# IDENTIFICATION OF THE FLORAL SOURCE OF NEW ZEALAND HONEYS

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
**Master of Science in Chemistry**
at
**The University of Waikato**
by

**LAURA ELEANOR PETCHELL**

The University of Waikato
2009

# Abstract

Depending on the nectar source, honey is either unifloral (derived mostly from one plant type), or polyfloral (derived from multiple plant types). Unifloral honey has characteristic sensory properties, and is therefore of greater commercial value. Currently, identification of floral source involves pollen counting, a specialised and labour intensive process. The current research was aimed at developing an alternative, rapid, chemistry-based method of floral identification.

The aroma of honey depends on volatile compounds present; these may be derived from the plant from which nectar was taken. Therefore by identifying volatiles in honey it could be possible to identify floral source.

Solid-phase microextraction (SPME) is a technique that is useful for the headspace analysis of volatile compounds; when coupled with GC-MS it provides a powerful tool for fingerprinting volatiles in honey. GC-MS chromatograms of ten New Zealand unifloral honey types were obtained after headspace SPME extraction.

Statistical analysis of the GC-MS chromatographic data was used to discriminate between floral types. Probability plots were used to identify compounds indicative of floral source; this method discriminated between honey types with 90% success. Hierarchical cluster analysis and principal component analysis were used to study the structure of the data.

Learning algorithms in Weka (machine-learning software) were used to build models of data to classify honey types. The logistic model tree algorithm classified 89.8% of samples correctly. Such a model has the potential to be used to classify future honey samples, once further samples have been tested to validate the model.

# Acknowledgements

This study would not have been possible without the help of a number of people. I would like to thank them all, including: my supervisors, Dr. Merilyn Manley-Harris of The University of Waikato and Dr. Bruce Morris of Hill Laboratories for all their help and advice; Technology New Zealand, for their generous funding of this research; Airborne Honey, Waitemata Honey and Haddrell's of Cambridge for providing honey samples for testing; Ray Littler and Lyn Hunt of the University of Waikato for their help with the statistical analyses involved in this study; Tony Greaves of Hill Laboratories for his advice on Weka; Carol Goss, for helping me to get started on my research; the staff of Hill Laboratories' Food and Bioanalytical division for allowing me to undertake this research and for their collective help, and Rebecca Fitzgerald of Hill Laboratories and Terry Braggins, formerly of Hill Laboratories, who helped to set up this project in the first place. Finally, thank you to John Revell and both of our families for their love and support during my time at university.

# Table of Contents

Abstract ..... ii
Acknowledgements ..... iii
Table of Contents ..... iv
List of Figures ..... viii
List of Tables ..... x
List of Abbreviations ..... xi
List of Structures ..... xiii
List of Equations ..... xiv

**Chapter One: Introduction ..... 1**
1.1 Methods for Determining the Botanical Origin of Honey ..... 1
1.1.1 Pollen Analysis ..... 1
1.1.2 Physico-Chemical Properties ..... 2
1.1.3 Amino Acids ..... 3
1.1.4 Phenolic Compounds ..... 4
1.1.5 Trace Elements ..... 4
1.1.6 Organic Acids ..... 4
1.1.7 Combinations of Parameters ..... 5
1.2 Methods for Extracting Volatiles from Honey ..... 5
1.2.1 Simultaneous Distillation-Extraction (SDE) ..... 6
1.2.2 Solid-Phase Extraction (SPE) ..... 6
1.2.3 Ultrasound-Assisted Extraction (USE) ..... 7
1.2.4 Headspace Techniques ..... 7
1.2.5 Electronic Nose ..... 8
1.3 Solid-Phase Microextraction ..... 9
1.3.1 Introduction ..... 9
1.3.2 Coatings ..... 10
1.3.3 Desorption into GC ..... 11
1.4 Review of Previous Work ..... 12
1.5 References ..... 17

**Chapter Two: Experimental and Method Development ..... 21**
2.1 Experimental ..... 21

2.1.1 Honey Samples ....21
2.1.2 Standards ....21
2.1.3 Materials and Instrumentation....23
2.1.4 Procedure....24
2.1.5 Identification ....25
2.2 Method Optimisation ....27
2.2.1 Fibre Type ....27
2.2.2 Sample Preparation ....29
2.2.3 Extraction Temperature....31
2.2.4 Agitation....32
2.2.5 Equilibration Time ....33
2.2.6 Summary ....34
2.3 Data Treatment....34
2.4 References ....36

**Chapter Three: Probability Plots as a Tool for Determining Floral Source ..37**
3.1 How to Read Probability Plots ....37
3.2 Methodology ....38
3.3 Results and Discussion....39
3.3.1 Pohutukawa ....39
3.3.2 Thyme ....41
3.3.3 Manuka....43
3.3.4 Honeydew ....45
3.3.5 Rata ....47
3.3.6 Kamahi ....49
3.3.7 Viper's Bugloss....51
3.3.8 Clover....53
3.3.9 Tawari ....55
3.3.10 Rewarewa....57
3.4 Statistical Significance of Results....58
3.5 Summary ....60
3.6 References ....62

**Chapter Four: Unsupervised Learning**....67
4.1 Introduction ....67

4.2 HCA and PCA Performed on Entire Data Set .....68
4.2.1 Hierarchical Cluster Analysis .....68
4.2.2 Principal Component Analysis.....69
4.3 HCA and PCA Performed on Data Set Consisting Only of Selected Compounds .....72
4.3.1 Hierarchical Cluster Analysis .....73
4.3.2 Principal Component Analysis.....74
4.4 Discussion .....75
4.4.1 Comment on Similarity Between Manuka and Honeydew Honeys .....75
4.5 References .....75

**Chapter Five: Supervised Learning .....76**
5.1 Introduction .....76
5.2 Weka Terminology.....76
5.2.1 Input .....76
5.2.2 Evaluation .....77
5.2.3 Output.....77
5.3 Method .....79
5.3.1 Decision Trees.....80
5.3.2 Functions .....80
5.3.3 Lazy Classifiers .....80
5.3.4 Metalearning Algorithms .....81
5.4 Results .....81
5.5 Conclusions .....84
5.6 References .....85

**Chapter Six: Discussion and Conclusions.....86**
6.1 Summary of Findings .....86
6.2 Discussion .....88
6.3 Previous Studies .....89
6.4 Conclusions .....89
6.5 Suggestions for Future Work .....90
6.6 References .....90

Appendix 1: Chromatograms of Standards .....91
Appendix 2: Chromatograms of Clover Honey Samples.....97
Appendix 3: Chromatograms of Honeydew Honey Samples .....113
Appendix 4: Chromatograms of Kamahi Honey Samples.....122
Appendix 5: Chromatograms of Manuka Honey Samples .....129
Appendix 6: Chromatograms of Pohutukawa Honey Samples.....139
Appendix 7: Chromatograms of Rata Honey Samples .....147
Appendix 8: Chromatograms of Rewarewa Honey Samples.....155
Appendix 9: Chromatograms of Tawari Honey Samples .....162
Appendix 10: Chromatograms of Thyme Honey Samples .....174
Appendix 11: Chromatograms of Viper's Bugloss Honey Samples.....179
Appendix 12: Compounds of Interest in Honey Samples.....187
Appendix 13: Summary of Selected Learning Algorithm Outputs.....194

# List of Figures

Figure 1.1: SPME device .....9

Figure 2.1: Chromatogram of manuka honey comparing fibres .....28

Figure 2.2: Chromatogram of tawari honey comparing sample preparation methods .....29

Figure 2.3: Chromatogram of pohutukawa honey comparing extraction temperatures .....31

Figure 2.4: Chromatogram of manuka honey comparing the effect of agitation...32

Figure 2.5: Chromatogram of manuka honey comparing equilibration times.....33

Figure 2.6: Graph of peak area vs. equilibration time for *ortho*-methoxyacetophenone in manuka honey .....34

Figure 3.1: Probability plot of (*E*)-cinnamaldehyde .....37

Figure 3.2: Probability plots of 3-methylbut-2-enal, dimethyl sulfoxide and (*E*)-cinnamaldehyde .....39

Figure 3.3: Chromatogram of pohutukawa honey showing 3-methylbut-2-enal, dimethyl sulfoxide and (*E*)-cinnamaldehyde .....40

Figure 3.4: Probability plots for hexanoic acid and thymol.....41

Figure 3.5: Chromatogram of thyme honey showing hexanoic acid and thymol ..42

Figure 3.6: Probability plots for 2-methylbenzofuran, 1-(2-methoxyphenyl)ethanol, myrtenal and 1-phenylethanol.....43

Figure 3.7: Chromatogram of manuka honey showing 1-phenylethanol, 2-methylbenzofuran, myrtenal and 1-(2-methoxyphenyl)ethanol.....44

Figure 3.8: Probability plots for linalool, phenol and 1-(2-methoxyphenyl)ethanol .....45

Figure 3.9: Chromatogram of honeydew honey showing phenol, linalool and 1-(2-methoxyphenyl)ethanol.....46

Figure 3.10: Probability plots for dimethyl sulfide and dimethyl sulfoxide.....47

Figure 3.11: Chromatogram of rata honey showing dimethyl sulfide and dimethyl sulfoxide.....48

Figure 3.12: Probability plots for 4-methyl-*H*-furan-2-one and pantoyl lactone...49

Figure 3.13: Chromatogram of kamahi honey showing 4-methyl-5*H*-furan-2-one .....50

Figure 3.14: Probability plot for *p*-benzoquinone.....51

Figure 3.15: Chromatogram of viper's bugloss showing *p*-benzoquinone and pantoyl lactone .....52

Figure 3.16: Probability plots for 3-methylpentanoic acid and *ortho*-methoxyacetophenone.....53

Figure 3.17: Chromatogram of clover honey showing 3-methylpentanoic acid and *ortho*-methoxyacetophenone.....54

Figure 3.18: Probability plots for 2-ethylhexanoic acid and *ortho*-methoxyacetophenone.....55

Figure 3.19: Chromatogram of tawari honey showing 2-ethylhexanoic acid and *ortho*-methoxyacetophenone.....56

Figure 3.20: Chromatogram of rewarewa honey .....57

Figure 4.1: Dendrogram showing clusters of honey samples .....69

Figure 4.2: Score plot showing honey types plotted against first two principal components .....70

Figure 4.3: Score plot of honey types plotted against first two principal components. Rata, kamahi, tawari and rewarewa are labelled as bush honey.....71

Figure 4.4: Dendrogram showing clusters of observations.....73

Figure 4.5: Score plot showing honey types plotted against first two principal components .....74

Figure 5.1: Example of a confusion matrix.....77

Figure 5.2: Confusion matrix for LMT algorithm on manually selected data .....83

# List of Tables

Table 1.1: Characteristic Volatiles in Unifloral Honeys from Eastern Sicily.....12

Table 2.1: Samples Used in Experimental Work.....22

Table 2.2: Retention Times of Standards.....25

Table 2.3: Quality of Hit for Compounds Identified by Mass Spectra.....26

Table 2.4: Summary of Optimised Method Variables.....34

Table 4.1: Compounds Important in Discriminating Between Honey Types.....72

Table 5.1: Statistics Used to Evaluate the Performance of Learning Algorithms .78

Table 5.2: Summary of Algorithm Performance on Three Data Sets.....82

Table 5.3: Evaluation Statistics for LMT Algorithm on Manually Selected Data.84

Table 6.1: Comparison of Samples Correctly Classified by the Probability Plots Method and LMT Algorithm.....88

# List of Abbreviations

## Extraction Procedures

| | |
|---|---|
| **HD** | Hydrodistillation |
| **INDEX** | Inside Needle Dynamic Extraction |
| **MSDE** | Microsimultaneous Distillation – Solvent Extraction |
| **SDE** | Simultaneous Distillation – Extraction |
| **SPE** | Solid-Phase Extraction |
| **SPDE** | Solid-Phase Dynamic Extraction |
| **SPME** | Solid-Phase Microextraction |
| **SPACE** | Solid-Phase Aroma Concentrate Extraction |
| **SHS** | Static Headspace Analysis |
| **USE** | Ultrasound-Assisted Extraction |

## Instrumentation

| | |
|---|---|
| **GC** | Gas Chromatography |
| **GC-FID** | Gas Chromatography – Flame Ionisation Detector |
| **GC-MS** | Gas Chromatography – Mass Spectrometry |
| **HPLC** | High Performance Liquid Chromatography |
| **ICP-AES** | Inductively Coupled Plasma – Atomic Emission Spectroscopy |
| **ICP-MS** | Inductively Coupled Plasma – Mass Spectrometry |
| **LC-MS** | Liquid Chromatography – Mass Spectrometry |
| **TOFMS** | Time of Flight – Mass Spectrometry |
| **UV** | Ultraviolet |

## SPME Fibre Coatings

| | |
|---|---|
| **CAR** | Carboxen |
| **CW** | Carbowax |
| **DVB** | Divinylbenzene |
| **PA** | Polyacrylate |
| **PDMS** | Polydimethylsiloxane |

## Statistical terms

| | |
|---|---|
| **DFA** | Discriminating Factor Analysis |
| **HCA** | Hierarchical Cluster Analysis |
| **PCA** | Principal Component Analysis |
| **SDA** | Stepwise Discriminant Analysis |

## Weka Terminology

| | |
|---|---|
| **KNN** | K-nearest-neighbour |
| **LMT** | Logistic model tree |
| **ROC** | Receiver Operated Characteristic |
| **TP rate** | True Positive rate |
| **FP rate** | False Positive rate |

# List of Structures

Structure I: 3-Methylbut-2-enal ..... 40
Structure II: Dimethyl sulfoxide ..... 40
Structure III: (*E*)-Cinnamaldehyde ..... 40
Structure IV: Hexanoic acid ..... 41
Structure V: Thymol ..... 41
Structure VI: 2-Methylbenzofuran ..... 44
Structure VII: 1-(2-Methoxyphenyl)ethanol ..... 44
Structure VIII: Myrtenal ..... 44
Structure IX: 1-Phenylethanol ..... 44
Structure X: Linalool ..... 45
Structure XI: Phenol ..... 45
Structure XII: Dimethyl sulfide ..... 47
Structure XIII: Pantoyl lactone ..... 50
Structure XIV: 4-Methyl-5*H*-furan-2-one ..... 50
Structure XV: *p*-Benzoquinone ..... 51
Structure XVI: 3-Methylpentanoic acid ..... 54
Structure XVII: *ortho*- Methoxyacetophenone ..... 54
Structure XVIII: 2-Ethylhexanoic acid ..... 55

# List of Equations

Equation 3.1: t-test equation .....58

Equation 3.2: Standard deviation used in t-test.....58

Equation 4.1: Euclidean distance .....68

# Chapter One: Introduction

The aim of this research was to investigate the use of headspace solid-phase microextraction (SPME) coupled with GC-MS and statistical analysis as a potential tool for the rapid analysis of the botanical origin of New Zealand honeys.

In this chapter alternative methods for determining the botanical origin of honey are discussed; also methods for extracting volatiles from honey, and research that has previously been undertaken in this field.

## 1.1 Methods for Determining the Botanical Origin of Honey

Depending from which plant or plants nectar is taken, honey is either unifloral (derived from mostly one type of plant), or mixed flower honey. Unifloral honey has characteristic sensory properties, and is therefore of greater commercial value. Realistically honey cannot be derived from just one floral source. Because of this, unifloral honey is generally defined as having at least 45% pollen from one source.[1] The development of methods for determining the botanical origin of honey has been widely researched; some of these methods are detailed in the following sections.

### 1.1.1 Pollen Analysis

Pollen analysis (melissopalnyology) has been the traditional method for determining the floral source of honey. This involves examining pollen grains in honey through a microscope, and requires a highly trained analyst. As well as being a slow and tedious process, problems are numerous. Different plant species produce different proportions of pollen.[2] This leads to under-represented honeys such as New Zealand thyme honey (less than 20 000 pollen grains per 10 g sample), and over-represented honeys such as manuka honey (More than 100 000 pollen grains per 10 g sample).[2] Only 20% of thyme pollen grains are required to classify a thyme unifloral honey, whereas at least 70% of manuka pollen grains are needed in order to classify a manuka unifloral honey.[2]

More difficulties with pollen analysis are that pollen can be filtered out in the bee's honey sac, or for the purpose of packaging for sale; bees can take pollen without collecting nectar; pollen can be collected from plants which cannot be the sources of honey, and pollen may be added fraudulently.[1]

### 1.1.2 Physico-Chemical Properties

The physico-chemical properties of honey can also be used to determine botanical origin. Many parameters can be studied such as colour, electrical conductivity, sugars and acidity. Analysing combinations of these can allow classification of floral source, and determine whether honey has been adulterated with sweeteners such as sucrose or corn syrup.[1]

Sensory characteristics such as colour, taste and aroma are the most basic indicators of botanical origin. For example, acacia and citrus honeys are light in colour whereas honeydew honeys are dark.[3]

Electrical conductivity is a relatively simple test and correlates well with mineral content; it has been reported that it has replaced ash content in the international standards for honey.[3]

Because honey is a sugar solution (sugars make up approximately 95% of honey by dry weight[3]), it is optically active, that is – it can rotate the plane of polarized light. Some sugars such as fructose have a negative optical rotation, whereas others such as glucose have a positive optical rotation.[3] A study of Saudi honeys found that floral honeys tended to have negative optical rotation due to high fructose content, while honeydew honeys and adulterated honeys had a positive optical rotation.[4]

The principal sugars in honey are glucose and fructose. About 25 oligosaccharides are present in minor amounts.[1] Sugar composition can be determined by HPLC with refractometric detection, ion exchange chromatography with pulsed amperometric detection and GC-FID. Honeydew honeys contain more oligosaccharides than blossom honeys.[3] A study of the oligosaccharide composition of British honeys by high-performance anion-exchange liquid chromatography method with pulsed amperometric detection and Canonical

Discriminant Analysis concluded that this method would need to be used in conjunction with other techniques.[5] In general, the analysis of sugars is more suitable for the determination of adulteration such as the addition of sucrose, high fructose corn syrup or invert syrup to honey rather than assignment of floral source.[1]

All honeys are acidic and tend to have a pH value between 3.5 and 5.5, due to the presence of organic acids. The main acid is gluconic acid, which is in equilibrium with the respective glucono-lactone. Free acidity, total acidity and pH have some classification power for unifloral honeys but lactone content does not, as it is highly variable. Determining free acidity by titration is not reproducible due to lactone hydrolysis during titration.[3]

### 1.1.3 Amino Acids

Honey contains approximately 0.2% protein, which is of bee and plant origin.[1] The analysis of amino acid profiles seems to be more suitable for the determination of floral source than protein composition; however it must be used in conjunction with other techniques.[1] Proline is the main amino acid in honey. It is added to honey by the bee and is an indication of ripeness. Proline shows characteristic values in different unifloral honeys however it is not possible to classify unifloral honeys using just proline.[3]

Floral and honeydew Spanish honeys could be distinguished by analyzing the tryptophan and glutamic acid content.[6] HPLC amino acid analysis and principal component analysis were used on seven unifloral honeys.[7] Lavender honeys were successfully classified but others were not. GC and a combination of statistical techniques were used to study amino acid patterns in six unifloral honeys.[8] For some honeys the amino acid profile enabled differentiation of floral source, however no amino acids could be selected as markers.

HPLC analysis of Spanish unifloral honeys showed the main amino acids to be proline, phenylalanine, tyrosine, lysine, arginine, glutamic acid, histidine and valine.[9] Again, it was concluded that amino acid composition was not able to absolutely determine botanical origin.

### 1.1.4 Phenolic Compounds

Phenolic acids and polyphenols are plant-derived secondary metabolites. Dark coloured honeys contain more phenolic acid derivatives but less flavonoids than light coloured honeys.[3] A study of the phenolic profiles of European unifloral honeys using HPLC found that specific markers and characteristic UV spectra were obtained for lime-tree, eucalyptus, rapeseed, chestnut and heather honeys.[10] Hesperetin was confirmed as a marker for citrus honey, as well as kaempferol for rosemary honey and quercetin for sunflower honey. It has also been reported that the methods used for flavonoid analysis are very time consuming as different purification steps are necessary before instrumental analysis.[10]

The phenolic fraction of New Zealand manuka honey was investigated using HPLC.[11] It was found that geography did not influence phenolic composition, and that methyl syringate dominated the phenolic fraction and could be a possible marker. HPLC chromatograms of the phenolic fractions of manuka, heather, clover and beech honeydew honeys were significantly different.

### 1.1.5 Trace Elements

ICP-MS and ICP-AES have been used to measure trace elements in honey.[12] However minerals and trace elements tend to be better indicators of geographical origin rather than botanical origin, as they are affected by environmental pollution.[1, 3]

### 1.1.6 Organic Acids

Solid-phase extraction and HPLC were used to analyse organic acids in honeys of different botanical origin.[13] Several organic acids were identified: pyruvic, citric, galacturonic, gluconic, citramalic, glycolic, formic, acetic, butyric, tartaric, malonic, malic, quinic, fumaric, succinic, lactic and propionic. These acids were shown to occur in different proportions in unifloral honeys, depending on floral source.

32 aliphatic organic acids as their methyl esters have been identified in New Zealand rewarewa honeys using GC-MS.[14] 2-Methoxybutanedioic acid and 4-hydroxy-3-methyl-*trans*-2-pentenedioic acid were proposed as markers. It has

been reported that most acids in honey originate from bees, so acids may not be good markers for unifloral honeys.[3]

### 1.1.7 Combinations of Parameters

As mentioned earlier, combinations of physico-chemical parameters can be analysed in classification studies of honey. In a study of Polish honeys using three physico-chemical properties (total ash content, total acidity and dynamic viscosity), 98.67% correct classification was achieved on the 73 honey samples analysed.[15] Discriminant analysis was used to classify the honeys.

A study of Spanish honeys found that the most discriminant variables for differentiation between floral and honeydew honeys were colour, electrical conductivity, acidity, ash content and pH.[16] In a separate study of unifloral honeys, physico-chemical analyses were performed on honeys followed by principal component analysis and stepwise discriminant analysis.[17] 100% correct classification was achieved using conductivity, pH, free acidity and percentage of fructose, glucose and raffinose as variables.

pH, acidity, diastase, water content, conductivity, dextrose, fructose, colour, specific rotation, hydroxymethylfurfural (HMF) content and the difference in isotope ratio ($\delta\ ^{13}C/^{12}C$) were measured in Italian honeys of six floral origins.[18] Pattern recognition techniques were used to classify honeys successfully.

## 1.2 Methods for Extracting Volatiles from Honey

The aroma of honey depends on the volatile fraction, which originates mostly from the plant from which the honey was produced.[3] Therefore knowledge of the volatile composition should be useful in determining botanical origin.[1]

Methods used to extract volatiles from honey include simultaneous distillation-extraction, solid-phase extraction, ultrasound-assisted extraction, headspace techniques and the use of electronic noses.

