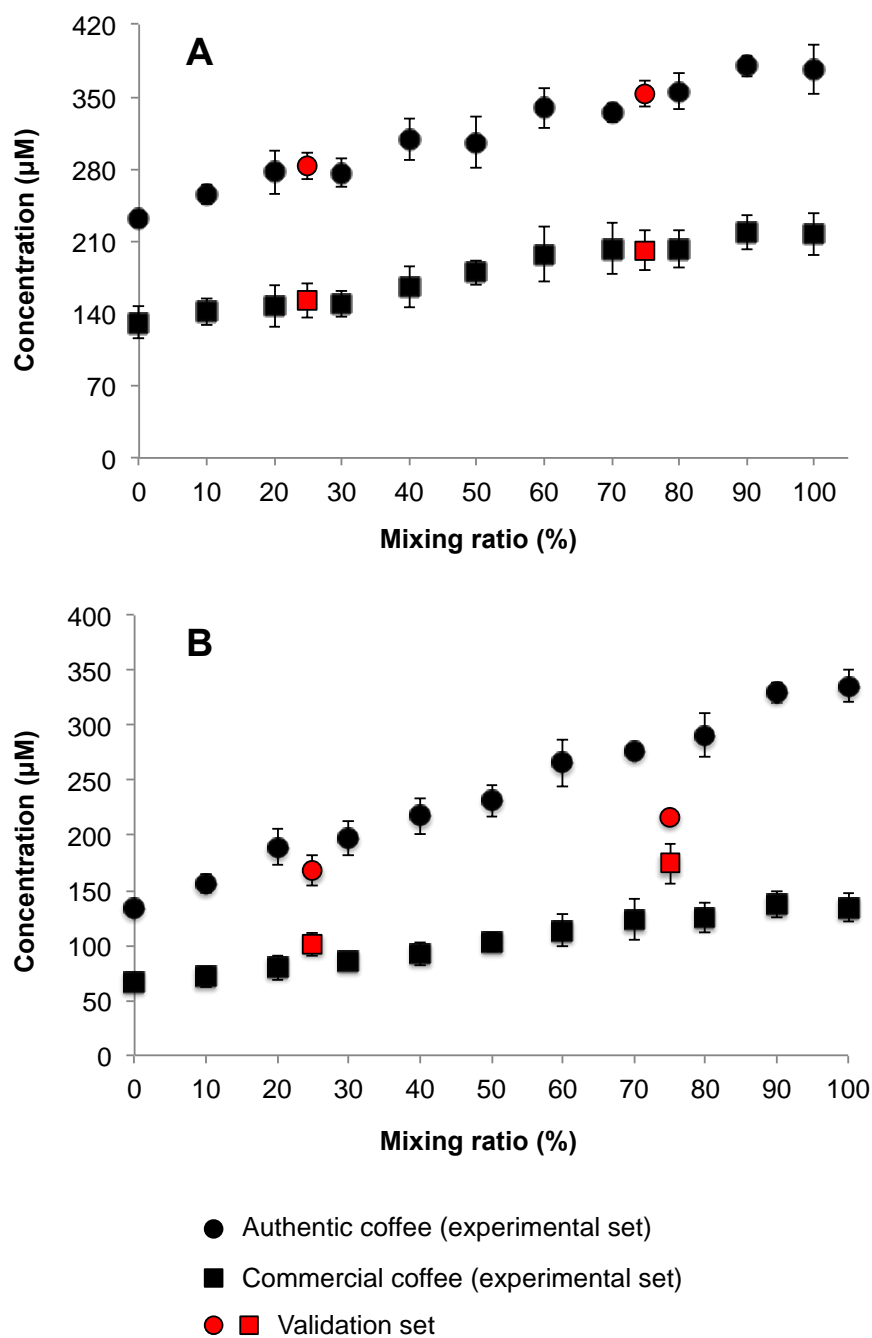
Supplementary Fig. 4-1. PCA loading plot derived from coffee samples with high- and low-mixing ratio, (A) authentic and (B) commercial coffee.