### 1.2.1 Simultaneous Distillation-Extraction (SDE)

SDE was developed by Likens and Nickerson,[3] and has been subsequently modified for various purposes. SDE is used for isolating volatile compounds from a matrix.[1] A major problem with SDE is that heat treatment can lead to Maillard reaction products and the formation of other artefacts produced by heat. One modification of SDE involves doing the extraction under vacuum at room temperature. It has been shown that using heat produces the artefacts HMF and furfural, whereas these products are avoided when the extraction is done at room temperature under vacuum.[1] Another problem with SDE and other extractions that use organic solvents is that solvents can solubilise non-volatile compounds, which contaminate the GC injection port, and mask compounds of interest.[19]

It was found that dichloromethane extraction under an inert atmosphere followed by simultaneous steam distillation-dichloromethane extraction was an appropriate method for quantifying volatiles in honey.[20]

A similar method was used to obtain extracts of the volatiles from lime tree honey and chestnut honey.[21] Volatiles were analysed with GC-MS, and it was shown that these honeys can be characterized by phenols and benzene derivatives.

The volatile fraction of an aqueous solution of Haze honey was extracted using adsorptive column chromatography and SDE.[22] Alcohols, aldehydes, ketones, esters, acids, hydrocarbons, furanoids and pyranoids were identified.[22]

### 1.2.2 Solid-Phase Extraction (SPE)

SPE was used to extract the volatiles from eucalyptus honey.[23] Extractions were optimized using honey (20 g) eluted with dichloromethane (60 mL). This method allowed quantification of 35 volatile compounds, including terpenes and derivatives, furanoids and pyranoids, ketones, benzene derivatives, acids and norisoprenoids.

SDE, liquid-liquid extraction and SPE were used to extract the volatile fraction from rosemary honey, which was then analysed by GC-MS.[24] It was found that SDE was suitable for extracting terpenes and esters, however highly polar

compounds with low volatility such as acids, 2,3-butanediol and benzene derivatives had poor recoveries. Liquid-liquid extraction gave a satisfactory quantitative analysis – high yields and low standard deviations were achieved with this method. Esters were poorly extracted with SPE, and standard deviations were higher than for liquid-liquid extraction using SPE.

### 1.2.3 Ultrasound-Assisted Extraction (USE)

USE is a solvent extraction performed at room temperature; therefore artefacts produced by heat are avoided. When volatiles were extracted from citrus honey by USE, it was found that linalool derivatives predominated in the volatile extract.[25]

USE followed by GC-MS was used to analyse the volatiles from *Salvia officinalis* L. honey.[26] Of the 54 volatile compounds identified, the high percentage of benzoic acid and phenylacetic acid appeared to be useful in characterizing this type of honey.

The following extraction techniques have been compared: Hydrodistillation (HD), microsimultaneous distillation-solvent extraction (MSDE – a modification of SDE), USE and solid-phase microextraction (SPME – see **Section 1.3**).[27] It was found that SPME and USE generated the best results. HD and MSDE used heat which lead to the formation of artefacts and Maillard reaction products, and degraded sensitive compounds.

### 1.2.4 Headspace Techniques

Headspace techniques are promising because they don't use toxic and expensive organic solvents, and the need for solvent disposal is eliminated. Some headspace techniques include static headspace analysis, dynamic headspace purge-and-trap, solid-phase dynamic extraction (or inside needle dynamic extraction) and solid-phase microextraction.

Static headspace analyses (SHS) are not typically used for honey, because the volatiles in honey are present in low concentrations, and low recoveries are obtained by SHS for semi-volatile compounds.[19]

Dynamic headspace purge-and-trap has been used to extract honey volatiles.[28] Compounds were isolated in adsorbent resin and were thermally desorbed and concentrated in a cold trap. Aldehydes, ketones and alcohols were identified.

The volatiles of Sardinian strawberry-tree honey have been characterized using dynamic headspace extraction and GC-MS.[29] Norisoprenoid compounds such as α-isophorone, β-isophorone and 4-oxoisophorone were identified as floral origin markers.

Headspace solid-phase dynamic extraction (SPDE – otherwise known as INDEX – inside needle dynamic extraction) uses a hollow needle with an internal coating of a polymer. Volatile compounds are concentrated on the polymer by passing the gas through the device with repeated motions of the syringe plunger. This is a promising new technique but parameters optimizing these extractions need to be studied.[19] SPDE has been used for sampling food matrices.[30]

### 1.2.5 Electronic Nose

An electronic nose consists of an array of sensors, which produce a "fingerprint" characteristic of each sample when stimulated by its volatile fraction. The set of signals for all samples can be statistically processed.[3] This is a relatively new technique, and extensive studies are needed to confirm its reliability.[19]

In a study of Swiss unifloral honeys, botanical origin was classified using an electronic nose.[31] Results were compared with those produced by SPME, SHS and INDEX. 98% correct classification of model samples was achieved using the electronic nose in combination with SPME.

An electronic nose was tested on honey samples using two approaches.[32] The chromatogram approach treated the signal as a normal GC chromatogram. Relative peak areas were calculated and compared. Using the spectral approach, the whole aroma spectrum was considered. Data was analysed using principal component analysis and canonical discriminant analysis. The aroma of six honey varieties and two types of sugar solutions were discriminated, and 94% correct classification was achieved using a calibration model.

# 1.3 Solid-Phase Microextraction

## 1.3.1 Introduction

SPME is a relatively new technique that comprises sampling, sample preparation and instrument introduction steps.[33] Organic solvents are not needed for SPME, which is a huge advantage in terms of their cost, toxicity and disposal requirements.[33]

The SPME device consists of a syringe. Inside the needle is a stainless steel sheath, which can be exposed. Inside this is a silica capillary coated in a stationary phase, which can also be exposed. A diagram is shown in **Figure 1.1**.

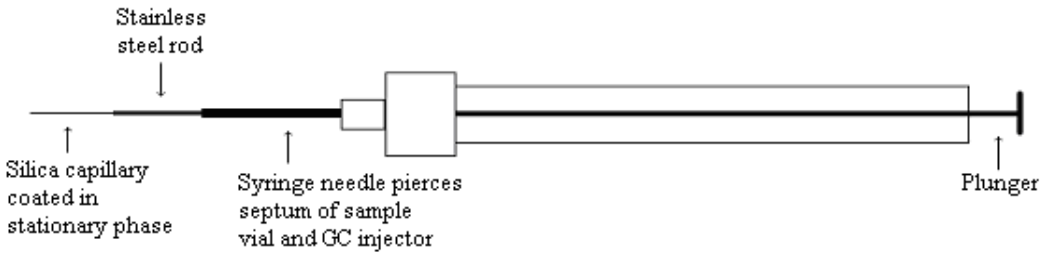



**Figure 1.1: SPME device[34]**

The basic SPME process is comprised of two steps:

1. Partitioning of analytes to the stationary phase. The fibre is exposed to the sample matrix or headspace.
2. Desorption of analytes into the instrument. Expansion of air caused by temperature increase in the instrument allows removal of desorbed analytes from the stationary phase.[33]

Once analytes have been desorbed into the instrument, the fibre is clean and ready for re-use. Fibres can be re-used hundreds of times for headspace analyses.[35]

There are three types of extractions that can be performed:

- Direct extraction – the stationary phase is immersed in the sample matrix
- Headspace sampling – protects the fibre from high molecular weight and other non-volatile compounds present in the sample matrix. Headspace sampling is appropriate for the extraction of honey volatiles.
- Membrane protection – for liquid sampling.[33]

SPME is a highly versatile technique. It can be applied to polar and non-polar compounds in gas, liquid and solid samples, and can couple to GC, GC-MS, HPLC and LC-MS. The use of SPME in environmental, food, flavour, fragrance, pheromone, pharmaceutical, clinical and forensic applications has been reviewed.[19]

Certain parameters are important in optimizing the performance of SPME, such as agitation, temperature, fibre coatings, headspace volume, salt concentration, pH, sample volume and desorption conditions. These parameters should also be kept constant for reproducible analyses, and will now be discussed in further detail.

### 1.3.2 Coatings

The silica fibre in a SPME device is coated with a thin polymeric film which concentrates analytes.[35] Better extractions are achieved when the polarity of the fibre matches the polarity of the analyte.[19] For example, polydimethylsiloxane (PDMS) is a non-polar coating therefore it extracts non-polar analytes well. It is a robust liquid coating and can stand injector temperatures of up to 300 °C.[33]

Both PDMS and polyacrylate (PA) coatings extract analytes via absorption – analytes dissolve and diffuse into the stationary phase.[35] Coatings such as carbowax-divinylbenzene (CW/DVB), PDMS/carboxen (PDMS/CAR) and PDMS/DVB extract via adsorption of analytes.[35]

Before a new fibre is used it should be conditioned. This can be done by exposing it to its maximum desorption temperature for 0.5 to 4 hours prior to use.[35] The thinnest coating which achieves the required detection limits should be used, as this influences time and sensitivity.[33]

The following parameters apply to extraction of analytes from the sample:

- Temperature

  Increasing temperature decreases the extraction time, but also decreases recovery. The highest temperature that does not decrease recovery should be used to increase analyte concentration in the headspace.[33]

- Agitation

  In some cases, agitation can decrease equilibration time and allow a faster extraction of less volatile species. Usually a magnetic stirrer or sonicator is used.[3]

- Salt concentration

  Adding soluble salts such as sodium chloride can improve extraction efficiency, and decrease the solubility of organic hydrophilic compounds in the aqueous phase.[19]

- pH

  The form analytes take in the sample depends on the pH of the matrix relative to the analyte, and influences extraction efficiency. Use of a non-volatile acid or base is desirable for headspace SPME.[19]

- Volume

  Headspace volume should be minimized, as very volatile compounds will accumulate in the headspace rather than on the fibre when the headspace is large. However, minimizing the headspace is limited by the length of the fibre.[33]

Vial size and sample volume should be kept constant. It has been reported that equilibrium is reached three times more quickly if 1 mL of liquid is placed in a 5 mL vial than if 10 mL of liquid is placed in a 50 mL vial.[19]

### 1.3.3 Desorption into GC

Desorption depends on analyte volatility, the thickness of the fibre coating, and injection depth, temperature and exposure time. Generally the optimum desorption temperature is equal to the boiling point of the least volatile analyte.[19] Injections should be splitless. This is because desorption is rapid, and a loss of analytes may occur if injections are split.[33] A narrow-bore GC injector insert is

required to ensure a high linear flow.[35] A high linear flow is desirable so that analytes are quickly removed from the coating, preventing further interactions that may slow down desorption. The fibre should be exposed immediately after the needle is introduced into the insert, and the needle depth should be adjusted to place the fibre in the centre of the hot injector zone.[35]

## 1.4 Review of Previous Work

Several papers have been published regarding the use of solid-phase microextraction of honey volatiles for identification of floral origin, which are subsequently reviewed.

A study of unifloral honeys from Eastern Sicily aimed to find volatile marker compounds indicative of floral source.[36] Characteristic volatile compositions were obtained for each type of honey, summarized in **Table 1.1**:

**Table 1.1: Characteristic Volatiles in Unifloral Honeys from Eastern Sicily[36]**

| **Honey** | **Characteristic volatiles** |
|---|---|
| Eucalyptus | 5-Hexen-2-ol and 2,3-dimethyl-5-hexen-2-ol |
| Orange | Hortrienol, methyl anthranilate |
| Wildflower | 3-Methylbutanoic acid, myrcenol, *cis*-carveol |
| Sulla | Terpinen-4-ol and α-terpineol |
| Chestnut | Benzaldehyde, camphor, acetophenone |

Compounds such as nonanol, nonanal, nonanoic acid, hexanol, linalol and hexanoic acid were found in relatively large amounts in multiple honeys.

Liquid sampling was compared with headspace sampling. It was found that although a larger amount of extracted compounds was obtained with liquid sampling, the method lacked repeatability since the sugars were adsorbed onto the fibre. Another disadvantage of liquid sampling was that no more than 3 analyses were possible with the same fibre.[36]

In a study of Spanish honeys, CAR/PDMS and PDMS/DVB fibres were compared for SPME-GC-MS analyses.[37] It was found that CAR/PDMS fibres gave

preferable results. Higher numbers and concentrations of honey volatiles were obtained with this fibre type, mainly volatiles with retention times of 0 – 4 minutes.

Extraction temperatures of 50 °C, 70 °C and 85 °C were compared. It was found that more chromatographic signals were obtained at 70 °C than at 50 °C. At 85 °C a high signal of HMF was detected, indicating fructose decomposition, hence an extraction temperature of 70 °C was used.

Using several statistical techniques, all honey types (orange, lavender, eucalyptus, rosemary and thyme) were able to be distinguished. The Kolmogorov-Smirnov test was used to evaluate normality. Combined with the result of the M. De Box test, it was found that no data transformation was needed. Canonical discriminant and stepwise discriminant analyses were performed in order to establish characteristic variables. Canonical discriminant analysis showed that the first four functions provided good discrimination. Finally, using a Jack-knifed classification matrix, all the honeys were correctly classified.[37]

Italian citrus, chestnut, eucalyptus, lime tree, thyme and dandelion honeys were analysed by SPME-GC-MS and SPME-GC-FID.[38] PDMS, PA and carboxen fibres were compared. The PDMS fibre gave poor results in the first part of the chromatogram and the PA fibre adsorbed more volatiles than the PDMS fibre. Peak distortion was observed with the carboxen fibre, possibly due to high desorption resulting from the high polarity of this phase. The PA fibre was used, as it gave the best results overall.

GC-MS was used to identify honey volatiles. Particular features of the spectra such as peak intensity and distribution were observed for each honey type. GC-FID was used for quantitative analysis. High variability between honey samples from the same floral source was observed, and attributed to quantitative limitations of SPME, or the complexity of unifloral honeys.[38]

A SPME-GC-MS method was developed and optimized to characterize Brassica honeys.[39] The optimized method can be summarized as follows: Honey (4 g) was added to a vial (10 mL). NaCl was added (1.05 g), and the water content was

adjusted to 36.0 g/100 g. The vial was incubated in a water bath (30 min, 70 °C) under constant stirring at 1100 rpm. A 50/30 μm DVB/CAR/PDMS fibre was exposed to the vial headspace (35 min) while keeping vial temperature and stirring constant, and the fibre was desorbed into the GC injector (10 mins). A splitless injection was used.

It was investigated whether two varieties of Brassica honey (winter rape and spring turnip rape) could be differentiated by SPME. Using SPME-GC-FID and a combination of statistical tests, significant differences were found in the concentrations of 2-furancarboxaldehyde, benzaldehyde, benzeneacetaldehyde, benzyl alcohol, benzeneethanol and benzeneacetonitrile in the two honeys at the 95% confidence level. Analysis by SPME-GC-MS showed that octanal, pentadecane, α-terpineol and 5-methyl-2-(1-methylethyl)-phenol were only present in spring turnip rape honey, and 4-methyl-phenol and 1,4-dichloro-benzene were only found in winter rape honey (however the latter was a contaminant).[39]

A study was undertaken in which five unifloral honeys, and mixed-flower honeys were analysed by SPME-GC-MS.[40] Four fibre types were tested (PDMS, PA, CW/DVB and C/PDMS where C = carbon molecular sieve). The best results were obtained using PA and C/PDMS fibres. Operating conditions were optimized for these two fibres, and both fibre types led to similar optimized experimental conditions. Honey samples were analysed by this method, and principal component analysis was used to discriminate between samples.[40]

Swiss honeys (acacia, chestnut, dandelion, lime, rape and fir) were analysed with an electronic MS-based nose, static headspace (SHS), SPME and Inside Needle Dynamic Extraction (INDEX) sampling modes.[31] The results for each extraction method were compared. SPME gave the best results, as it was able to extract heavier volatiles, and at a higher concentration than the other two methods. SHS was satisfactorily able to discriminate lime and dandelion honeys, and as this is a faster technique, only acacia, chestnut, rape and fir honey samples were analysed using SPME. Principal component (PCA) and discriminating factor (DFA) analyses were performed on the results obtained by SPME-electronic nose, and a 98% correct classification model was obtained.[31] DFA usually affords a better

classification rate than PCA. This is because class membership is applied, so a model is built that tries to discriminate between honey groups. Caution should be exercised with DFA as models can be over-fitted.[31]

Mixed flower honeys from Galacia were analysed by SPME-GC-MS.[41] Method optimization was studied with respect to headspace sampling compared with direct immersion sampling. It was found that extracted compounds showed greater peak areas in the direct immersion technique; however headspace sampling has the advantage of avoiding contamination and increasing the fibre lifetime. The optimized method proved to be good for the analysis of monoterpenes, and the test set showed high precision and repeatability.[41]

USE and SPME were compared for the analysis of Greek cotton honey.[42] More than 35 phenolic compounds were detected using the USE method that could serve as markers. Fewer compounds were identified in SPME. Benzenepropanol and (*E*)-Cinnamaldehyde were detected by a DVB/CAR/PDMS fibre as possible floral markers.[42]

Spanish eucalyptus, rosemary, heather and citrus unifloral honeys were analysed using SPME-GC-MS and a CAR/PDMS fibre.[43] Samples were qualitatively and quantitatively evaluated, and multivariate statistical analyses were performed. Acetoin and diacetyl were found in eucalyptus honey in high concentrations, which have previously been reported as markers for eucalyptus honeys. Rosemary honey gave a characteristic composition including alcohols such as 3-methyl-3-buten-1-ol and 3-methyl-1-butanol. Compounds of low volatility were more abundant in heather honey than in other types, and furfuryl alcohol, benzyl alcohol and 2-phenylethanol were prominent. Citrus honey was characterized by four lilac aldehydes. Statistical models yielded good results for eucalyptus and citrus honeys; however rosemary and heather were more difficult to characterize.[43]

Alfalfa, sunflower, white clover, carob and caldén unifloral honeys were analysed using SPME-GC-MS and a range of statistical techniques.[44] The aim of using such chemometric techniques was to select a minimum number of compounds

which were able to give a correct classification. Using a minimum number of compounds eliminated irrelevant information.

Using hierarchical cluster analysis (HCA), volatile organic compounds (VOCs) were clustered by similarities among them, not by floral source. This helped to identify compounds that could help explain differences between honey samples based on their VOC profile. Stepwise discriminant analysis (SDA) was used to determine which compounds were closely associated with floral origin.

Combining results from HCA and SDA, six volatiles were selected as being the most appropriate to differentiate between honeys of varying floral origin. These were octanal, benzeneacetaldehyde, 1-octanol, 2-methoxyphenol, nonanal and 2-*H*-1-benzopyran-2-one. 93% correct classification was achieved when *K*-nearest neighbor (KNN) classification was applied to these six VOCs. This study showed that better results were obtained by analysing VOC profiles as a whole instead of looking for particular markers for each floral origin.[44]

Unifloral Greek and Italian citrus honeys were analysed by SPME-GC-MS and a DVB/CAR/PDMS fibre.[45] Lilac aldehydes proved to be powerful markers, followed by dill ether, methyl anthranilate and the third and fourth isomers of 1-*p*-menthen-9-al. Greek and Italian honeys could not be discriminated by this method.[45]

The volatile composition of thyme honey was studied using SPME-GC-MS.[46] Volatiles were extracted using a 100 µm PDMS fibre. Honeys were analysed by GC-MS and descriptive flavor profile analysis. High amounts of thymol and carvacrole indicated adulteration by thyme essential oil. 3,4,5-Trimethoxybenzaldehyde was found only in pure thyme honey, but more samples would need to be analysed to verify this as a marker.[46]

# 1.5 References


1. Anklam, E., A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chemistry* **1998,** 63, (4), 549-562.
2. Moar, N. T., Pollen analysis of New Zealand honey. *Journal of Agricultural research* **1985,** 28, 39-70.
3. Bogdanov, S.; Ruoff, K.; Persano Oddo, L., Physico-chemical methods for the characterisation of unifloral honeys: a review. *Apidologie* **2004,** 35, 4-17.
4. Al-Khalifa, A. S.; Al-Arify, I. A., Physicochemical characteristics and pollen spectrum of some Saudi honeys. *Food Chemistry* **1999,** 67, (1), 21-25.
5. Goodall, I.; Dennis, M. J.; Parker, I.; Sharman, M., Contribution of high-performance liquid chromatographic analysis of carbohydrates to authenticity testing of honey. *Journal of Chromatography A International Ion Chromatographic Symposium 1994* **1995,** 706, (1-2), 353-359.
6. Iglesias, M. T.; de Lorenzo, C.; Polo, M. C.; Martin-Alvarez, P. J.; Pueyo, E., Usefulness of amino acid composition to discriminate between honeydew and floral honeys. Application to honeys from a small geographic Area. *Journal of Agricultural and Food Chemistry* **2004,** 52, (1), 84-89.
7. Cotte, J. F.; Casabianca, H.; Giroud, B.; Albert, M.; Lheritier, J.; Grenier-Loustalot, M. F., Characterization of honey amino acid profiles using high-pressure liquid chromatography to control authenticity. *Analytical & Bioanalytical Chemistry J1 - Analytical & Bioanalytical Chemistry* **2004,** 378, (5), 1342-1350.
8. Pirini, A.; Conte, L. S.; Francioso, O.; Lercker, G., Capillary gas chromatographic determination of free amino acids in honey as a means of discrimination between different botanical sources. *Journal of High Resolution Chromatography* **1992,** 15, 165-170.
9. Hermosin, I.; Chicon, R. M.; Cabezudo, M. D., Free amino acid composition and botanical origin of honey. *Food Chemistry* **2003,** 83, (2), 263-268.
10. Tomás-Barberán, F. A.; Martos, I.; Ferreres, F.; Radovic, B. S.; Anklam, E., HPLC flavonoid profiles as markers for the botanical origin of European unifloral honeys. *Journal of the Science of Food and Agriculture* **2001,** 81, (5), 485-496.
11. Weston, R. J.; Brocklebank, L. K.; Lu, Y., Identification and quantitative levels of antibacterial components of some New Zealand honeys. *Food Chemistry* **2000,** 70, 427-435.
12. Caroli, S.; Forte, G.; Iamiceli, A. L.; Galoppi, B., Determination of essential and potentially toxic trace elements in honey by inductively coupled plasma-based techniques. *Talanta* **1999,** 50, (2), 327-336.
13. del Nozal, M. J.; Bernal, J. L.; Marinero, P.; Diego, J. C.; Frechilla, J. I.; Higes, M.; Llorente, J., High performance liquid chromatographic determination of organic acids in honeys from different botanical origin. *Journal of Liquid Chromatography & Related Technologies* **1998,** 21, (20), 3197 - 3214.

14. Wilkins, A. L.; Lu, Y.; Tan, S., Extractives from New Zealand honeys. 5. Aliphatic dicarboxylic acids in New Zealand Rewarewa (*Knightea excelsa*) honey. *Journal of Agricultural and Food Chemistry* **1995,** 43, 3021-3025.
15. Popek, S., A procedure to identify a honey type. *Food Chemistry* **2002,** 79, 401-406.
16. Soria, A. C.; Gonzalez, M.; de Lorenzo, C.; Martínez-Castro, I.; Sanz, J., Characterization of artisanal honeys from Madrid (Central Spain) on the basis of their melissopalynological, physiochemical and volatile composition data. *Food Chemistry* **2004,** 85, 121-130.
17. Devillers, J.; Morlot, M.; Pham-Delegue, M. H.; Dore, J. C., Classification of monofloral honeys based on their quality control data. *Food Chemistry* **2004,** 86, (2), 305-312.
18. Marini, F.; Magri, A. L.; Balestrieri, E.; Fabretti, F.; Marini, D., Supervised pattern recognition applied to the discrimination of the floral origin of six types of Italian honey samples. *Analytica Chimica Acta* **2004,** 515, (1), 117-125.
19. Cuevas-Glory, L. F.; Pino, J. A.; Santiago, L. S.; Sauri-Duch, E., A review of volatile analytical methods for determining the botanical origin of honey. *Food Chemistry* **2007,** 103, (3), 1032-1043.
20. Bouseta, A.; Collin, S., Optimized Likens-Nickerson methodology for quantifying honey flavors. *Journal of Agricultural and Food Chemistry* **1995,** 43, (7), 1890-1897.
21. Guyot, C.; Bouseta, A.; Scheirman, V.; Collin, S., Floral origin markers of chestnut and lime tree honeys. *Journal of Agricultural and Food Chemistry* **1998,** 46, (2), 625-633.
22. Shimoda, M.; Wu, Y.; Osajima, Y., Aroma compounds from aqueous solution of haze (*Rhus succedanea*) honey determined by adsorptive column chromatography. *Journal of Agricultural and Food Chemistry* **1996,** 44, (12), 3913-3918.
23. Vazquez, L. C.; Diaz-Maroto, M. C.; Guchu, E.; Perez-Coello, M. S., Analysis of volatile compounds of eucalyptus honey by solid-phase extraction followed by gas chromatography coupled to mass spectrometry. *European Food Research and Technology* **2006,** 224, (1), 27-31.
24. Castro-Vazquez, L.; Perez-Coello, M. S.; Cabezudo, M. D., Analysis of volatile compounds of rosemary honey. Comparison of different extraction techniques. *Chromatographia* **2003,** 57, (3-4), 227-233.
25. Alissandrakis, E.; Daferera, D.; Tarantilis, P. A.; Polissiou, M.; Harizanis, P. C., Ultrasound-assisted extraction of volatile compounds from citrus flowers and citrus honey. *Food Chemistry* **2003,** 82, (4), 575-582.
26. Jerkovic, I.; Mastelic, J.; Marijanovic, Z., A variety of volatile compounds as markers in unifloral honey from Dalmatian sage (Salvia officinalis L.). *Chemistry & Biodiversity* **2006,** 3, (12), 1307-1316.
27. Alissandrakis, E.; Tarantilis, P.; Harizanis, P.; Polissiou, M., Evaluation of four isolation techniques for honey aroma compounds. *Journal of the Science of Food and Agriculture* **2005,** 85, (1), 91-97.
28. Radovic, B. S.; Careri, M.; Mangia, A.; Musci, M.; Gerboles, M.; Anklam, E., Contribution of dynamic headspace GC-MS analysis of aroma compounds to authenticity testing of honey. *Food Chemistry* **2001,** 72, 511-520.

29. Bianchi, F.; Careri, M.; Musci, M., Volatile norisoprenoids as markers of botanical origin of Sardinian strawberry-tree (*Arbutus unedo* L.) honey: Characterisation of aroma compounds by dynamic headspace extraction and gas chromatography-mass spectrometry. *Food Chemistry* **2005,** 89, 527-532.
30. Bicchi, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B., Automated headspace solid-phase dynamic extraction to analyse the volatile fraction of food matrices. *Journal of Chromatography A* **2004,** 1024, (1-2), 217-226.
31. Ampuero, S.; Bogdanov, S.; Bosset, J.-O., Classification of unifloral honeys with an MS-based electronic nose using different sampling modes: SHS, SPME and INDEX. *European Food Research and Technology* **2004,** 218, 198-207.
32. Lammertyn, J.; Veraverbeke, E. A.; Irudayaraj, J., zNose(TM) technology for the classification of honey based on rapid aroma profiling. *Sensors and Actuators B: Chemical* **2004,** 98, (1), 54-62.
33. Pawliszyn, J., *Solid Phase Microextraction: Theory and Practice*. Wiley-VCH: New York, 1997.
34. Mendham, J.; Denney, R. C.; Barnes, J. D.; Thomas, M. J. K., *Vogel's Textbook of Quantitative Chemical Analysis*. 6th ed.; Pearson Educated Limited: Essex, 2000.
35. Vas, G.; Vekey, K., Solid-phase microextraction: a powerful sample preparation tool prior to mass spectrometric analysis. *Journal of Mass Spectrometry* **2004,** 39, (3), 233-254.
36. Verzera, A.; Campisi, S.; Zappala, M.; Bonaccorsi, I., SPME-GC-MS analysis of honey volatile components for the characterization of different floral origin. *American Laboratory* **2001**, 18-21.
37. Perez, R. A.; Sanchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L., Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography-mass spectrometry. *Journal of Agricultural and Food Chemistry* **2002,** 50, (9), 2633-2637.
38. Piasenzotto, L.; Gracco, L.; Conte, L., Solid phase microextraction (SPME) applied to honey quality control. *Journal of the Science of Food and Agriculture* **2003,** 83, (10), 1037-1044.
39. Ruoff, K. Solid-Phase Microextraction of Honey Volatiles: A Method for the Determination of the Botanical Origin of Honey. University of Helsinki, 2003.
40. Soria, A.; Martínez-Castro, I.; Sanz, J., Analysis of volatile composition of honey by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2003,** 26, (9-10), 793-801.
41. Peña, R.; Barciela, J.; Herrero, C.; García-Martín, S., Solid-phase microextraction gas chromatography-mass spectrometry determination of monoterpenes in honey. *Journal of Separation Science* **2004,** 27, (17-18), 1540-1544.
42. Alissandrakis, E.; Kibaris, C.; Tarantilis, P.; Harizanis, P.; Polissiou, M., Flavour compounds of Greek cotton honey. *Journal of the Science of Food and Agriculture* **2005,** 85, (9), 1444-1452.

43. de la Fuente, E.; Martínez-Castro, I.; Sanz, J., Characterization of Spanish unifloral honeys by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2005,** 28, (9-10), 1093-1100.
44. Baroni, M. V.; Nores, M. L.; Diaz, M. D. P.; Chiabrando, G. A.; Fassano, J. P.; Costa, C.; Wunderlin, D. A., Determination of volatile organic compound patterns characteristic of five unifloral honey by solid-phase microextraction-gas chromatography-mass spectrometry coupled to chemometrics. *Journal of Agricultural and Food Chemistry* **2006,** 54, (19), 7235-7241.
45. Alissandrakis, E.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M., Aroma investigation of unifloral Greek citrus honey using solid-phase microextraction coupled to gas chromatographic-mass spectrometric analysis. *Food Chemistry* **2007,** 100, (1), 396-404.
46. Mannas, D.; Altug, T., SPME/GC/MS and sensory flavour profile analysis for estimation of authenticity of thyme honey. *International Journal of Food Science and Technology* **2007,** 42, (2), 133-138.

# Chapter Two: Experimental and Method Development

This chapter details the samples, materials and instrumentation used in measuring honey volatiles. The experiments that were undertaken in order to develop an optimised method are discussed; also how data was treated once it was obtained.

## 2.1 Experimental

### 2.1.1 Honey Samples

Honey Samples were obtained from Airborne Honey Ltd., suppliers of New Zealand unifloral and polyfloral honeys. Information about the honey samples received is given in **Table 2.1**. Additional samples were obtained from Haddrell's of Cambridge, Waitemata Honey Co. Ltd., and the New Zealand Honey Food and Ingredient Advisory Service.

### 2.1.2 Standards

Standards of compounds important in determining floral source were tested: 3-Methylpentanoic acid, *ortho*-methoxyacetophenone, 2-methylbenzofuran, myrtenal, hexanoic acid and 3-methylbut-2-enal (Aldrich); Thymol (BDH); Phenol (Univar); (*E*)-Cinnamaldehyde (May & Baker); 1-(2-methoxyphenyl)ethanol (Acros); 2-ethylhexanoic acid (Supelco); Linalool (Fluka) and 1-phenylethanol (Sigma).

**Table 2.1: Samples Used in Experimental Work**

| Honey type | Number of samples | Taxonomic classification | Comments |
|---|---|---|---|
| Clover | 144 | *Trifolium repens* | Main species is white clover. Honey may also contain red, strawberry and subterranean clover.[1] |
| Honeydew | 26 | | Refers to honey produced by bees that have collected nectar that was previously exuded by another insect. Most NZ honeydew honey is from black beech (*Nothofagus solandri*) and red beech (*N. fusca*).[1] |
| Kamahi | 18 | *Weinmannia racemosa* | |
| Manuka | 29 | *Leptospermum scoparium* | Manuka honey can also contain kanuka honey (from *Kunzea ericoides*).[1] The two are difficult to distinguish. |
| Pohutukawa | 19 | *Mestrosideros excelsa* | |
| Rata | 23 | *Metrosideros umbellata* | |
| Rewarewa | 16 | *Knightia excelsa* | |
| Tawari | 28 | *Ixerba brexioides* | |
| Thyme | 12 | *Thymus vulgaris* | |
| Viper's Bugloss | 23 | *Echium vulgare* | |

Airborne Honey classified the majority of the unifloral honeys listed in **Table 2.1** on the basis of pollen content, colour, moisture, pH levels, conductivity, sugar profiles and organoleptic properties.[1] These methods were discussed in **Section 1.1**.

### 2.1.3 Materials and Instrumentation

Sample Preparation

Honey samples were prepared using:

- 2 mL crimp neck glass vials.
- 11 mm caps containing teflon/silicone septa. Prior to use, caps were baked overnight at 150 °C to remove siloxanes (a potential source of contamination).

SPME

Extractions were performed using:

- 85 µm carboxen/polydimethylsiloxane fibres, manufactured by Supelco.
- A Shimadzu AOC-5000 SPME auto injector.

GC-MS

Instrumental analyses were performed using the following equipment:

- Agilent Technologies 6890 GC and 5793 MS. The GC was equipped with a split-splitless inlet and CTC combi-pal robotic autosampler.
- DB-VRX capillary column (Agilent 122-1534) of length 30 m, internal diameter 250 µm and film thickness 1.4 µm.
- Helium gas (> 99.995% purity, BOC) - this was the carrier gas used in the GC-MS.
- Nitrogen gas (> 99.99% purity, BOC) – this was used to condition the SPME fibre prior to extraction.

### 2.1.4 Procedure

Sample Preparation

Honey samples (0.50 g ± 0.01 g) were weighed into vials with clean pasteur pipettes. Vials were sealed immediately. A blank (vial containing no honey) was included with every group of ten honey samples, to check for contamination. Honeys were stored below 1 °C when not in use.

Extraction

At the start of each sequence, the SPME fibre was conditioned in the injection port of the GC-MS (250 °C, 5 min) to remove contaminants. Honey samples were heated (70 °C, 5 min). During this time the SPME fibre was heated (250 °C) in the conditioning station under a flow of $N_2$ gas to desorb contaminants arising from solvents present in the laboratory. The SPME fibre was exposed to the headspace of honey samples via the vial septum, allowing volatile analytes to adsorb onto the fibre (70 °C, 20 min).

GC-MS analysis

The SPME fibre was injected into the GC-MS in splitless mode, with an injector temperature of 250 °C. The oven temperature was held at 35 °C (4 min) and was increased at 15 °C/min over 15 minutes. The total GC run time was 23 minutes. The mass spectrometer was run in full scan mode, using a mass range of *m/z* 33 – 350.

### 2.1.5 Identification

In the probability plots method used to identify floral source (**Chapter Three**), floral sources were discriminated between by presence and absence of specific compounds. Standards of these compounds were tested by SPME-GC-MS to confirm their identity. Retention times are given in **Table 2.2** and chromatograms can be found in **Appendix 1**.

**Table 2.2: Retention Times of Standards**

| **Compound** | **Retention time (min)** | **Compound** | **Retention time (min)** |
|---|---|---|---|
| 3-Methylbut-2-enal | 8.129 | 2-Methylbenzofuran | 12.909 |
| 3-Methylpentanoic acid | 10.399 | Myrtenal | 13.846 |
| Hexanoic acid | 10.960 | 1-(2-Methoxyphenyl)ethanol | 14.554 |
| Phenol | 11.069 | Thymol | 14.595 |
| 1-Phenylethanol | 12.081 | (*E*)-Cinnamaldehyde | 14.742 |
| 2-Ethylhexanoic acid | 12.506 | *ortho*-Methoxyacetophenone | 14.780 |
| Linalool | 12.517 | | |

For the multivariate statistical analyses discussed in **Chapters Four** and **Five**, compound identification was not as important, so compounds used in this method were identified by their mass spectra only, using the NIST 98 spectral library. The quality of hit is given in **Table 2.3.**

**Table 2.3: Quality of Hit for Compounds Identified by Mass Spectra**

| Compound | Quality (%) | Compound | Quality (%) |
|---|---|---|---|
| Dimethyl sulfide | 98 | Nonanal | 80 |
| Dimethyl sulfoxide | 93 | 2-Phenylethanol | 97 |
| 1,3,5,7-Cyclooctatetraene | 96 | 2,6-Dimethyl-1,3,5,7-octatetraene | 98 |
| 4-Hydroxybutanoic acid | 83 | Lilac aldehyde A | 59 |
| *p*-Benzoquinone | 78 | Octanoic acid | 86 |
| Furan-2-carbaldehyde | 81 | Lilac aldehyde B | 50 |
| Pentanoic acid | 80 | Benzoic acid | 95 |
| Tetrahydro-2,5-dimethyl furan | | Lilac aldehyde C | 91 |
| 5-Methyloxolan-2-one | 83 | Lilac aldehyde D | 58 |
| Benzaldehyde | 97 | 2,2,6-Trimethylcyclohexane-1,4-dione | 90 |
| 1,1'-Bicyclopentyl | 83 | 2,6-Dimethyl-3,7-octadiene-2,6-diol | 90 |
| 2,2,4,4-Tetramethylcyclobutane-1,3-dione | 87 | α-Terpineol | 59 |
| Benzyl alcohol | 95 | Lilac alcohol B | 50 |
| Pantoyl lactone | 90 | Lilac alcohol C | 59 |
| Heptanoic acid | 47 | Lilac alcohol A | 50 |
| Benzeneacetaldehyde | 94 | Lilac alcohol D | 64 |
| 4-Methyl-5*H*-furan-2-one | 95 | Benzeneacetic acid | 91 |
| *cis*-Linaloloxide | 80 | Nonanoic acid | 72 |
| Acetophenone | 96 | Nonan-4-one | 64 |
| Linalool oxide | 64 | Decanoic acid | 72 |
| Hortrienol | 87 | 2,6-Dimethyl-2,7-octadiene-1,6-diol | 80 |

## 2.2 Method Optimisation

A number of methods for extracting honey volatiles by SPME-GC-MS have been detailed[2-9] involving various fibre types and extraction parameters. Experiments were undertaken to improve recoveries in New Zealand honeys, particularly regarding fibre type, sample preparation, extraction temperature, sample agitation, and equilibration time. These experiments are detailed below.

### 2.2.1 Fibre Type

Several fibre types are commercially available, manufactured by Supelco. Six such fibres were tested on manuka honey (obtained from Manuka Health), and pohutukawa, rewarewa and tawari honeys (all obtained from HNZ):

CAR/PDMS 85 µm
DVB/CAR/PDMS 50/30 µm
PDMS 100 µm
CW/DVB 70 µm
PDMS/DVB 65 µm
PA 85 µm

CAR = carboxen, PDMS = polydimethylsiloxane, DVB = divinylbenzene, CW = carbowax, PA = polyacrylate.

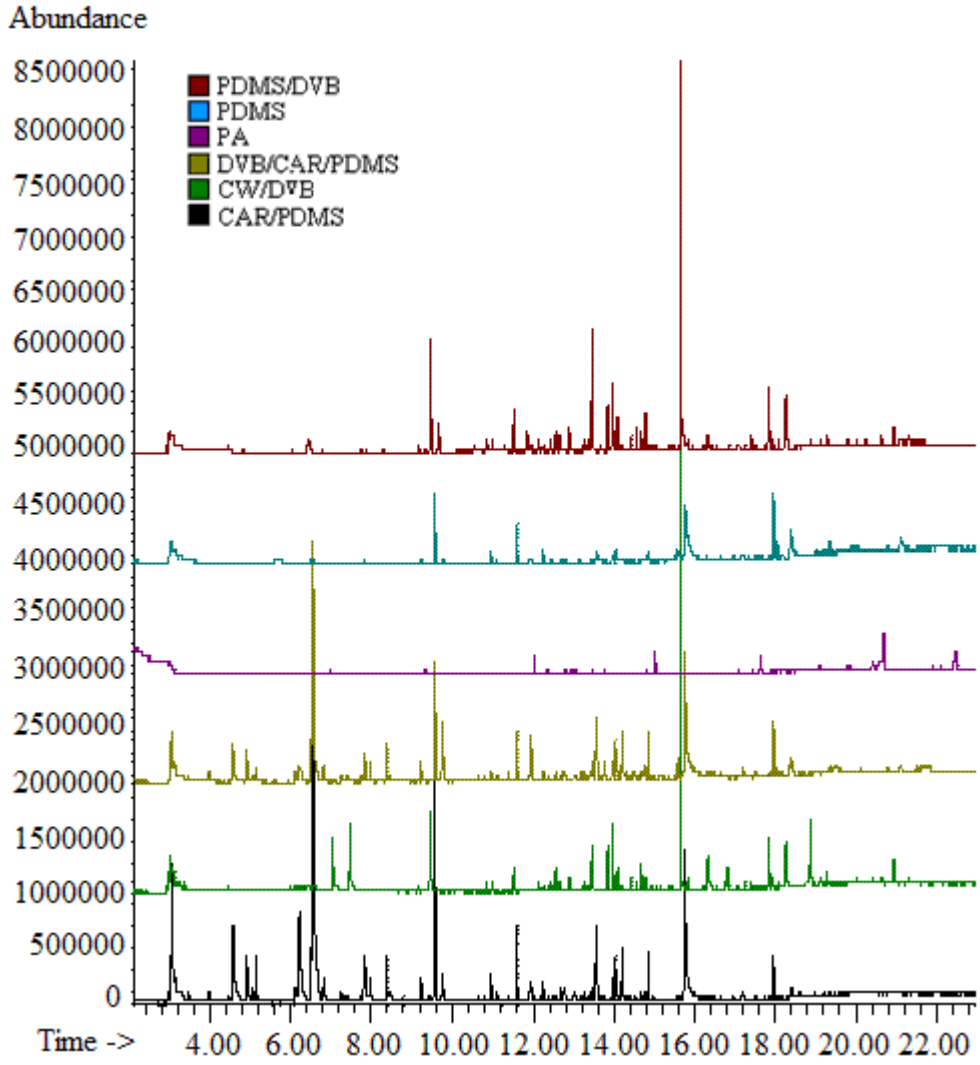



**Figure 2.1: Chromatogram of manuka honey comparing fibres**

For analytes that have a molecular weight less than 90, CAR/PDMS fibres are recommended, regardless of functional groups.[10] Carboxen particles act as a molecular sieve. They are porous, and retain small analytes that come into contact with the pores. CAR/PDMS fibres extract analytes by adsorption. Fibres coated in a liquid phase (such as PDMS, CW and PA) are different in that they operate by an absorption mechanism. For larger analytes, polarity of the fibre becomes increasingly important. CW/DVB and PA fibres are highly polar.[10]

**Figure 2.1** shows a GC chromatogram of manuka honey, extracted with CAR/PDMS, DVB/CAR/PDMS, PDMS, CW/DVB and PDMS/DVB fibres. The CAR/PDMS fibre showed better recoveries for lower molecular weight compounds. Other fibres such as DVB/CAR/PDMS, PDMS/DVB and PDMS showed greater sensitivity for high molecular weight compounds, but such compounds tended to be impurities from vial caps. The CW/DVB and PA fibres

yielded quite a different chromatogram from the others, as they are polar fibres whereas the others are non-polar. It was decided that the fibre to be used would be a CAR/PDMS fibre of thickness 85 µm.

### 2.2.2 Sample Preparation

Manuka, pohutukawa and tawari honeys were tested by SPME-GC-MS using varying amounts of honey and water, in order to test which method of sample preparation gave the best recovery.

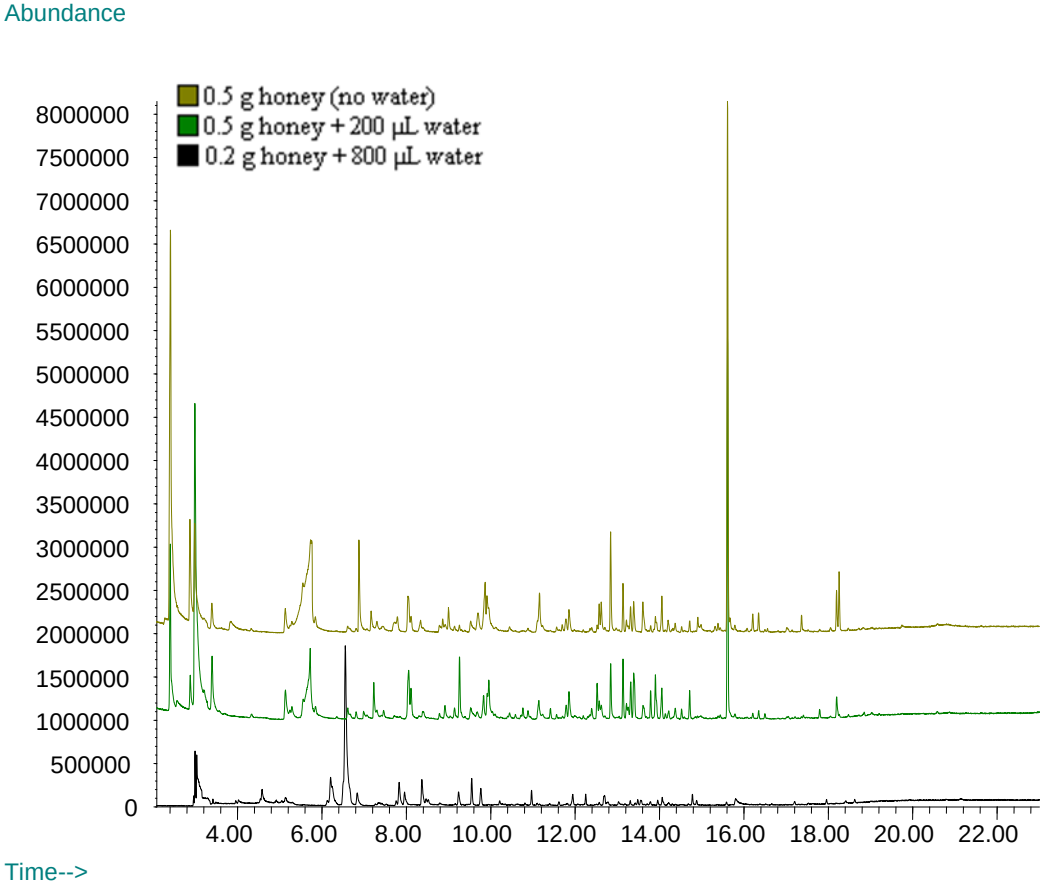



**Figure 2.2: Chromatogram of tawari honey comparing sample preparation methods**

As shown in **Figure 2.2**, better sensitivity was achieved for most compounds when honey was analysed with no extra water present. Some compounds such as hydrocarbons had higher recoveries when water was used, but these were likely to be contaminants from the laboratory environment.