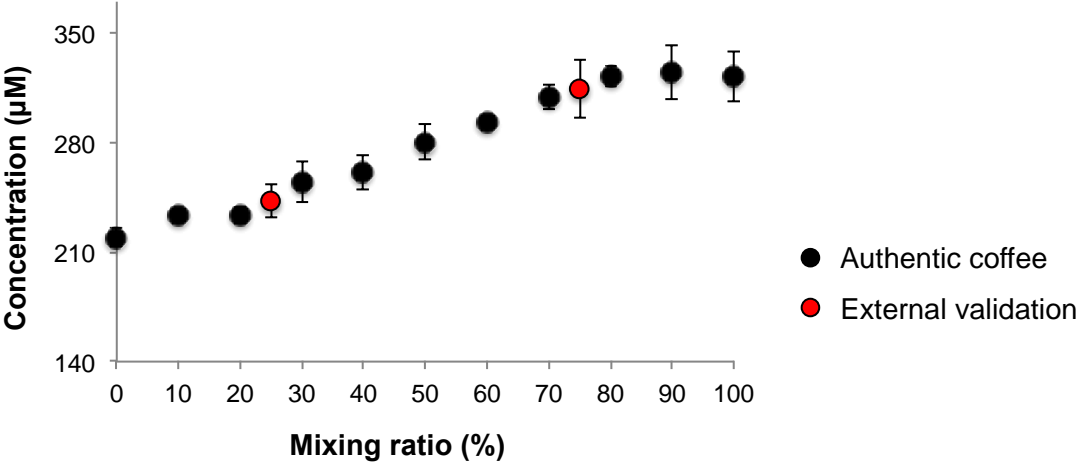
Supplementary Fig. 4-2. PCA loading plot derived from coffee samples with high- and low-mixing ratio after block-wise scaling, (A) authentic and (B) commercial coffee.



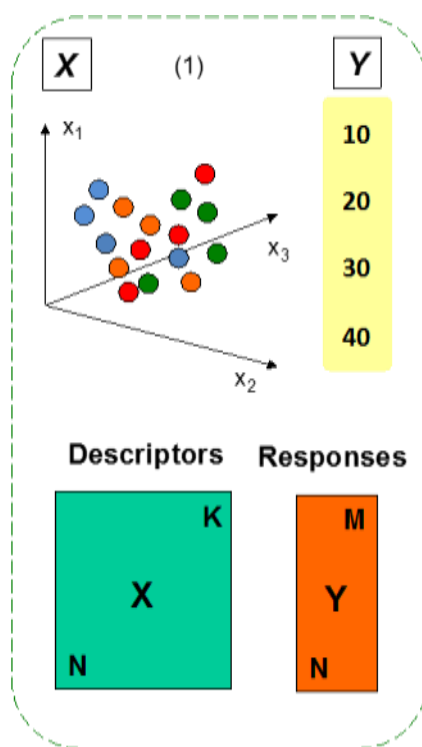
Supplementary Fig. 4-3. Concentration of citric acid (A) elevated proportionally with the increase of coffee mixing ratio; the level of the external validation samples fit the experimental coffee in both data sets. In contrary, malic acid is only feasible for quantification of coffee blends with ratio less than 25%.