Some researchers have used water to decrease the density of the honey matrix,[2, 5, 8, 9] or added salt to decrease the solubility of hydrophilic compounds in the aqueous

phase.[7, 9] However if water is not used in the first place, salt is not needed to counteract the increased retention of hydrophilic compounds.

It was decided that 0.5 g of honey would be used, with no water added.

## 2.2.3 Extraction Temperature

Increasing the extraction temperature increases the concentration of analytes in the headspace, and therefore increases recovery.[11] However the temperature should not be so high that artefacts such as Maillard reaction products form. By extracting manuka, pohutukawa and tawari honeys at 30 °C, 50 °C and 70 °C for 20 minutes (**Figure 2.3**), it was found that an extraction temperature of 70 °C gave a higher recovery for all compounds.

It was decided that an extraction temperature of 70 °C would be used.

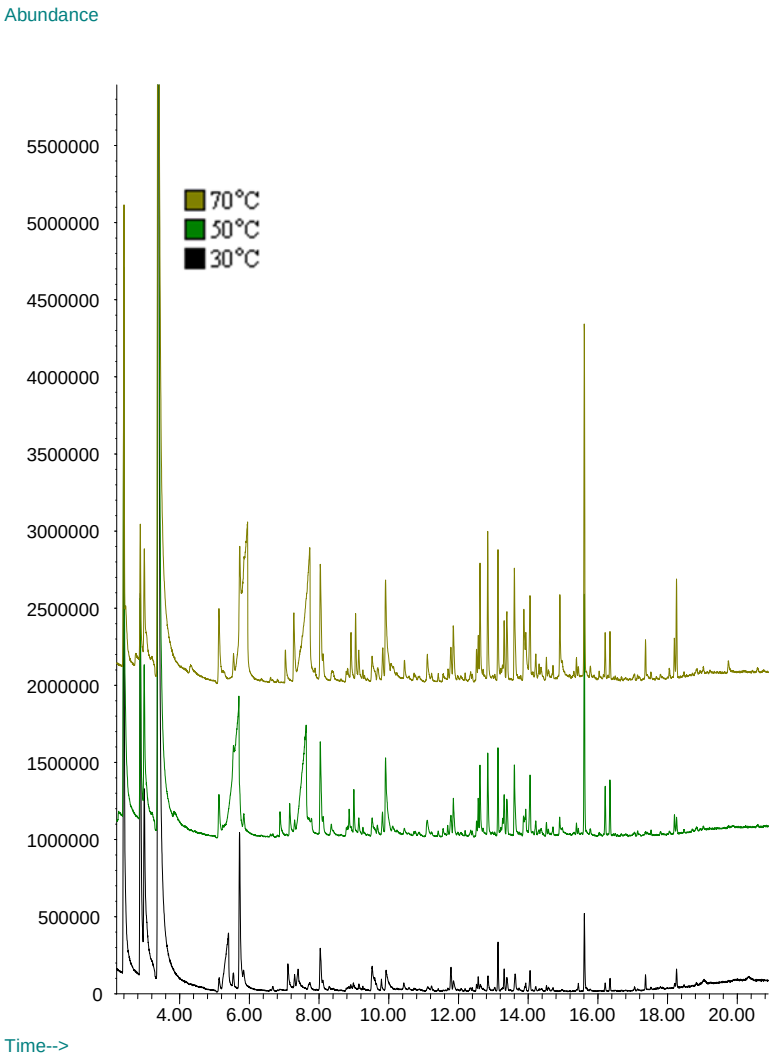



**Figure 2.3: Chromatogram of pohutukawa honey comparing extraction temperatures**

### 2.2.4 Agitation

When honey samples are dissolved in water, agitation can help release volatile analytes from the honey matrix and improve recovery. However when manuka, pohutukawa and tawari honey samples were extracted without water, agitation did not seem to make a significant difference, as shown in **Figure 2.4**. Agitation lead to increased sensitivity for some analytes, and poorer sensitivity for others.

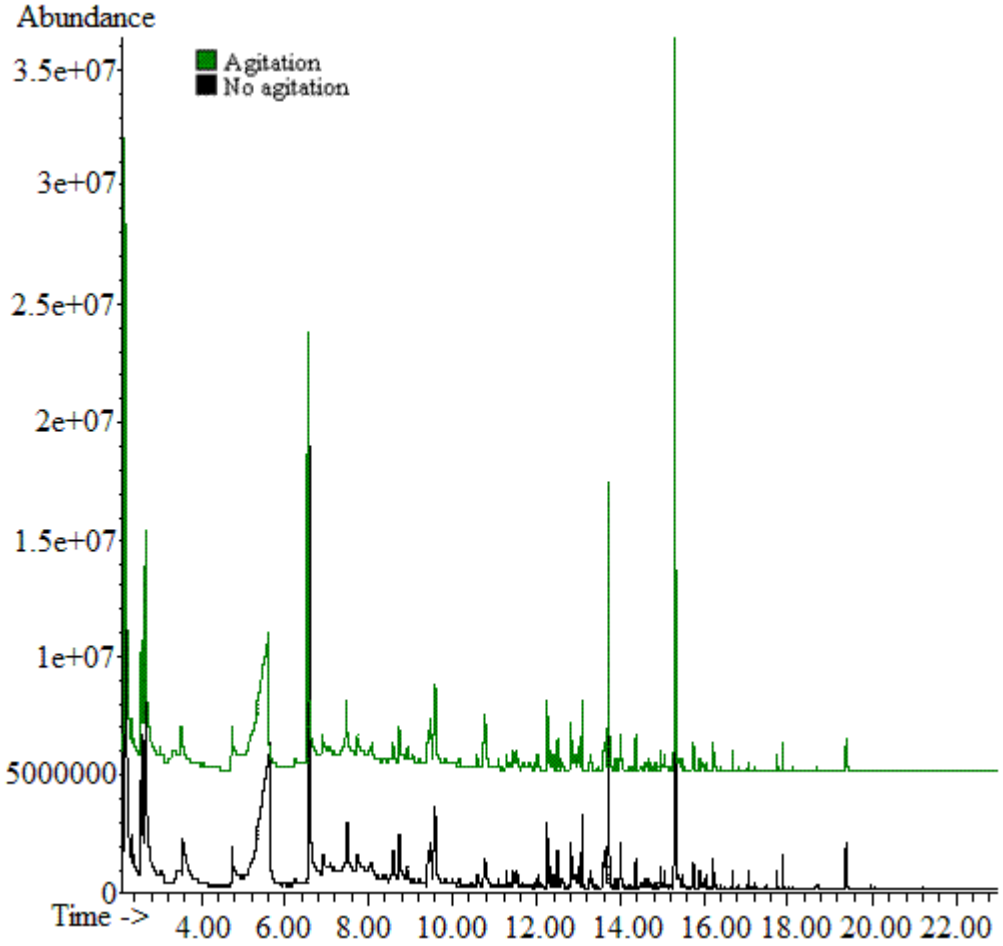



**Figure 2.4: Chromatogram of manuka honey comparing the effect of agitation**

It was decided that agitation would not be used.

## 2.2.5 Equilibration Time

Varying equilibration times were tested on manuka honey.

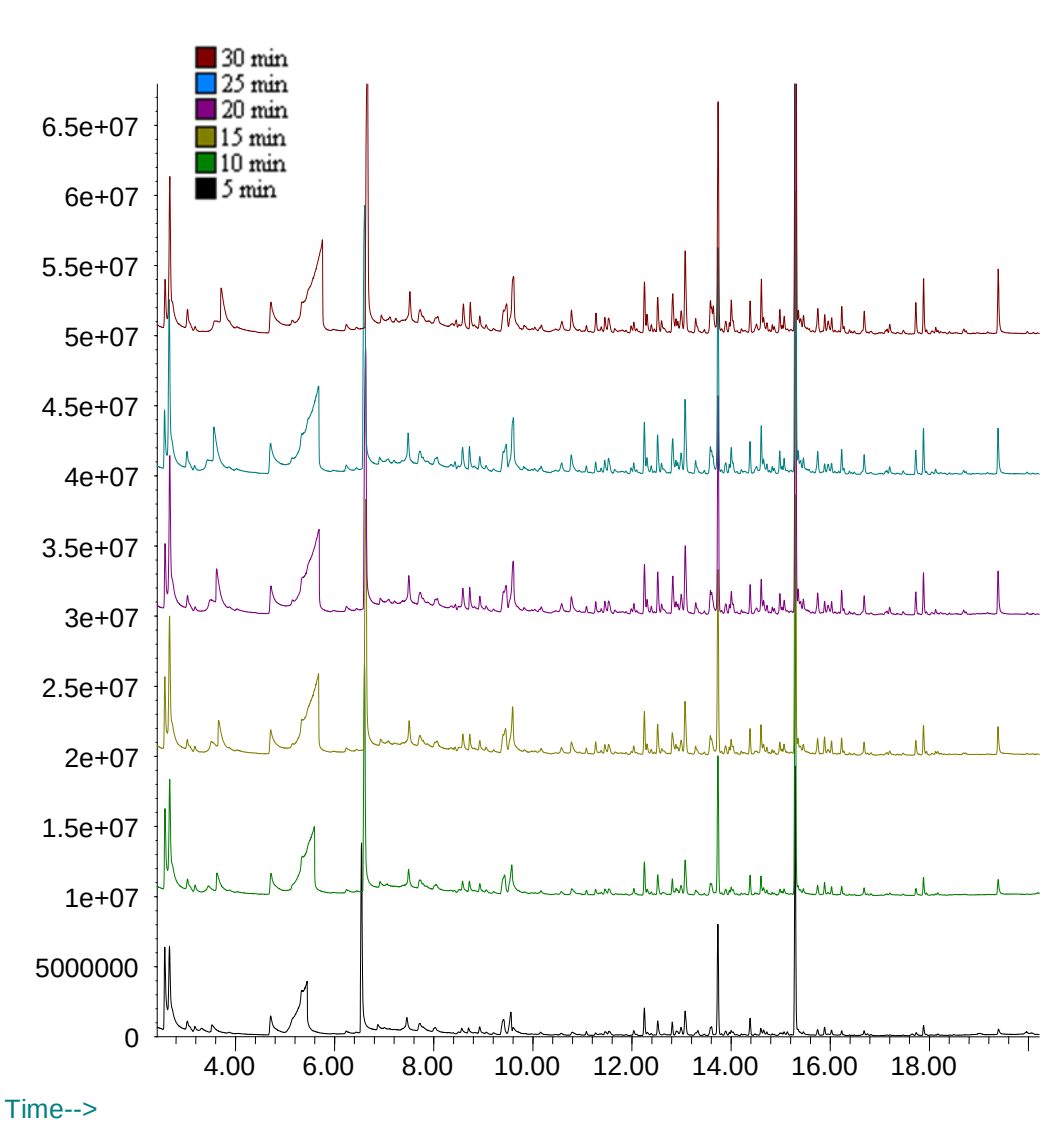



**Figure 2.5: Chromatogram of manuka honey comparing equilibration times**

The longer a sample is left to equilibrate, the higher the analyte recovery, as shown in **Figure 2.5**. However after a certain time, the increase in recovery is minimal, as shown in **Figure 2.6**. This is because at equilibrium, the transfer of analytes into the coating is the same as the transfer of analytes desorbing from the coating to the headspace.[11] Taking these factors into account, as well as the need to minimise the time for each extraction where possible, it was decided that samples would be allowed to equilibrate for 20 minutes.

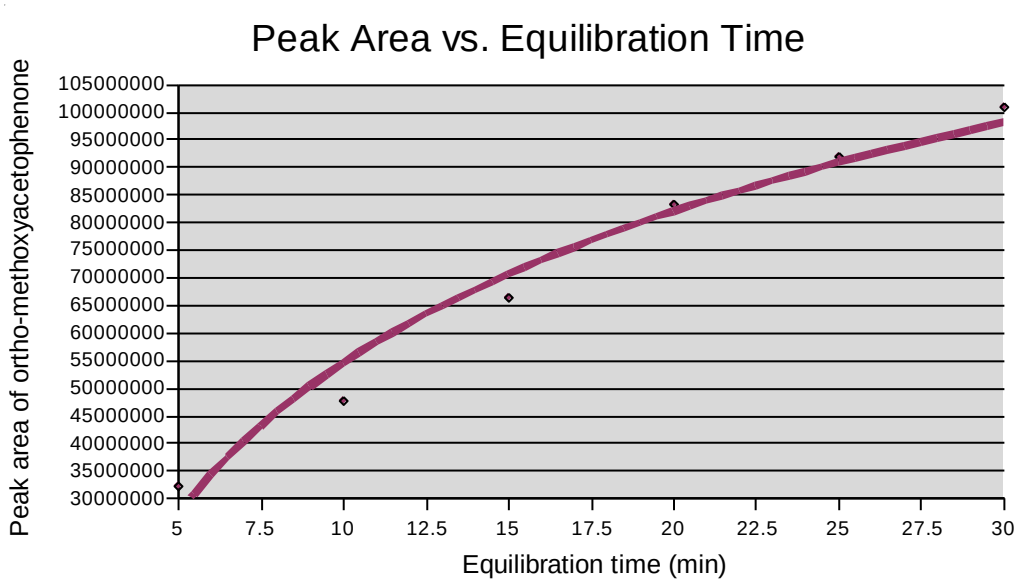



**Figure 2.6: Graph of peak area vs. equilibration time for *ortho*-methoxyacetophenone in manuka honey**

### 2.2.6 Summary

A summary of optimised method variables is given in **Table 2.4**.

**Table 2.4: Summary of Optimised Method Variables**

| **Variable** | **Selected Parameter** |
|---|---|
| Fibre type | CAR/PDMS 85 µm |
| Sample preparation | 0.5 g honey |
| Extraction temperature | 70 °C |
| Agitation | No agitation |
| Equilibration time | 20 minutes |

## 2.3 Data Treatment

This section details the processes used to extract information from the large number of chromatograms obtained after SPME-GC-MS analysis (**Appendices 2 – 11**). The intention was to reduce the hundreds of complex chromatograms to a list of compounds that could be used to discriminate between floral source.

Two chromatograms from each floral source were examined in detail. All of the peaks in these chromatograms were identified by the NIST 98 mass spectral

library and recorded. A list of 240 compounds was obtained. A further five chromatograms from each floral source were processed using this list, with the intention of gaining more information on the types of compounds present in the honey samples. It was observed that a large number of compounds could be eliminated. These were mainly:

- Compounds in the first five minutes of the chromatograms. These tended to be common to all samples and included compounds such as acetone and ethyl acetate.
- Compounds in the last five minutes of the chromatograms. These were difficult to identify due to siloxane contamination from vial caps.
- Compounds that appeared in one sample of one honey type only.

The final list contained 55 compounds. For the full list see **Appendix 12**. All chromatograms were worked up based on this list, using Enviroquant software. Peaks were identified using their mass spectra, and integrated. Finally the integrated peak area for each compound was tabulated for all samples. Because an internal standard was not run with the honey samples, peak areas could not be fully quantified; hence this approach was semi-quantitative.

Due to the large spread of values, the logarithm to base ten was calculated for each value in order to make the data set more manageable. Data was subjected to analysis by probability plots (**Chapter Three**), multivariate pattern recognition techniques (**Chapter Four**) and data mining algorithms (**Chapter Five**). Minitab 15 was used to generate probability plots and perform multivariate analysis. The data mining software used was Weka version 3.5.4.

# 2.4 References


1. Airborne Honey Ltd. Airborne's New Zealand Honey Collections. http://www.airborne.co.nz (03/12/07)

2. Alissandrakis, E.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M., Aroma investigation of unifloral Greek citrus honey using solid-phase microextraction coupled to gas chromatographic-mass spectrometric analysis. *Food Chemistry* **2007,** 100, (1), 396-404.

3. Ampuero, S.; Bogdanov, S.; Bosset, J.-O., Classification of unifloral honeys with an MS-based electronic nose using different sampling modes: SHS, SPME and INDEX. *European Food Research and Technology* **2004,** 218, 198-207.

4. Baroni, M. V.; Nores, M. L.; Diaz, M. D. P.; Chiabrando, G. A.; Fassano, J. P.; Costa, C.; Wunderlin, D. A., Determination of volatile organic compound patterns characteristic of five unifloral honey by solid-phase microextraction-gas chromatography-mass spectrometry coupled to chemometrics. *Journal of Agricultural and Food Chemistry* **2006,** 54, (19), 7235-7241.

5. de la Fuente, E.; Martínez-Castro, I.; Sanz, J., Characterization of Spanish unifloral honeys by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2005,** 28, (9-10), 1093-1100.

6. Perez, R. A.; Sanchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L., Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography-mass spectrometry. *Journal of Agricultural and Food Chemistry* **2002,** 50, (9), 2633-2637.

7. Piasenzotto, L.; Gracco, L.; Conte, L., Solid phase microextraction (SPME) applied to honey quality control. *Journal of the Science of Food and Agriculture* **2003,** 83, (10), 1037-1044.

8. Soria, A.; Martínez-Castro, I.; Sanz, J., Analysis of volatile composition of honey by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2003,** 26, (9-10), 793-801.

9. Verzera, A.; Campisi, S.; Zappala, M.; Bonaccorsi, I., SPME-GC-MS analysis of honey volatile components for the characterization of different floral origin. *American Laboratory* **2001**, 18-21.

10. Supelco How to choose the proper SPME fibre. http://www.sigmaaldrich.com/Brands/Supelco/Home/Literature/Analytical_Technique/Solid_Phase_Microextraction.html (03/12/07)

11. Pawliszyn, J., *Solid Phase Microextraction: Theory and Practice*. Wiley-VCH: New York, 1997.

# Chapter Three: Probability Plots as a Tool for Determining Floral Source

Probability plots were generated for compounds in all honey samples. Visual analysis of the plots and a process of elimination allowed pohutukawa, thyme, manuka, honeydew, rata, kamahi, viper's bugloss, clover and tawari honeys to be identified.

## 3.1: How to Read Probability Plots

Probability plots show abundance of each compound across the *x* axis (the semi-quantitated concentration), and the percentage of samples falling within a region (based on a Normal probability distribution) up the *y* axis. An example of a probability plot is shown in **Figure 3.1:**

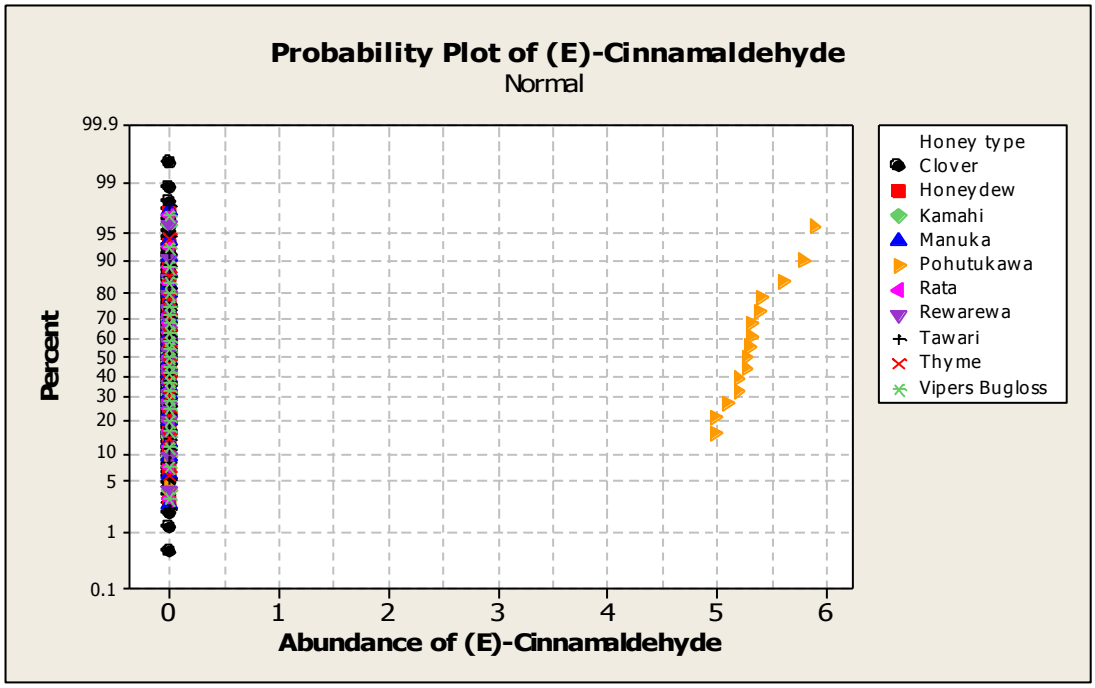



**Figure 3.1: Probability plot of (*E*)-cinnamaldehyde**

**Figure 3.1** shows that pohutukawa is the only type of honey with (*E*)-cinnamaldehyde in it, and that it was detected in approximately 90% of samples. The points at zero on the *x* axis show that 10% of pohutukawa samples do not have (*E*)-cinnamaldehyde in them (although individual points are difficult to distinguish), and 100% of all other honey types do not contain this compound.

## 3.2: Methodology

Probability plots were generated for all compounds under investigation in the set of honey samples (see **Appendix 11** for a list of compounds). Visual examination of the plots allowed groups of compounds to be chosen which represented a particular floral source in some way. That floral type was then eliminated from the set, so that discrimination became apparent for other floral sources.

# 3.3: Results and Discussion

## 3.3.1 Pohutukawa

Pohutukawa honey was indicated by the presence of 3-methylbut-2-enal, dimethyl sulfoxide and (*E*)-cinnamaldehyde (See **Figure 3.2** and **Structures I - III**).

Using this method, 15/19 samples were classified correctly. There were no samples of other floral sources that were incorrectly classified as pohutukawa. Of the four misclassified pohutukawa samples, two had been labelled by the beekeeper as a poor example of pohutukawa honey, as contamination with bush honey was likely. These two samples were removed from the data set. The other two samples were re-tested by SPME-GC-MS; however results remained the same.

*Proportion of samples classified correctly: 15/17.*

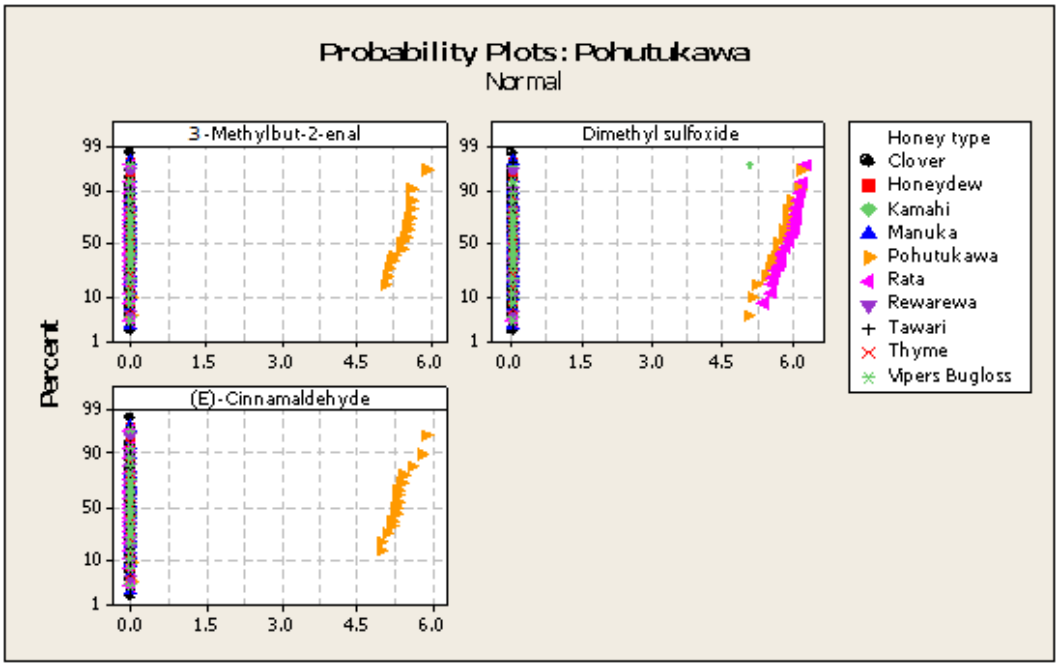
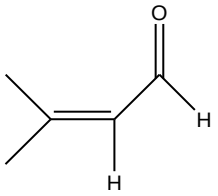
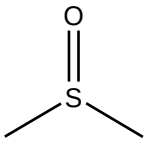


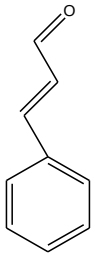

**Figure 3.2: Probability plots of 3-methylbut-2-enal, dimethyl sulfoxide and (*E*)-cinnamaldehyde**




**Structure I: 3-Methylbut-2-enal**




**Structure II: Dimethyl sulfoxide**




**Structure III: (*E*)-Cinnamaldehyde**

**Figure 3.3** shows a chromatogram of a pohutukawa honey sample, with important peaks labelled.

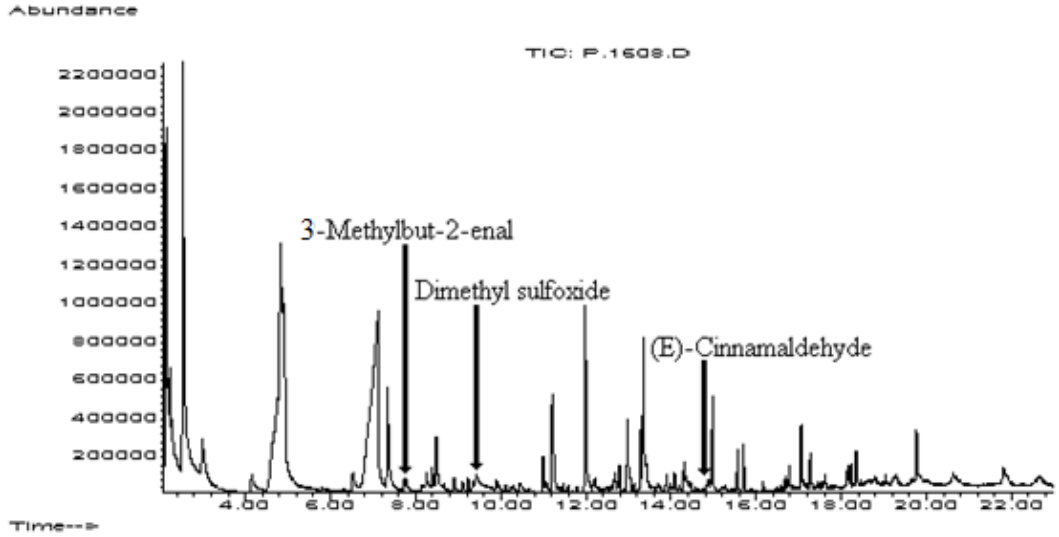



**Figure 3.3: Chromatogram of pohutukawa honey showing 3-methylbut-2-enal, dimethyl sulfoxide and (*E*)-cinnamaldehyde**

3-Methylbut-2-enal has previously been found in Spanish citrus, rosemary and polyfloral honeys using purge-and-trap GC-MS.[1] It was also found in a study of blended honeys of various European origins, which were analysed using SPME followed by two-dimensional GC-TOFMS.[2]

Dimethyl sulfoxide has been found in rosemary honey after liquid-liquid extraction; it was suggested that it forms from the oxidation of dimethyl sulfide.[3]

(*E*)-Cinnamaldehyde was proposed as a marker of Greek Cotton honey after it was isolated by both SPME and USE.[4] More recently it has been suggested that (*E*)-cinnamaldehyde forms from the corresponding cinnamyl alcohol.[5] In the present study cinnamyl alcohol was not observed in GC chromatograms after SPME, most likely because of its high water solubility.

### 3.3.2 Thyme

Thyme honey was indicated by the presence of thymol and hexanoic acid (see **Figure 3.4** as well as **Structures IV** and **V**).

These two compounds enabled 11/12 thyme samples to be classified correctly. The sample that was misclassified was re-tested; however the same result was obtained. There were no samples of other floral sources that were incorrectly classified as thyme honey.

*Proportion of samples classified correctly: 11/12.*

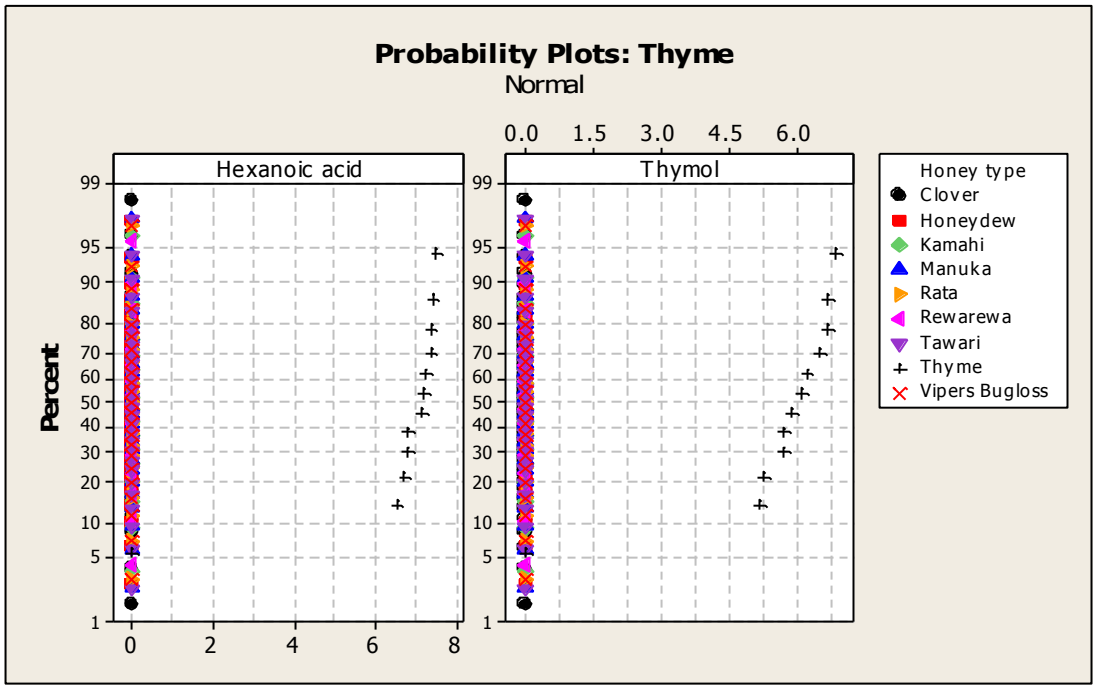
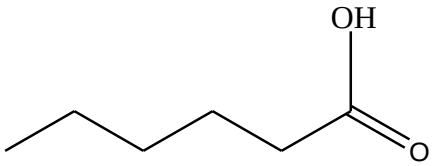
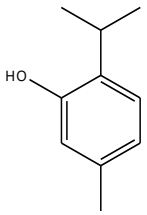



**Figure 3.4: Probability plots for hexanoic acid and thymol**




**Structure IV: Hexanoic acid**




**Structure V: Thymol**

**Figure 3.5** shows a chromatogram of a thyme honey sample with the hexanoic acid and thymol peaks labelled.

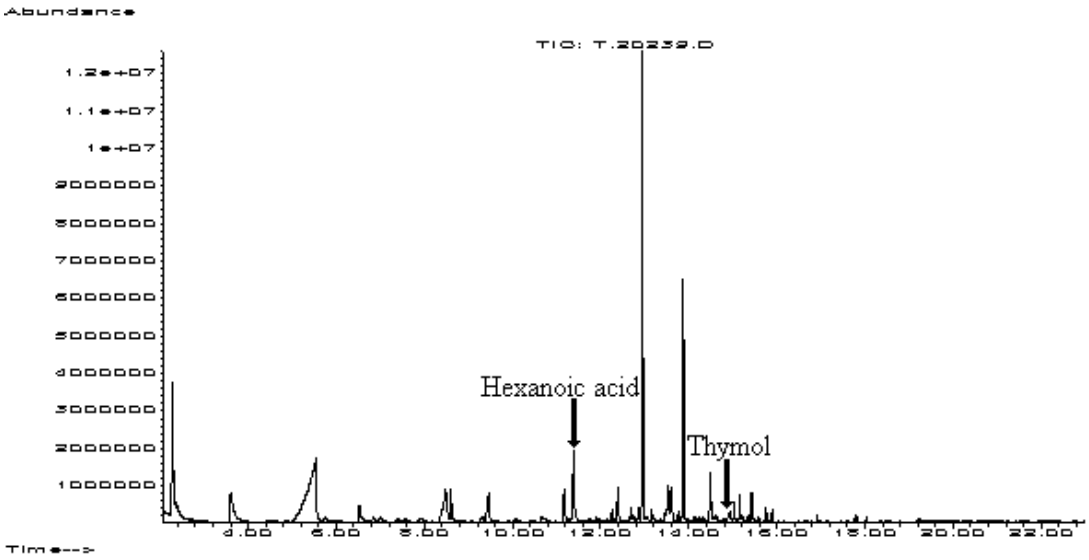



**Figure 3.5: Chromatogram of thyme honey showing hexanoic acid and thymol**

Thymol has been found in many foreign honeys using a range of extraction procedures,[2, 6-9] including Greek thyme honey analysed by SPME-GC-MS.[10] Similarly, hexanoic acid has been found in a diverse range of honeys such as citrus, chestnut and honeydew.[5-7, 11-15]

A potential limitation of using thymol to classify thyme honey is its use as a control agent for the varroa bee mite. Thymol is an active ingredient in two control agents (Apiguard and Apilife VAR) which have been used to control varroa in certain countries.[16] Trials have been conducted in New Zealand to study the effectiveness of thymol on varroa.[17] If thymol becomes a widely-used control agent in New Zealand, it is possible that honey produced in such hives could be contaminated with residual thymol.

### 3.3.3 Manuka

Manuka honey was identifiable by the presence of 2-methylbenzofuran and 1-(2-methoxyphenyl)ethanol, as well as either myrtenal or 1-phenylethanol (see **Figure 3.6** and **Structures VI - IX**).

This method classified 27/29 samples correctly. There were no samples of other floral sources that were incorrectly classified as manuka. Of the two manuka samples that were misclassified, reference to pollen data showed that one sample didn't have manuka pollen in it. This sample was removed from the data set. The other sample was re-tested; however the same result was obtained.

*Proportion of samples classified correctly: 27/28.*

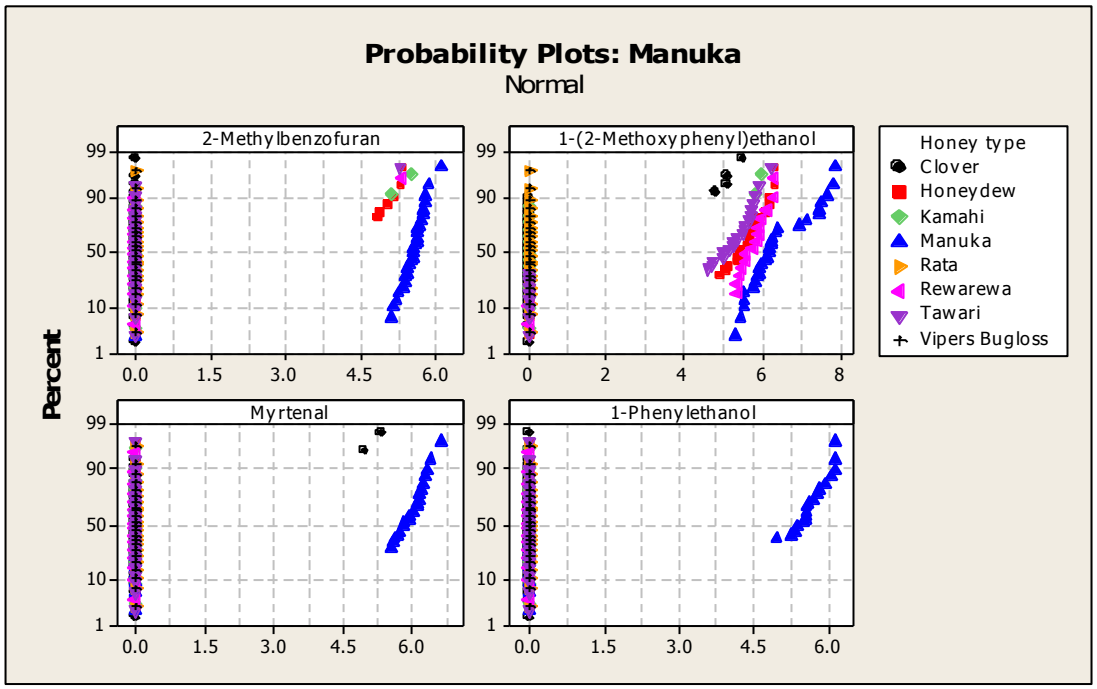
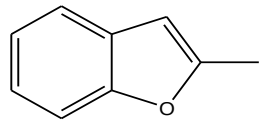
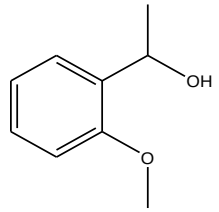


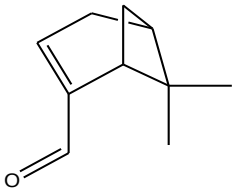

**Figure 3.6: Probability plots for 2-methylbenzofuran, 1-(2-methoxyphenyl)ethanol, myrtenal and 1-(2-methoxyphenyl)ethanol**



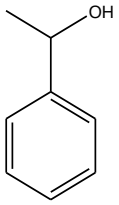

**Structure VI: 2-Methylbenzofuran**




**Structure VII: 1-(2-Methoxyphenyl)ethanol**




**Structure VIII: Myrtenal**




**Structure IX: 1-Phenylethanol**

**Figure 3.7** shows the chromatogram of a manuka honey sample that contained all four compounds important in classifying manuka honey.

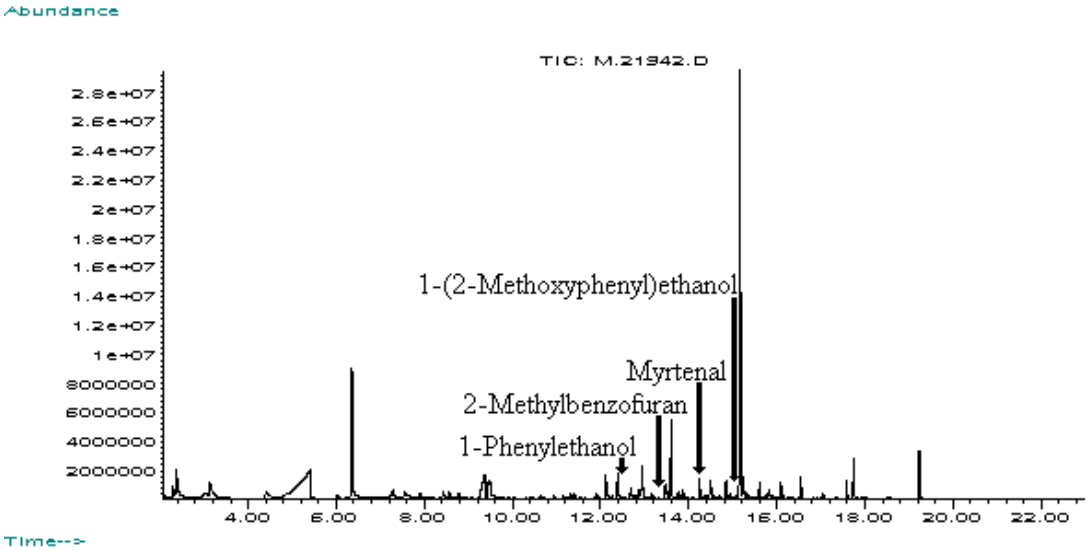



**Figure 3.7: Chromatogram of manuka honey showing 1-phenylethanol, 2-methylbenzofuran, myrtenal and 1-(2-methoxyphenyl)ethanol**

1-Phenylethanol and 2-methylbenzofuran have been found in polyfloral honey by SPME-GC-MS.[18] The latter was also found in polyfloral honey using SPME and GC × GC-TOFMS.[2] Myrtenal and 1-(2-methoxyphenyl)ethanol have not been reported as honey constituents before; however myrtenal has been found in bee propolis.[19]

### 3.3.4 Honeydew

Honeydew honey was indicated by the presence of either linalool and phenol, or linalool and 1-(2-methoxyphenyl)ethanol (see **Figure 3.8** as well as **Structures X, XI** and **VII**).

Using this method, 24/26 samples were classified correctly. The two samples that were incorrectly classified were re-tested, and were subsequently classified as honeydew. Two tawari samples were misclassified as honeydew. These were re-tested, and found not to meet the criteria for honeydew honey.

*Proportion of samples classified correctly: 26/26*

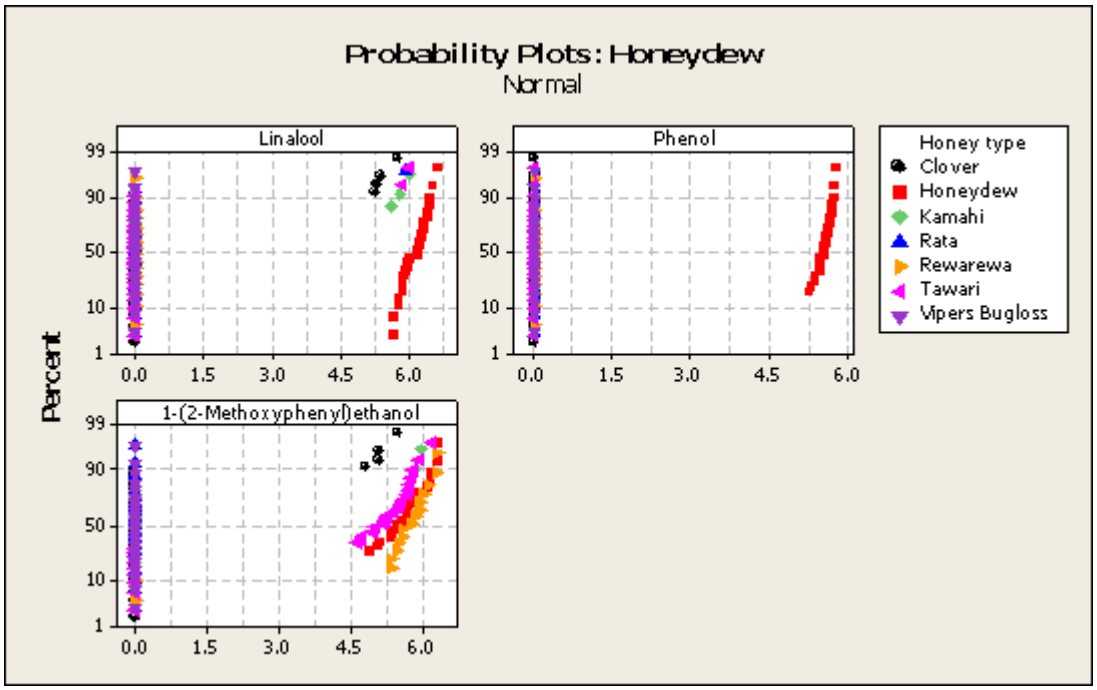
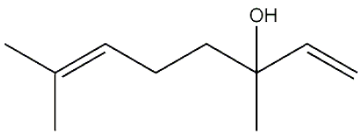
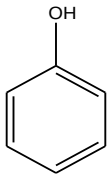



**Figure 3.8: Probability plots for linalool, phenol and 1-(2-methoxyphenyl)ethanol**




**Structure X: Linalool**




**Structure XI: Phenol**

**Figure 3.9** shows a chromatogram of a sample of honeydew honey. Linalool, phenol and 1-(2-methoxyphenyl)ethanol are labelled.