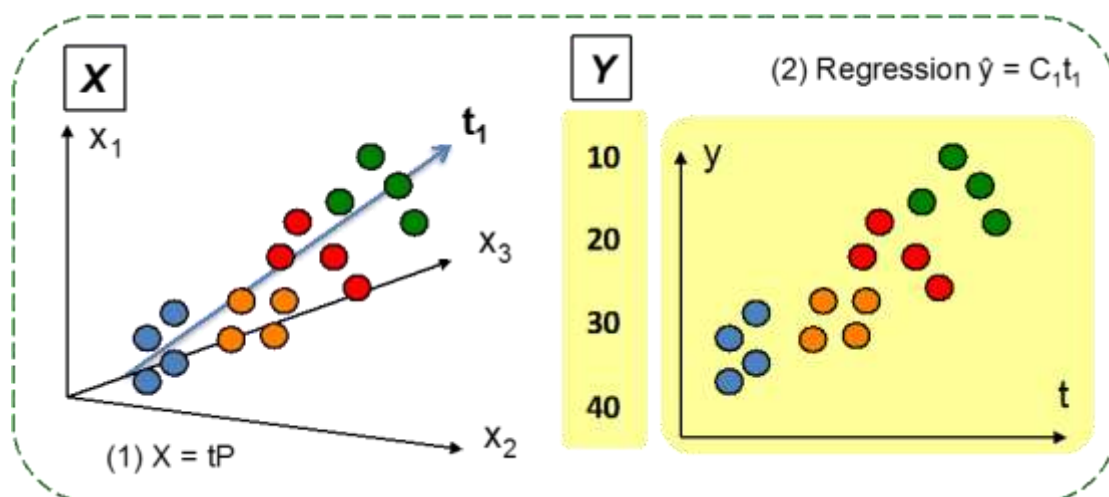
Supplementary Fig. 4-4. Reproducibility test of citric acid concentration in authentic civet coffee utilizing new calibration curve from intra-day experiment.



Supplementary Principle 1. Orthogonal partial to latent structures (OPLS) and orthogonal partial to latent structures-discriminant analysis (OPLS-DA)

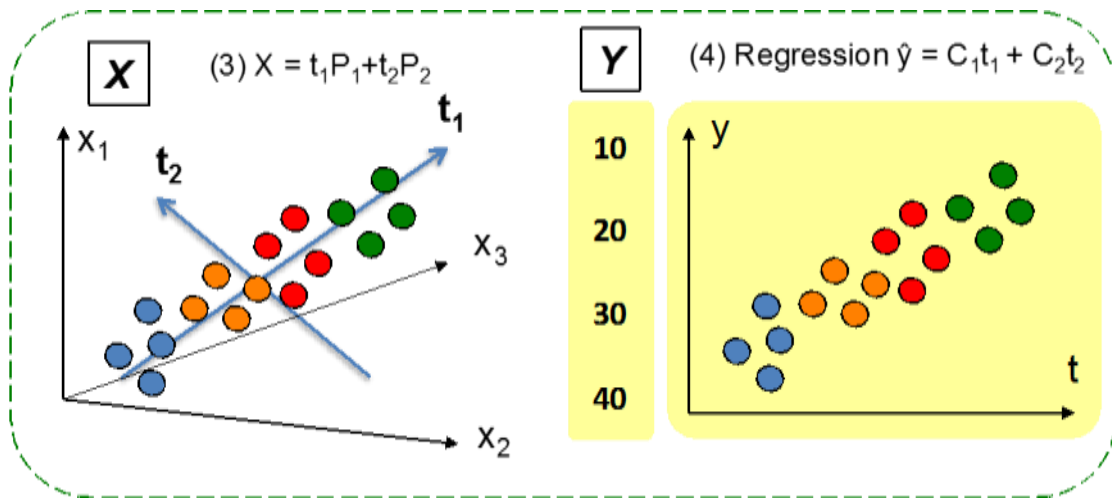


(1) Constructing X (descriptors or observations) matrix with N quantities (*i.e.*, samples) and K variables (*i.e.*, mass-to-charge ratio) and Y (responses) matrix (*i.e.*, coffee mixing ratio). Each variable in descriptors corresponds to 2 data point, X and Y .