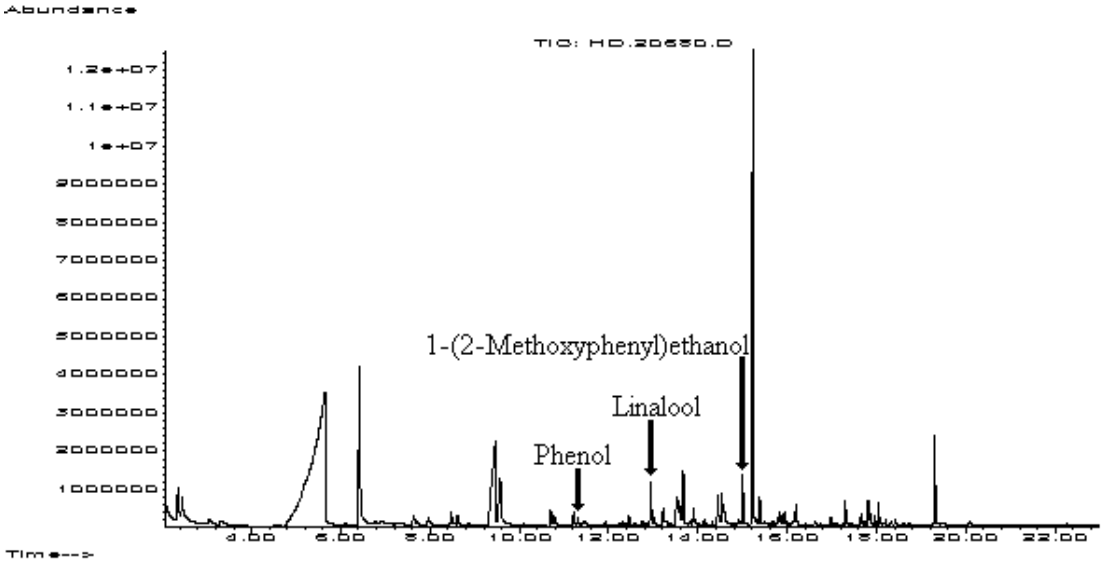



**Figure 3.9: Chromatogram of honeydew honey showing phenol, linalool and 1-(2-methoxyphenyl)ethanol**

Phenol has been found in a number of different unifloral honeys using a range of extraction procedures,[8, 9, 18, 20-22] including New Zealand heather honey.[23] Heather honey was not analysed in the present study. It has been suggested that phenols are produced by the biochemical degradation of phenolic acids in honey.[20]

Linalool has been widely characterised as a volatile compound in honey.[2-4, 8, 10, 15, 18, 24-32] It has been found in large amounts in Sicilian wildflower and sulla honeys,[14] and was reported in a study of citrus honey and citrus trees as being the major compound in extracts of Greek orange, tangerine and sour tree blossoms.[33] It was noted that linalool gives rise to other terpenoids found in honey.[6, 33]

Sixteen linalool derivatives were found in New Zealand nodding thistle honey after liquid-liquid extraction, and proposed as markers for this honey type.[34]

As noted in **Section 3.3.3**, 1-(2-methoxyphenyl)ethanol has not previously been reported in honey.

### 3.3.5 Rata

Rata honey was identifiable by the presence of dimethyl sulfide and dimethyl sulfoxide (see **Figure 3.10** as well as **Structures XII** and **II**).

The presence of these two compounds enabled 22/23 samples to be classified correctly. The one misclassified sample was re-tested and found to meet the criteria for rata honey. There were no samples of other floral sources that were incorrectly classified as rata.

*Proportion of samples classified correctly: 23/23*

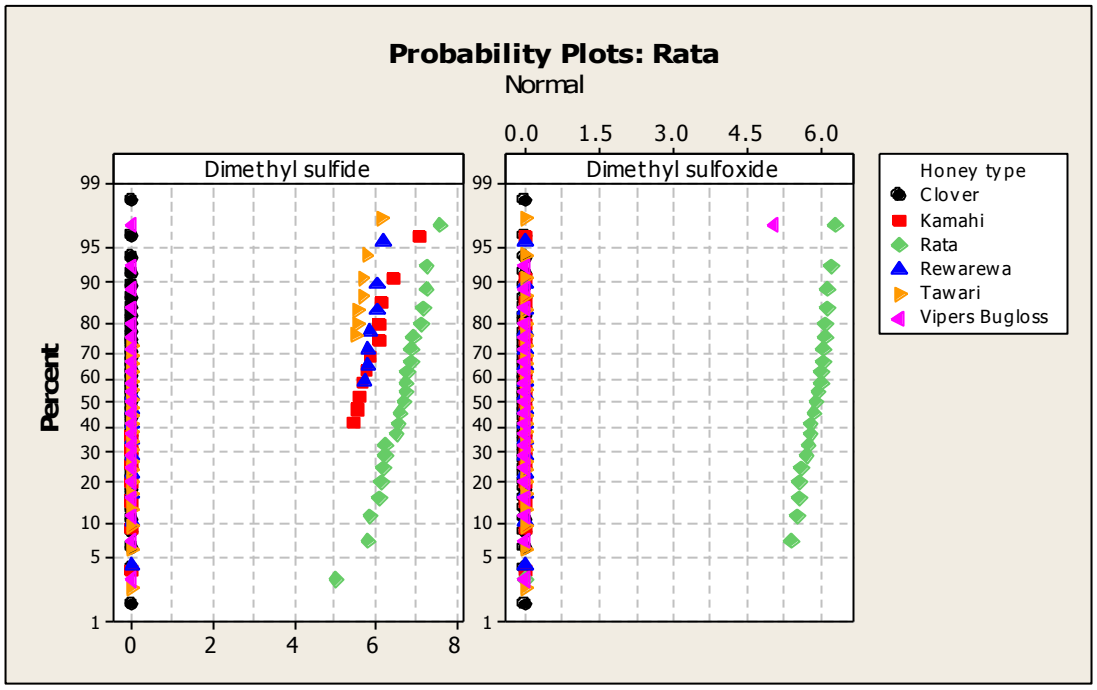
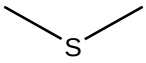



**Figure 3.10: Probability plots for dimethyl sulfide and dimethyl sulfoxide**




**Structure XII: Dimethyl sulfide**

**Figure 3.11** shows the chromatogram of a sample of rata honey, with the peaks dimethyl sulfide and dimethyl sulfoxide labelled.

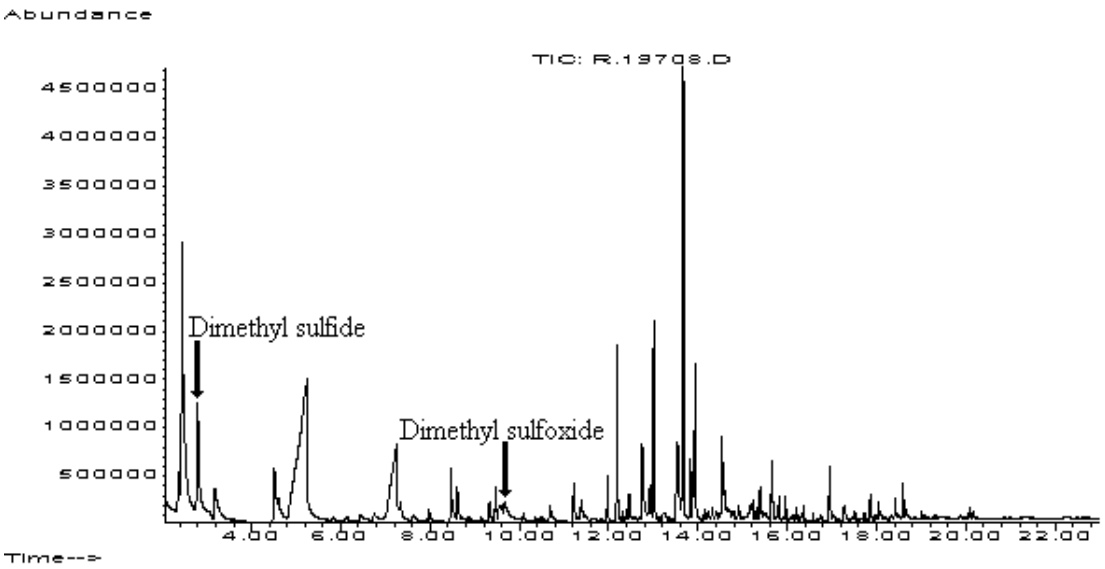



**Figure 3.11: Chromatogram of rata honey showing dimethyl sulfide and dimethyl sulfoxide**

Dimethyl sulfide was found by SPME-GC-MS in rare Spanish honeys including avocado and oak.[35] In an earlier study of SPME-GC-MS on Spanish honeys, dimethyl sulfide was found in common honeys such as rosemary, orange and eucalyptus.[36]

The occurrence of dimethyl sulfoxide in honey was discussed in **Section 3.3.1**.

### 3.3.6 Kamahi

Kamahi honey was indicated by the absence of pantoyl lactone and an abundance of 4-methyl-5*H*-furan-2-one greater than 6 (see **Figure 3.12** as well as **Structures XIII** and **XIV**).

This allowed 17/19 samples to be classified correctly. Of the two misclassified samples, one had been noted by the beekeeper to be a poor example of kamahi honey due to contamination from other honey in the frames from which the sample was collected. This sample was removed from the data set. The other sample was re-tested; however the result did not change.

One rewarewa sample was incorrectly classified as being kamahi. However the beekeeper had noted that this sample was a poor example of rewarewa honey, as it was contaminated with kamahi. The sample was removed from the data set. *Proportion of samples classified correctly: 17/18.*

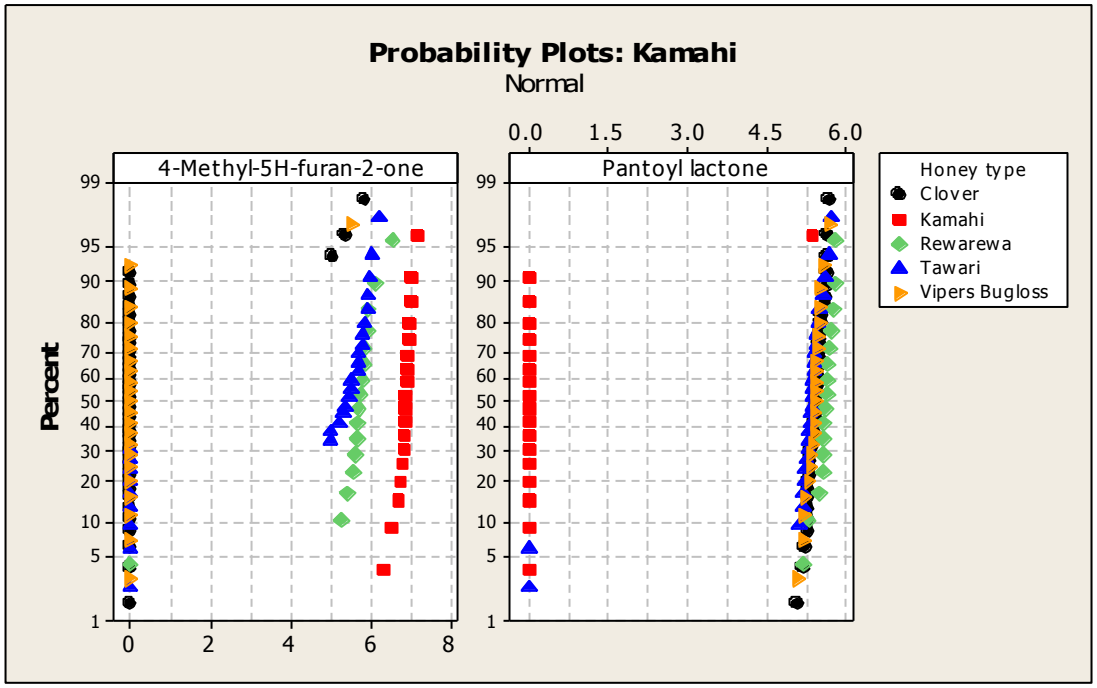
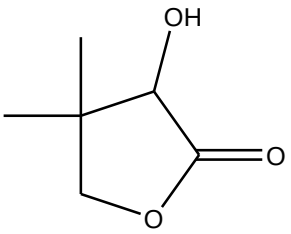
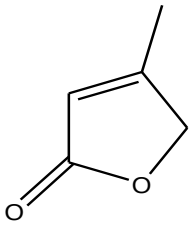



**Figure 3.12: Probability plots for 4-methyl-5*H*-furan-2-one and pantoyl lactone**




**Structure XIII: Pantoyl lactone**




**Structure XIV: 4-Methyl-5*H*-furan-2-one**

**Figure 3.13** shows a chromatogram of a kamahi sample with the 4-methyl-5*H*-furan-2-one peak labelled. Pantoyl lactone is shown in **Figure 3.15**, the chromatogram for viper's bugloss honey.

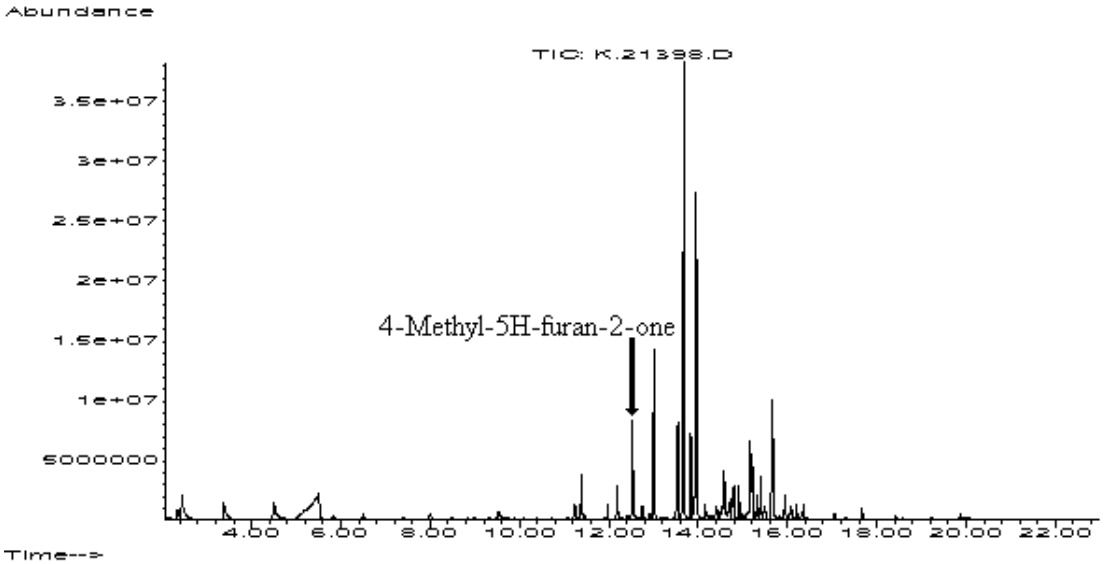



**Figure 3.13: Chromatogram of kamahi honey showing 4-methyl-5*H*-furan-2-one**

Pantoyl lactone and 4-methyl-5*H*-furan-2-one have not been reported as honey constituents before. However similar compounds to 4-methyl-5*H*-furan-2-one have been found; 3-methyl-3*H*-furan-2-one has been found in Japanese haze honey[15] and 5-methyl-3*H*-furan-2-one was reported in Spanish citrus honey.[6] 4-Methyl-5*H*-furan-2-one has been found in snake fruit[37] and roasted coffee been aroma analysed by SPACE (Solid-Phase Aroma Concentrate Extraction; a modification by SPME).[38]

Pantoyl lactone has been found in red wine from Uruguay,[39] and in bulbs of the plant *Fritillaria imperialis*.[40]

### 3.3.7 Viper's Bugloss

The presence of *p*-benzoquinone indicated viper's bugloss honey (see **Figure 3.14** and **Structure XV**).

Of the viper's bugloss samples, 19/23 were classified correctly. The four anomalous samples were re-tested; however results did not change. One clover sample was incorrectly classified as viper's bugloss, although upon inspection of pollen data it was revealed that information was not available for that particular sample, so it was discarded from the sample set.

*Proportion of samples classified correctly: 19/23.*

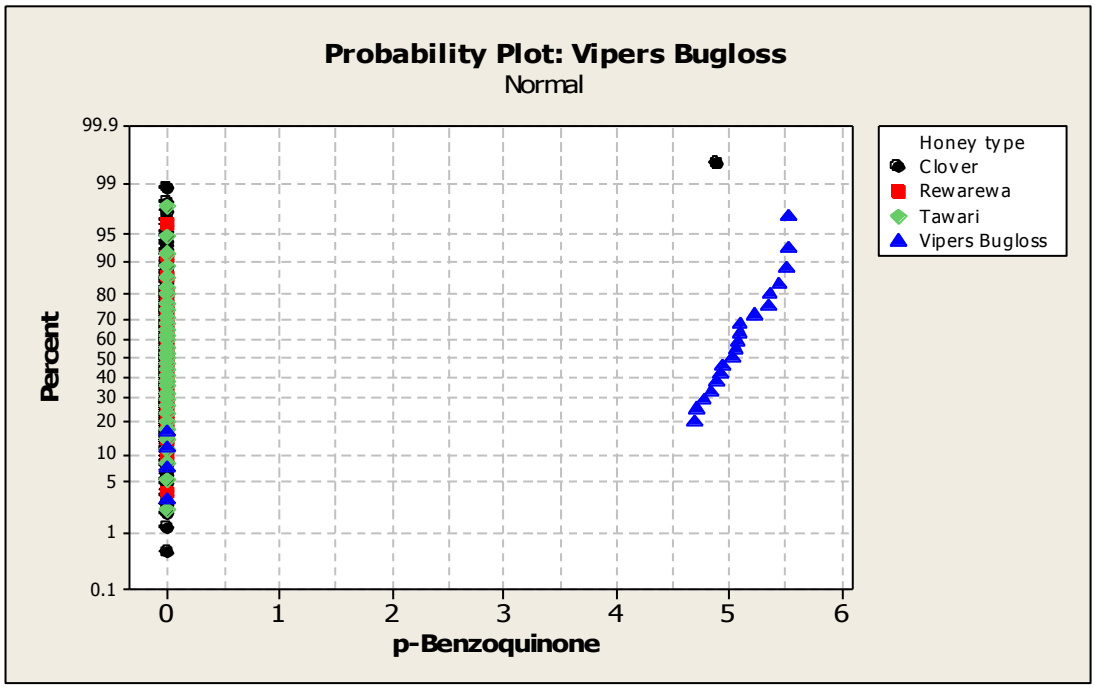
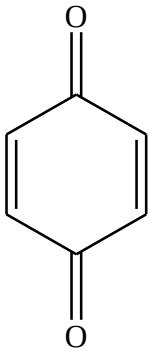



**Figure 3.14: Probability plot for *p*-benzoquinone**




**Structure XV: *p*-Benzoquinone**

**Figure 3.15** shows a chromatogram of a viper's bugloss honey sample with *p*-benzoquinone and pantoyl lactone. The latter was included to show whereabouts in the chromatogram it occurs; its absence is an indicator of kamahi honey (**Section 3.3.6).**

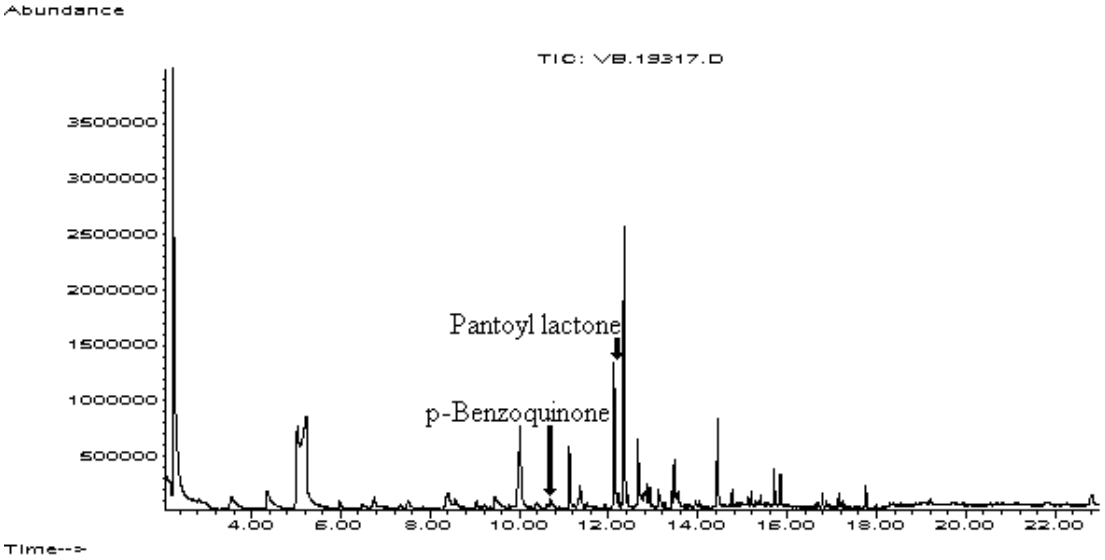



**Figure 3.15: Chromatogram of viper's bugloss honey showing *p*-benzoquinone and pantoyl lactone**

*p*-Benzoquinone has been found in New Zealand viper's bugloss honey.[36] One result of that study (which used liquid-liquid extraction) was that hydroquinone was proposed as a marker of viper's bugloss honey. Hydroquinone is not seen in the GC chromatograms of viper's bugloss honey after SPME due to its high water solubility.

### 3.3.8 Clover

Initially it was thought that it would be outside the scope of this method to classify clover honey. However upon further inspection of information provided with the samples, it was discovered that most of the samples labelled clover did not have pollen analysis data to support this classification. These samples were removed, leaving a set of 43 clover honeys. This smaller data set enabled clover honey to be classified using probability plots. Clover honey was indicated by the presence of 3-methylpentanoic acid and an abundance of *ortho*-methoxyacetophenone of less than six. (See **Figure 3.16** as well as **Structures XVI** and **XVII**).

This enabled 35/43 samples to be classified correctly. Two tawari samples were incorrectly classified as clover. Non-conforming samples were re-tested. Results did not change except for one of the tawari samples, which no longer met the criteria for clover honey.

*Proportion of samples classified correctly: 35/43.*

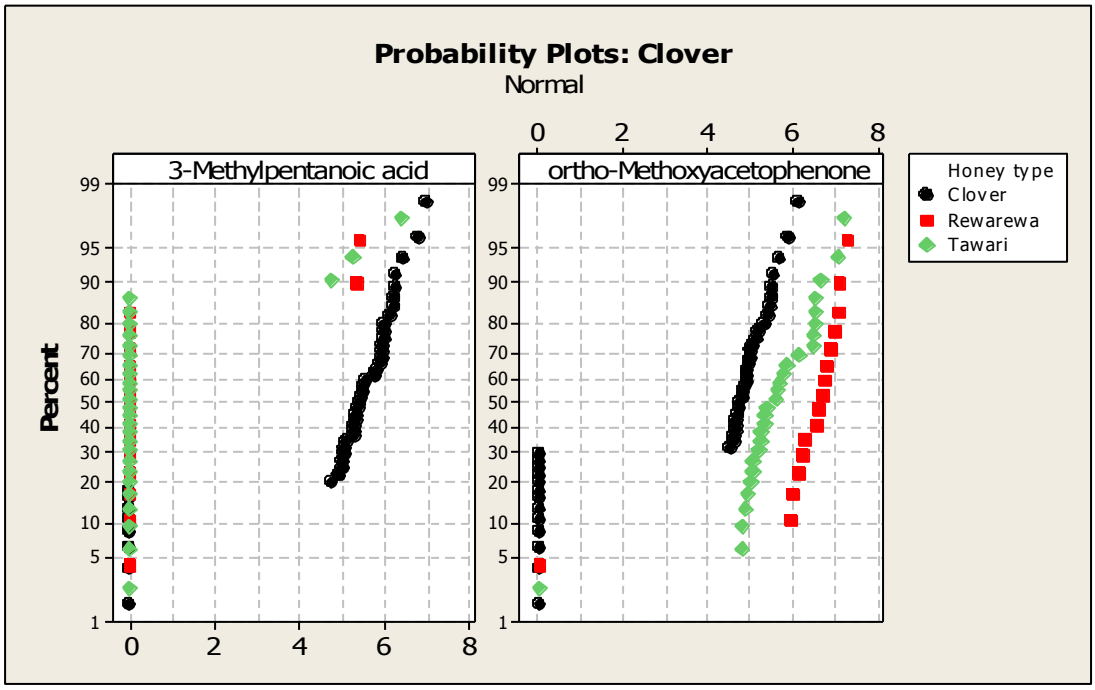
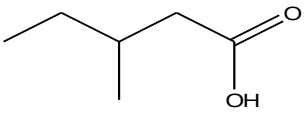
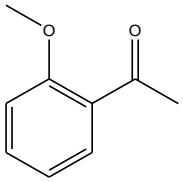



**Figure 3.16: Probability plots for 3-methylpentanoic acid and *ortho*-methoxyacetophenone**




**Structure XVI: 3-Methylpentanoic acid**




**Structure XVII: *ortho*-Methoxyacetophenone**

**Figure 3.17** shows a chromatogram of clover honey with important compounds labelled.

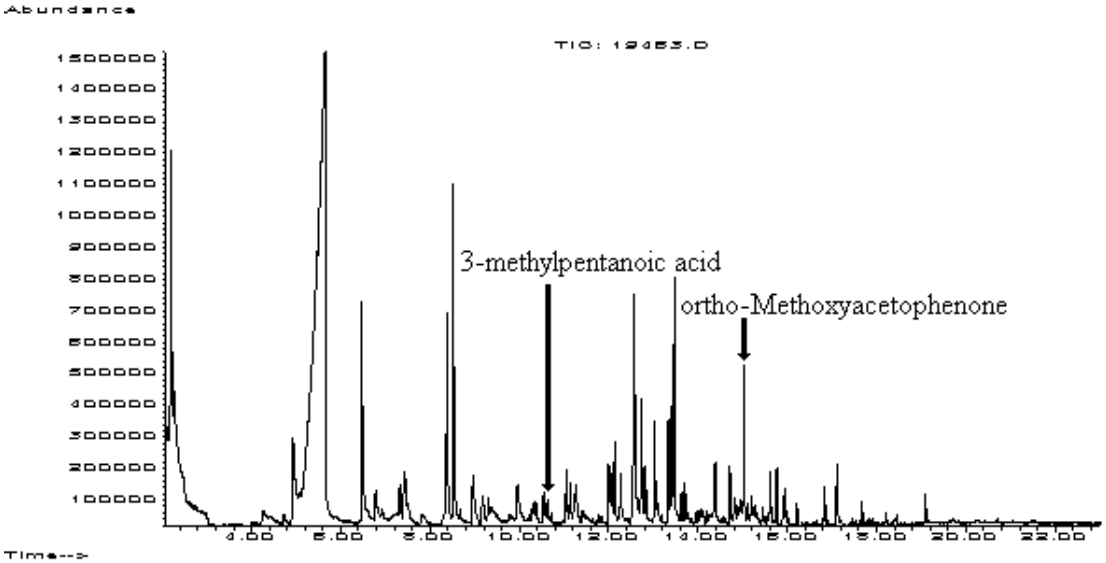



**Figure 3.17: Chromatogram of clover honey showing 3-methylpentanoic acid and *ortho*-methoxyacetophenone**

*Ortho*-methoxyacetophenone has been found in New Zealand willow, clover, manuka and kanuka honey by liquid-liquid extraction. [42, 43] It was found in Portuguese polyfloral honey using SPME-GC-MS.[18]

3-Methylpentanoic acid has not previously been reported as a volatile constituent of honey, although it has been found in palm wine[44] and roasted coffee.[45] It has been reported that it contributes (along with two other volatile compounds) to the sweaty odour of snake fruit.[46]

### 3.3.9 Tawari

Tawari honey was marked by the presence of 2-ethylhexanoic acid, or an abundance of *ortho*-methoxyacetophenone of less than six (see **Figure 3.18** as well as **Structures XVIII** and **XVII**).

Using this method, 24/26 samples were classified correctly. Three rewarewa samples were incorrectly classified as tawari. The anomalous samples were re-tested; one of the anomalous samples was found to meet the criteria for tawari honey and two of them were found not to meet the criteria.

*Proportion of samples classified correctly: 22/26* (Due to the two tawari samples that were classified as clover)

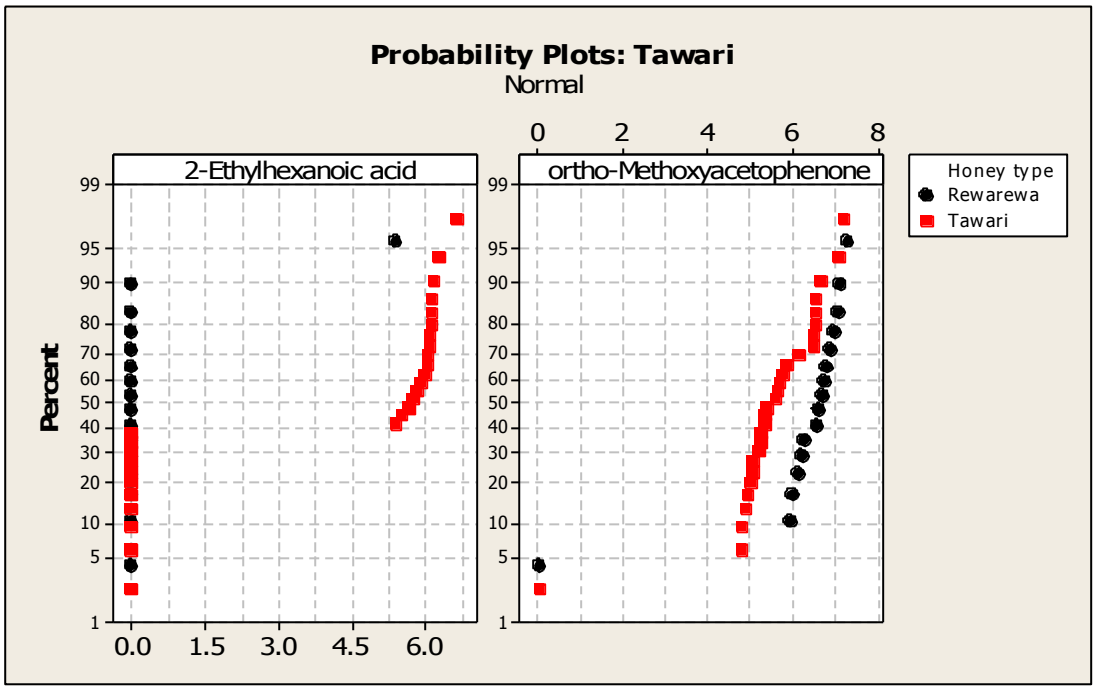
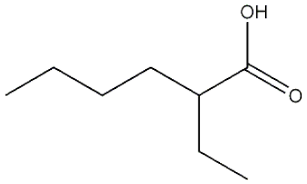



**Figure 3.18: Probability plots for 2-ethylhexanoic acid and *ortho*-methoxyacetophenone**




**Structure XVIII: 2-Ethylhexanoic acid**

**Figure 3.19** shows the chromatogram of a sample of tawari honey, with 2-ethylhexanoic acid and *ortho*-methoxyacetophenone labelled.

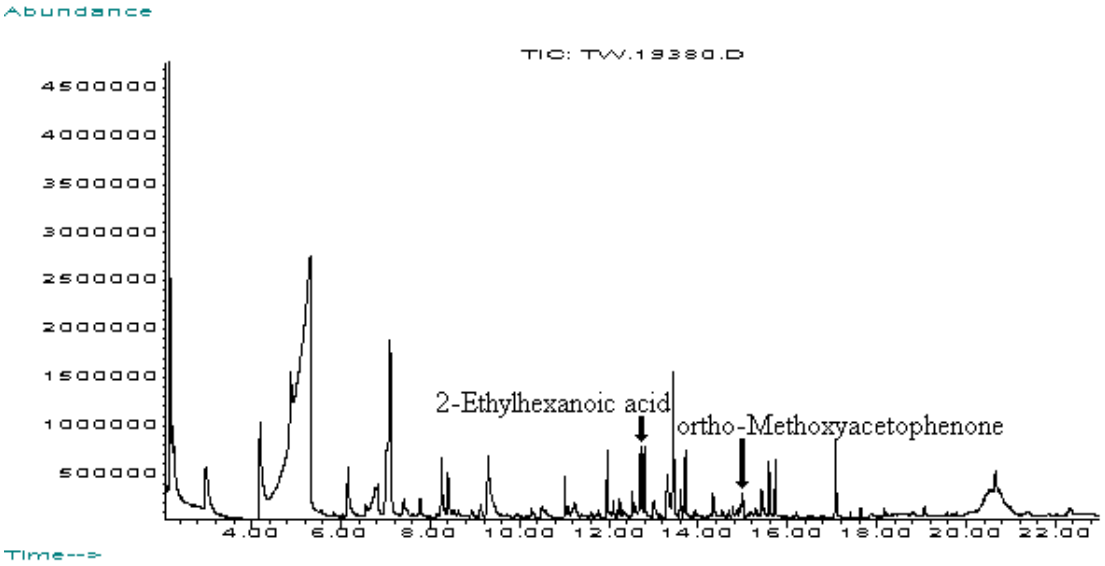



**Figure 3.19: Chromatogram of tawari honey showing 2-ethylhexanoic acid and *ortho*-methoxyacetophenone**

In a comparison of extraction techniques using Spanish rosemary honey, 2-ethylhexanoic acid was found using SPE and SDE but not liquid-liquid extraction.[3] 2-Ethylhexanoic acid has also been detected in honeys using SPME-GC-MS.[18, 25]

The occurrence of *ortho*-methoxyacetophenone in honey was discussed in **Section 3.3.8**.

### 3.3.10 Rewarewa

The remaining unifloral honey type studied was rewarewa. It was not possible to classify rewarewa honey by means of presence or absence of specific compounds, because there were no honey types left in the data set to compare it with. Rewarewa was left until last because it didn't exhibit any defining features. A typical rewarewa chromatogram is shown in **Figure 3.20**.

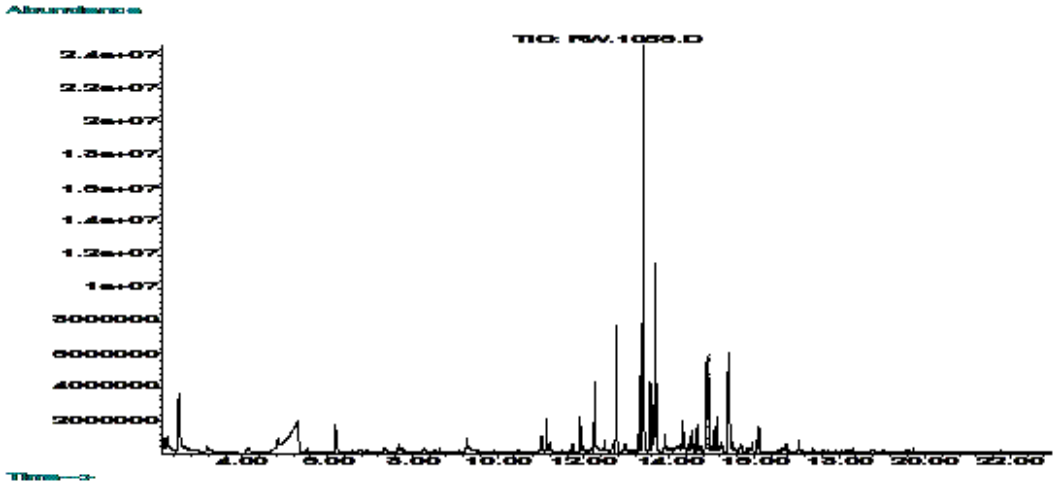



**Figure 3.20: Chromatogram of rewarewa honey**

# 3.4: Statistical Significance of Results

Student t statistics were calculated for compounds indicated by the probability plots as being important in determining floral source of honey.

The independent two-sample t-test was used (**Equations 3.1 and 3.2**), due to unequal sample sizes and unequal variances:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

**Equation 3.1: t-test equation**

Where:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

**Equation 3.2: Standard deviation used in t-test**

$\bar{x}$ is the mean of the abundances of selected compounds, *s* the standard deviation and *n* the number of samples.

There was a significant difference between the abundance of 3-methylbut-2-enal in pohutukawa honey and all other honey types ($p < 0.05$). That is, $t_{calculated}$ was greater than $t_{tabulated}$. The same was true for (*E*)-cinnamaldehyde. However for dimethyl sulfoxide – the third compound important in identifying pohutukawa honey – there was a significant difference between its abundance in pohutukawa honey and all other honey types ($p < 0.05$) except for rata. This was not a problem because pohutukawa honey is identified by other compounds (3-methylbut-2-enal and (*E*)-cinnamaldehyde) which are not found in rata honey.

Pohutukawa honey was removed from the test set and t-tests were performed in a sequential manner for the remaining honey types. Compounds of importance were shown to have a significant difference in their abundance for all honey types ($p < 0.05$), with only three exceptions:

- There was no significant difference ($p < 0.05$) in the abundance of 1-(2-methoxyphenyl)ethanol in honeydew and rewarewa honey; also between honeydew and tawari honey.
- There was no significant difference ($p < 0.05$) in the abundance of *ortho*-

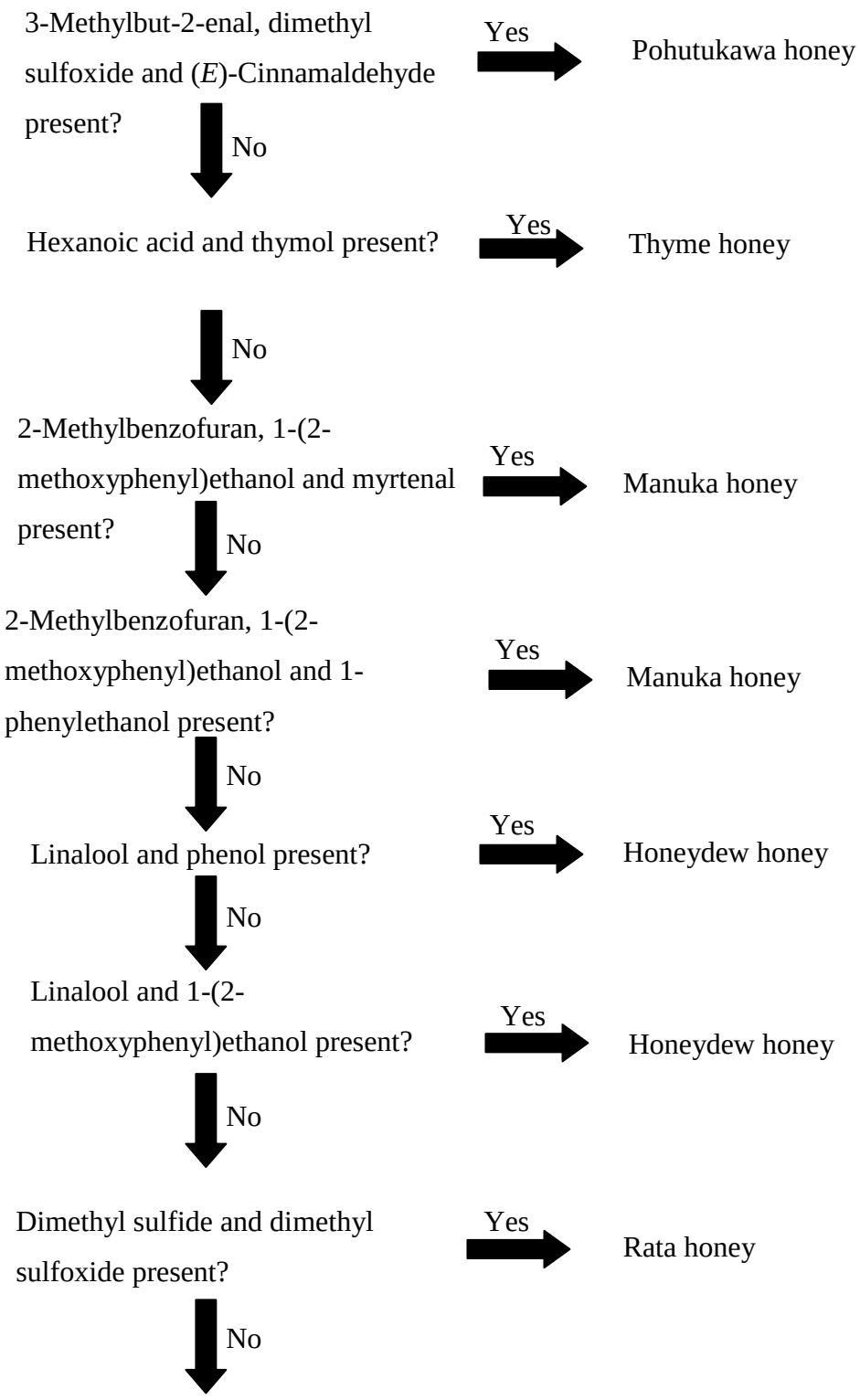
methoxyacetophenone in tawari and rewarewa honey.

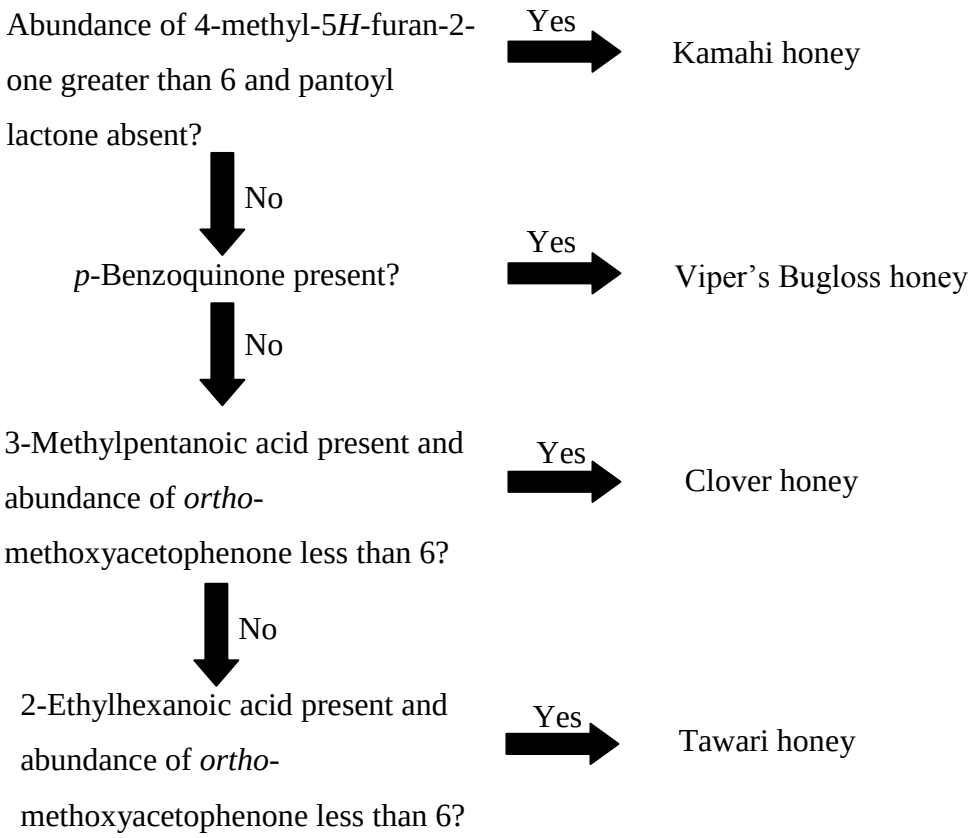
Because honeydew and tawari honeys have more than one compound important in identifying them, this was not thought to be a problem.

## 3.5: Summary

This method proved to be useful in classifying unifloral pohutukawa, thyme, manuka, honeydew, rata, kamahi, viper's bugloss, clover and tawari honeys. 91% of samples were correctly classified by this method.

**Flowchart Showing Classification of Pohutukawa, Thyme, Manuka, Honeydew, Rata, Kamahi, Viper's Bugloss, Clover and Tawari Honeys**




Abundance of 4-methyl-5*H*-furan-2-one greater than 6 and pantoyl lactone absent?
Yes
Kamahi honey
No
*p*-Benzoquinone present?
Yes
Viper's Bugloss honey
No
3-Methylpentanoic acid present and abundance of *ortho*-methoxyacetophenone less than 6?
Yes
Clover honey
No
2-Ethylhexanoic acid present and abundance of *ortho*-methoxyacetophenone less than 6?
Yes
Tawari honey

# 3.6 References


1. Escriche, I.; Visquert, M.; Juan-Borrás, M.; Fito, P., Influence of simulated industrial thermal treatments on the volatile fractions of different varieties of honey. *Food Chemistry* **2009,** 112, 329-338.

2. Cajka, T.; Hajslova, J.; Cochran, J.; Holadova, K.; Klimankova, E., Solid phase microextraction-comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry for the analysis of honey volatiles. *Journal of Separation Science* **2007,** 30, (4), 534-546.

3. Castro-Vazquez, L.; Perez-Coello, M. S.; Cabezudo, M. D., Analysis of volatile compounds of rosemary honey. Comparison of different extraction techniques. *Chromatographia* **2003,** 57, (3-4), 227-233.

4. Alissandrakis, E.; Kibaris, C.; Tarantilis, P.; Harizanis, P.; Polissiou, M., Flavour compounds of Greek cotton honey. *Journal of the Science of Food and Agriculture* **2005,** 85, (9), 1444-1452.

5. Jerkovic, I.; Mastelic, J.; Marijanovic, Z.; Klein, Z.; Jelic, M., Comparision of hydrodistillation and ultrasonic solvent extraction for the isolation of volatile compounds from two unifloral honeys of *Robinia pseudoacacia* L. and *Castanea sativa* L. *Ultrasonics Sonochemistry* **2007,** 14, 750-756.

6. Castro-Vazquez, L.; Diaz-Maroto, M. C.; Perez-Coello, M. S., Aroma composition and new chemical markers of Spanish citrus honeys. *Food Chemistry* **2007,** 103, 601-606.

7. Castro-Vazquez, L.; Diaz-Maroto, M. C.; Perez-Coello, M. S., Volatile composition and contribution to the aroma of Spanish honeydew honeys. Identification of a new chemical marker. *Journal of Agricultural and Food Chemistry* **2006,** 54, (13), 4809-4813.

8. Bonvehi, J. S.; Coll, F. V., Flavour index and aroma profiles of fresh and processed honeys. *Journal of the Science of Food and Agriculture* **2003,** 83, 275-282.

9. Guyot, C.; Bouseta, A.; Scheirman, V.; Collin, S., Floral origin markers of chestnut and lime tree honeys. *Journal of Agricultural and Food Chemistry* **1998,** 46, (2), 625-633.