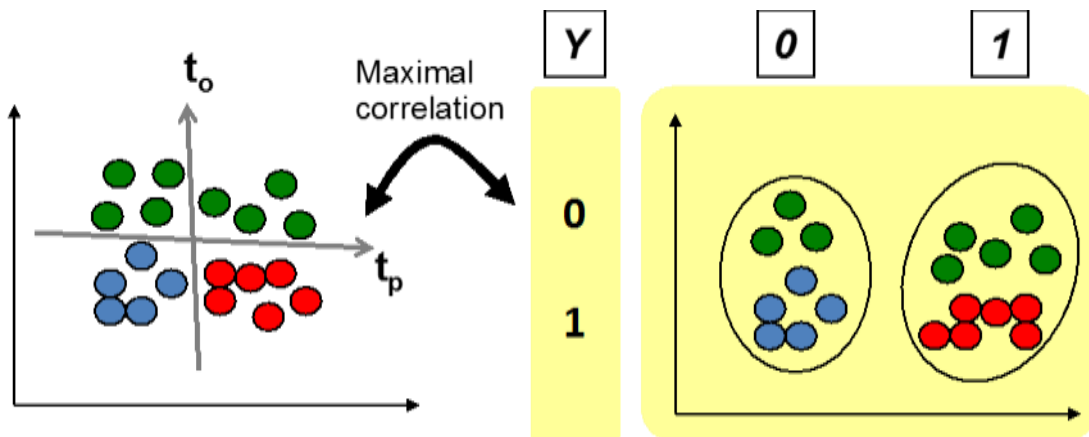
(2) The first component is computed. The first component of the model will orient itself so that it can describe the point in X -space while at the same time giving a maximal correlation with Y . The projections of the descriptors onto the line in the X -space give the score for each descriptor and summarize into score vector t_1 . P is the relationship

between X and t . When t is used to predict Y , regression equation is formed as $\hat{y} = C_1t_1$ with C represents the weight of the Y -vector.



3) The second component is then computed. Two components Y modeling is preferable than by one because the conformity between descriptors and estimated responses data is better with two components.

OPLS-DA is modification of OPLS modeling specifically in its Y variable. In OPLS-DA, Y is a binary vector with the value of 0 for civet coffee class and 1 for the regular coffee.



Geometrically, OPLS-DA separates or somehow rotates variations in X into two parts, one that is linearly correlated with Y (t_o) and one that is uncorrelated or orthogonal to Y (t_p). OPLS-DA will then finds variations or new axis in X dimension that is maximally correlated with Y .

List of Publications

Conferences

- 1) Udi Jumhawan, Sastia Prama Putri, Yusianto, Erly Marwani, Takeshi Bamba, Eiichiro Fukusaki. Discrimination of roasted Kopi Luwak and regular coffee based on geographical origins and variety using gas chromatography coupled with quadrupole mass spectrometry (Oral). The 2012 Annual Conference of the Japan Society for Bioscience, Biotechnology, and Agrochemistry. March 2012. Kyoto, Japan.
- 2) Udi Jumhawan, Sastia Prama Putri, Yusianto, Erly Marwani, Takeshi Bamba, Eiichiro Fukusaki. Validation of the applicability of discriminant marker for authenticity assessment of Luwak coffee (Oral). The 2013 Annual Conference of the Japan Society for Bioscience, Biotechnology, and Agrochemistry. March 2013. Sendai, Japan.
- 3) Udi Jumhawan, Sastia Prama Putri, Yusianto, Erly Marwani, Takeshi Bamba, Eiichiro Fukusaki. Identification and characterization of key chemical constituents for the authentication of Asian palm civet coffee (Kopi Luwak) using GC-Q/MS (Oral). Joint Symposium of Osaka University – RWTH Aachen. March 2013. Osaka, Japan.
- 4) Udi Jumhawan, Sastia Prama Putri, Yusianto, Takeshi Bamba, Eiichiro Fukusaki. Prediction of the coffee blend mixing ratio by GC-MS-based metabolomics and its application to the authentication protocol for Kopi Luwak (Poster). International Conference of Metabolomics Society. June 2014. Tsuruoka, Japan.
- 5) Udi Jumhawan, Sastia Prama Putri, Yusianto, Erly Marwani, Takeshi Bamba, Eiichiro Fukusaki. Metabolomics approach to select discriminant markers for authentication of Asian palm civet coffee (Kopi Luwak) (Poster). 248th American Chemical Society National Meeting. August 2014. San Francisco, USA.
- 6) Udi Jumhawan, Sastia Prama Putri, Yusianto, Takeshi Bamba, Eiichiro Fukusaki. Metabolomics application for rapid screening and authentication of Asian palm civet coffee (Kopi Luwak) (Poster). 250th American Chemical Society National Meeting. August 2015. Boston, USA.

Book chapter

- 1) Walter A. Lavina, Yusuke Fujieda, Udi Jumhawan, Sastia Prama Putri, and Eiichiro Fukusaki. Case Studies. In “Mass Spectrometry-Based Metabolomics: A Practical Guide”. CRC Press, Taylor and Francis, Boca Raton, USA. 2015.

Original papers

- 1) Udi Jumhawan, Sastia Prama Putri, Yusianto, Erly Marwani, Takeshi Bamba, Eiichiro Fukusaki. Selection of discriminant markers for authentication of Asian palm civet coffee (Kopi Luwak): a metabolomics approach. *Journal of Agricultural and Food Chemistry* **61**, 7994-8001 (2013) (*selected for press release*)
- 2) Udi Jumhawan, Sastia Prama Putri, Yusianto, Takeshi Bamba, Eiichiro Fukusaki. Application of gas chromatography/flame ionization detector-based metabolite fingerprinting for authentication of Asian palm civet coffee (Kopi Luwak). *Journal of Bioscience and Bioengineering* (2015) (In press, DOI: 10.1016/j.jbiosc.2015.03.005).
- 3) Sastia Prama Putri, Udi Jumhawan, Eiichiro Fukusaki. Application of GC/MS and GC/FID-based metabolomics for authentication of Asian palm civet coffee (Kopi Luwak). *Shimadzu Journal* (2015).
- 4) Udi Jumhawan, Sastia Prama Putri, Yusianto, Takeshi Bamba, Eiichiro Fukusaki.. Quantification of coffee blends for authentication of Asian palm civet coffee (Kopi Luwak) via metabolomics. (submitted).

Acknowledgements

First and Foremost, praises and thanks to Allah, the Almighty, for His showers of blessings and guidance throughout my life, including a chance for me to complete the research successfully.

I sincerely thank Prof. Eiichiro Fukusaki for giving me opportunity to pursue one of my countless dreams and for his supervision throughout this research. His dynamism and vision have taught me the real meaning of hard work, patience and being independent. I thank my thesis co-supervisor, Prof. Kazuhito Fujiyama and Prof. Toshiya Muranaka for their supportive comments and advices. I also express my thanks to Prof. Takeshi Bamba and Assoc. Prof. Shuichi Shimma for their valuable counsels. I am extending my heartfelt thanks to Assist. Prof. Hisayo Ono for her caring and humble assistance.

I would also like to express my gratitude to MEXT for providing me financial support for five years period of my study and researchers from ICCRI, Dr. Teguh Wahyudi, Dr. Surip Mawardi, Mr. Yusianto and Ms. Ariza Budi Tunjung Sari, for the samples and their collaborative works. To Dr. Sastia Putri and Mr. Danang Waluyo for guiding me to the colorful world of *Kopi Luwak*.

I am extremely thankful to my family: my parents and my brother, for their love, prayers, caring and sacrifices for my whole life. Their everlasting love means world to me. This remarkable journey would not have been possible without the support from my all lab mates, particularly Lee, Kobayashi-*senpai*, Sanda, Zana, Shao, Kakuta-*san*, Nagasawa-*san*, An, Risako, Shiga-*sensei*, Matsuo-*san*, Yang, Hanghang, Mega, Anja. Thank you for helping me to stay sane through these unbelievably challenging years. Finally, to all my Indonesian friends, especially *mas* Gagus, *mas* Andi, and *mas* Anto, for their companion along this long and winding road.

So, verily, with every hardship, there is ease.

Verily, with every difficulty, there is relief.

[Qur'an 94:6]