10. Alissandrakis, E.; Tarantilis, P.; Harizanis, P.; Polissiou, M., Comparison of the volatile composition in thyme honeys from several origins in Greece. *Journal of Agricultural and Food Chemistry* **2007,** 55, 8152-8157.

11. Castro-Vazquez, L.; Diaz-Maroto, M. C.; Guchu, E.; Perez-Coello, M. S., Analysis of volatile compounds of eucalyptus honey by solid phase extraction followed by gas chromatography coupled to mass spectrometry. *European Food Research and Technology* **2006,** 224, 27-31.

12. Bahaffi, S. O.; Al-Lihaibi, S. S., Determination of volatile organic compounds in local honey by gas chromatography-mass spectrometer. *Communications. Faculty of Sciences, University of Ankara. Series B* **2005,** 51, (2), 1-12.

13. Guyot Declerck, C.; Renson, S.; Bouseta, A.; Collin, S., Floral quality and discrimination of *Lavandula stoechas, Lavandula augustifolia,* and *Lavandula augustifolia* x *latifolia* honeys. *Food Chemistry* **2002,** 79, 453-459.

14. Verzera, A.; Campisi, S.; Zappala, M.; Bonaccorsi, I., SPME-GC-MS analysis of honey volatile components for the characterization of different floral origin. *American Laboratory* **2001**, 18-21.

15. Shimoda, M.; Wu, Y.; Osajima, Y., Aroma compounds from aqueous solution of haze (*Rhus succedanea*) honey determined by adsorptive column chromatography. *Journal of Agricultural and Food Chemistry* **1996,** 44, (12), 3913-3918.

16. Adamczyk, S.; Lazaro, R.; Perez-Arquillue, C.; Conchello, P.; Herrera, A., Evaluation of residues of essential oil components in honey after different anti-varroa treatments. *Journal of Agricultural and Food Chemistry* **2005,** 53, (26), 10085-10090.

17. Goodwin, R. M.; Taylor, M. A.; McBrydie, H. M.; Cox, H. M. *Development of technologies for the control of varroa. Ninth quarterly report to National Beekeepers Association.*; 2007.

18. Pontes, M.; Marques, J. C.; Camara, J. S., Screening of volatile composition from Portuguese multifloral honeys using headspace solid-phase microextraction-gas chromatography-quadrupole mass spectrometry. *Talanta* **2007,** 74, 91-103.

19. Pino, J. A.; Marbot, R.; Delgado, A.; Zumarraga, C.; Sauri, E., Volatile constituents of propolis from honey bees and stingless bees from Yucatan. *Journal of Essential Oil Research* **2006,** 18, (1), 53-56.

20. Jerkovic, I.; Mastelic, J.; Marijanovic, Z., A variety of volatile compounds as markers in unifloral honey from Dalmatian sage (*Salvia officinalis* L.). *Chemistry & Biodiversity* **2006,** 3, (12), 1307-1316.

21. Guyot, C.; Scheirman, V.; Collin, S., Floral origin markers of heather honeys: *Calluna vulgaris* and *Erica arborea*. *Food Chemistry* **1999,** 64, (1), 3-11.

22. Rowland, C. Y.; Blackman, A. J.; D'Arcy, B. R.; Rintoul, G. B., Comparison of organic extractives found in leatherwood (*Eucryphia lucida*) honey and leatherwood flowers and leaves. *Journal of Agricultural and Food Chemistry* **1995,** 43, 753-763.

23. Tan, S. T.; Wilkins, A. L.; Holland, P. T.; McGhie, T. K., Extractives from New Zealand unifloral honeys. 2. Degraded carotenoids and other substances from heather honey. *Journal of Agricultural and Food Chemistry* **1989,** 37, (5), 1217-1221.

24. Tananaki, C.; Thrasyvoulou, A.; Giraudel, J. L.; Montury, M., Determination of volatile characteristics of Greek and Turkish pine honey samples and their classification by using Kohonen self organising maps. *Food Chemistry* **2007,** 101, 1687-1693.

25. Odeh, I.; Abu-Lafi, S.; Dewik, H.; Al-Najjar, I.; Imam, A.; Dembitsky, V. M.; Hanus, L. O., A variety of volatile compounds as markers in Palestinian honey from *Thymus capitatus*, *Thymelaea hirsuta*, and *Tolpis virgata*. *Food Chemistry* **2007,** 101, (4), 1393-1397.

26. de la Fuente, E.; Martínez-Castro, I.; Sanz, J., Characterization of Spanish unifloral honeys by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2005,** 28, (9-10), 1093-1100.

27. Peña, R.; Barciela, J.; Herrero, C.; García-Martín, S., Solid-phase microextraction gas chromatography-mass spectrometry determination of monoterpenes in honey. *Journal of Separation Science* **2004,** 27, (17-18), 1540-1544.

28. Piasenzotto, L.; Gracco, L.; Conte, L., Solid phase microextraction (SPME) applied to honey quality control. *Journal of the Science of Food and Agriculture* **2003,** 83, (10), 1037-1044.

29. Soria, A.; Martínez-Castro, I.; Sanz, J., Analysis of volatile composition of honey by solid phase microextraction and gas chromatography-mass spectrometry. *Journal of Separation Science* **2003,** 26, (9-10), 793-801.

30. Moreira, R. F. A.; Trugo, L. C.; Pietroluongo, M.; De Maria, C. A. B., Flavor composition of cashew (*Anacardium occidentale*) and marmeleiro (*Croton* Species) honeys. *Journal of Agricultural and Food Chemistry* **2002,** 50, (26), 7616-7621.

31. Radovic, B. S.; Careri, M.; Mangia, A.; Musci, M.; Gerboles, M.; Anklam, E., Contribution of dynamic headspace GC-MS analysis of aroma compounds to authenticity testing of honey. *Food Chemistry* **2001,** 72, 511-520.

32. D'Arcy, B. R.; Rintoul, G. B.; Rowland, C. Y.; Blackman, A. J., Composition of Australian honey extractives. 1. Norisoprenoids, monoterpenes, and other natural volatiles from blue gum (*Eucalyptus leucoxylon*) and yellow box (*Eucalyptus melliodora*) honeys. *Journal of Agricultural and Food Chemistry* **1997,** 45, 1834-1843.

33. Alissandrakis, E.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M., Aroma investigation of unifloral Greek citrus honey using solid-phase microextraction coupled to gas chromatographic-mass spectrometric analysis. *Food Chemistry* **2007,** 100, (1), 396-404.

34. Wilkins, A. L.; Lu, Y.; Tan, S. T., Extractives from New Zealand honeys. 4. Linalool derivatives and other components from nodding thistle (*Carduus nutans*) honey. *Journal of Agricultural and Food Chemistry* **1993,** 41, (6), 873-878.

35. de la Fuente, E.; Sanz, M. L.; Martinez-Castro, I.; Sanz, J.; Ruiz-Matute, A. I., Volatile and carbohydrate composition of rare unifloral honeys from Spain. *Food Chemistry* **2007,** 105, 84-93.

36. Perez, R. A.; Sanchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L., Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography-mass spectrometry. *Journal of Agricultural and Food Chemistry* **2002,** 50, (9), 2633-2637.

37. Supriyadi; Suhardi; Suzuki, M.; Yoshida, K.; Muto, T.; Fujita, A.; Watanabe, N., Changes in the volatile compounds and in the chemical and physical properties of snake fruit (*Salacca edulis* Reinw) Cv. *Pondoh* during maturation. *Journal of Agricultural and Food Chemistry* **2002,** 50, (26), 7627-7633.

38. Ishikawa, M.; Ito, O.; Ishizaki, S.; Kurobayashi, Y.; Fujita, A., Solid-phase aroma concentrate extraction (SPACE): a new headspace technique for more sensitive analysis of volatiles. *Flavour and Fragrance Journal* **2004,** 19, (3), 183-187.

39. Boido, E.; Lloret, A.; Medina, K.; Farina, L.; Carrau, F.; Versini, G.; Dellacassa, E., Aroma composition of *Vitis vinifera* Cv. Tannat: the typical red wine from Uruguay. *Journal of Agricultural and Food Chemistry* **2003,** 51, (18), 5408-5413.

40. Helsper, J. P. F. G.; Buecking, M.; Muresan, S.; Blaas, J.; Wietsma, W. A., Identification of the volatile component(s) causing the characteristic foxy odor in various cultivars of *Fritillaria imperialis* L. (Liliaceae). *Journal of Agricultural and Food Chemistry* **2006,** 54, (14), 5087-5091.

41. Wilkins, A. L.; Tan, S.; Molan, P. C., Extractable organic substances from New Zealand unifloral viper's bugloss (*Echium vulgare*) honey. *Journal of Apicultural Research* **1995,** 34, (2), 73-78.

42. Tan, S. T.; Holland, P. T.; Wilkins, A. L.; Molan, P. C., Extractives from New Zealand honeys. 1. White clover, manuka and kanuka unifloral honeys. *Journal of Agricultural and Food Chemistry* **1988,** 36, (3), 453-460.

43. Tan, S. T.; Wilkins, A. L.; Holland, P. T.; McGhie, T. K., Extractives from New Zealand honeys. 3. Unifloral thyme and willow honey constituents. *Journal of Agricultural and Food Chemistry* **1990,** 38, (9), 1833-1838.

44. Lasekan, O.; Buettner, A.; Christlbauer, M., Investigation of important odorants of palm wine (*Elaeis guineensis*). *Food Chemistry* **2007,** 105, (1), 15-23.

45. Gonzalez-Rios, O.; Suarez-Quiroz, M. L.; Boulanger, R.; Barel, M.; Guyot, B.; Guiraud, J.-P.; Schorr-Galindo, S., Impact of ecological post-harvest processing on coffee aroma: II. Roasted coffee. *Journal of Food Composition and Analysis* **2007,** 20, (3-4), 297-307.

46. Wijaya, C. H.; Ulrich, D.; Lestari, R.; Schippel, K.; Ebert, G., Identification of potent odorants in different cultivars of snake fruit [*Salacca zalacca* (Gaert.) Voss] using gas

chromatography-olfactometry. *Journal of Agricultural and Food Chemistry* **2005,** 53, (5), 1637-1641.

# Chapter Four: Unsupervised Learning

The aim of this section was to uncover natural groupings of honeys by studying the structure of the data.

## 4.1 Introduction

Pattern recognition techniques fall into two classes; supervised and unsupervised. Supervised pattern recognition is used when class membership (for example, floral source of honey) is known. Unsupervised pattern recognition techniques do not use class membership, and can be useful (even when class membership is known, as in the present study) for revealing natural groupings within the data.[1]

The pattern recognition techniques discussed in this chapter are hierarchical cluster analysis (HCA) and principal component analysis (PCA). These are both unsupervised techniques.

HCA links observations together based on the interpoint distance between all samples. The output is in the form of a dendrogram, which shows clusters of similar samples.[1] The *y* axis shows similarity between clusters and the *x* axis labels observations.

PCA constructs combinations of variables, called components. Each component explains a percentage of variation. Principal components can be plotted against each other, to show the structure of data in a small number of dimensions.[1] Such graphs are called score plots.

# 4.2 HCA and PCA Performed on Entire Data Set

## 4.2.1 Hierarchical Cluster Analysis

HCA was performed on the data set obtained in **Section 2.3**. This contained information about the abundances of 55 compounds in ten unifloral honey types. Ward linkage and Euclidean distances (**Equation 4.1**) were used to cluster observations.

The Euclidean distance between samples $x$ and y is calculated in nvars (nvars = number of variables) dimensions as:

$$\text{Distance} = \sqrt{[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_{nvars} - y_{nvars})^2]}$$

**Equation 4.1: Euclidean distance**

Where $x_i$ and $y_i$ are the coordinates of samples $x$ and y in the $i^{th}$ dimension of the row space, and $i$ ranges from 1 to nvars.[1]

Ward linkage is one of several measures used to cluster observations in HCA. Each observation* is treated as a cluster, and similar clusters are joined together until they are all linked.[1]

* An observation is the data obtained from each individual chromatogram.

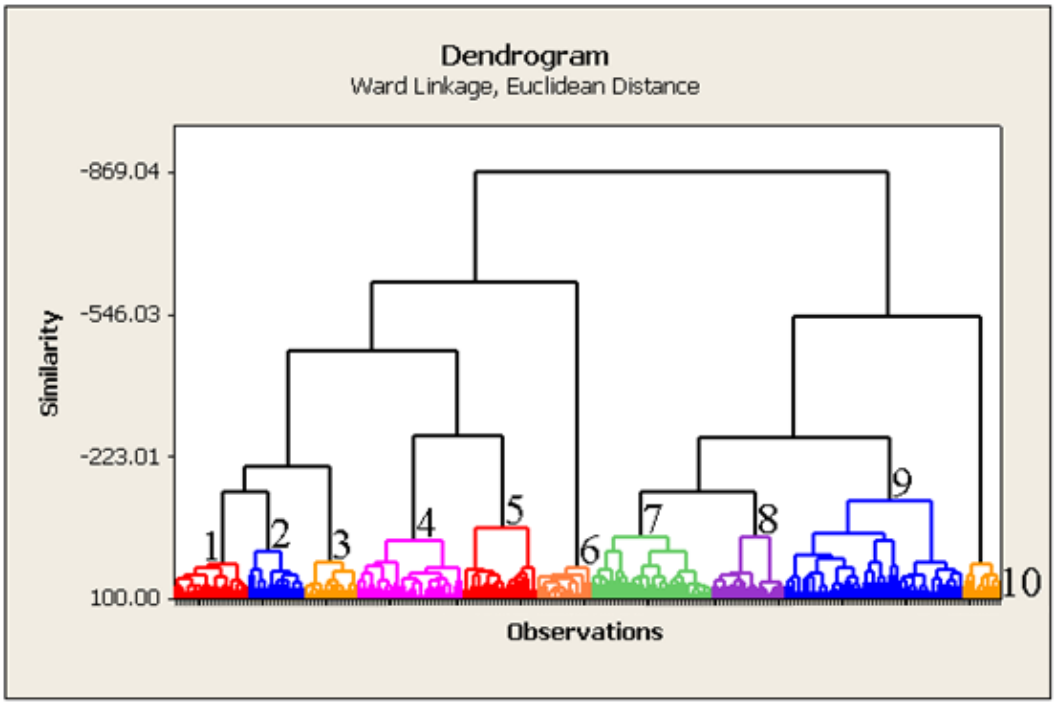



**Figure 4.1: Dendrogram showing clusters of honey samples**

A dendrogram of all honey samples showed ten main clusters, numbered in **Figure 4.1**. Clusters one and three consisted mostly of clover samples, with small amounts of viper's bugloss. Cluster two was 95% viper's bugloss. Cluster four was mostly honeydew samples, and only manuka samples were found in cluster five. Cluster six contained only pohutukawa samples. Cluster seven was a mixture of tawari, rewarewa and a small amount of manuka. Cluster eight was a mixture of tawari and rewarewa samples – the two were difficult to distinguish by HCA. Cluster nine was large, and made up of mostly kamahi and rata samples. Cluster ten consisted only of thyme samples.

The dendrogram indicated that discrimination between honey samples should be possible, as there were obviously natural groupings amongst the honey samples.

### 4.2.2 Principal Component Analysis

Principal component analysis was performed. The output of PCA is a set of equations called 'principal components'. Each component explains a continuously decreasing proportion of variance. The number of components reported is equal to the number required to explain all of the variation in the data set. By studying the principal components, compounds with a larger coefficient can be identified as explaining more variation than others.

A score plot is also reported; this is a plot of the first two principal components against each other. Separate classes (honey types) are highlighted on the plot. The first two components explain more variance than successive individual components. If the first two components explain a large proportion of variance, it is expected that classes on the score plot will be well defined.

Principal component analysis was performed. The first two components explained 30% of variation. 1,1'-Bicyclopentyl, 4-methyl-5*H*-furan-2-one, hortrienol, nonan-4-one and 2,6-dimethyl-2,7-octadiene-1,6-diol made a strong positive contribution to the first component. Lilac aldehyde C, 2,6,6-trimethylcyclohexane-1,4-dione and 1-(2-methoxyphenyl)ethanol made a strong positive contribution to the second component. Hexanoic acid, lilac aldehyde B and thymol made a strong negative contribution to the second component.

To explain 90% of variation, 26 components were needed. A limitation of PCA was that all compounds were used, even though some were probably not useful in explaining floral source. To remove compounds manually would be to risk biasing the result.

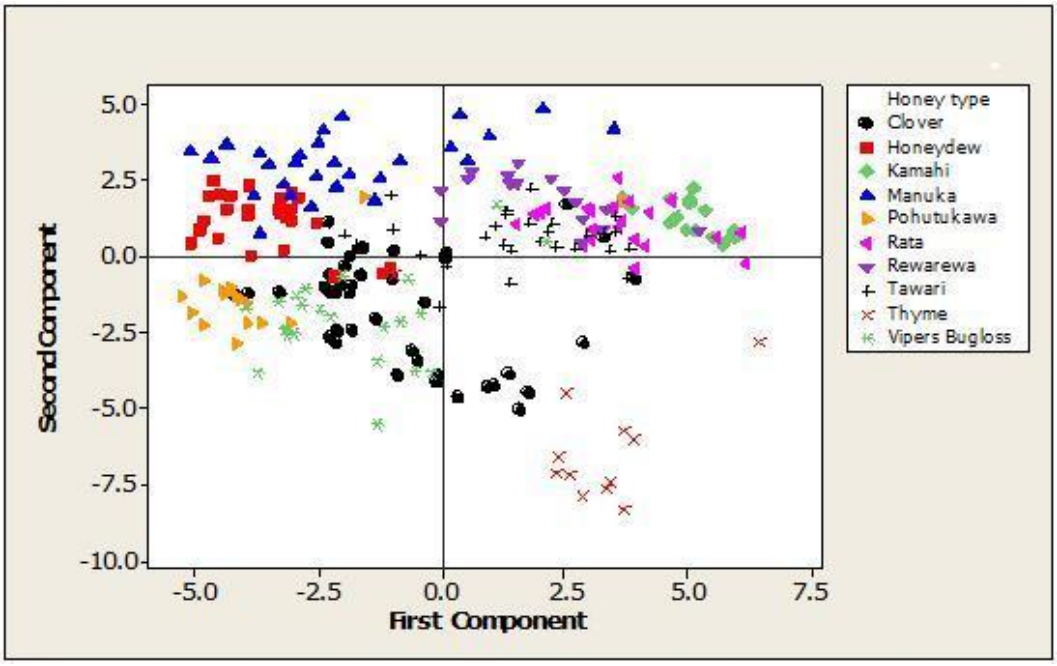



**Figure 4.2: Score plot showing honey types plotted against first two principal components**

The score plot (**Figure 4.2**) showed that in general there was a distinction between honey types, although it was not well defined. Given that the first and second component only explained 30% of variance combined, this was not surprising. The plot showed thyme honey to be well defined; pohutukawa, honeydew and

manuka less so. There was considerable overlap between kamahi, rata, tawari and rewarewa honeys. These are all bush honeys, and a better visualisation was obtained by labelling them as such (**Figure 4.3**). This shows the bush honeys grouped in the top right corner of the score plot.

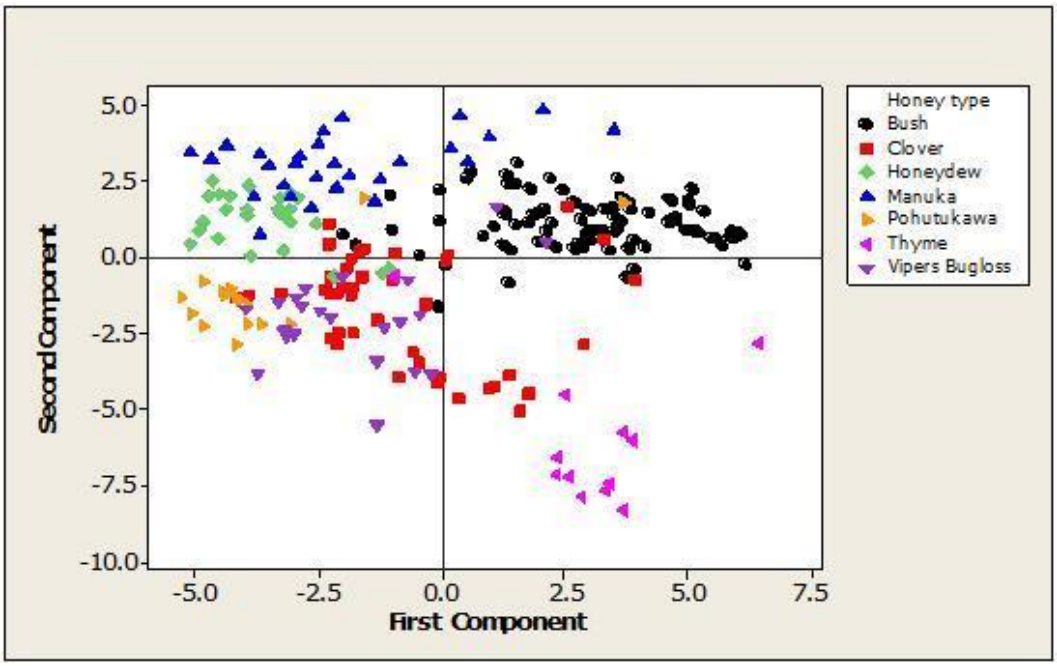



**Figure 4.3: Score plot showing honey types plotted against first two principal components. Rata, kamahi, tawari and rewarewa are labelled as bush honey.**

## 4.3 HCA and PCA Performed on Data Set Consisting Only of Selected Compounds

It was possible to obtain a clearer representation of data via HCA and PCA by using only the compounds selected by the probability plots (**Chapter Three**) as being useful for discrimination. These compounds, and the honey type(s) between which they distinguished, are shown in **Table 4.1.**

**Table 4.1: Compounds Important in Discriminating Between Honey Types**

| **Compound** | **Honey type identified** |
|---|---|
| Dimethyl sulfide | Rata |
| 3-Methylbut-2-enal | Pohutukawa |
| Dimethyl sulfoxide | Pohutukawa, rata |
| *p*-Benzoquinone | Viper's Bugloss |
| 3-Methylpentanoic acid | Clover |
| Hexanoic acid | Thyme |
| Phenol | Honeydew |
| Pantoyl lactone | Kamahi |
| 1-Phenylethanol | Manuka |
| 4-Methyl-5*H*-furan-2-one | Kamahi |
| 2-Ethylhexanoic acid | Tawari |
| Linalool | Honeydew |
| 2-Methylbenzofuran | Manuka |
| Myrtenal | Manuka |
| (*E*)-Cinnamaldehyde | Pohutukawa |
| Thymol | Thyme |
| 1-(2-Methoxyphenyl)ethanol | Manuka, honeydew |
| *ortho*-Methoxyacetophenone | Clover, tawari |

### 4.3.1 Hierarchical Cluster Analysis

HCA on the refined data set yielded a dendrogram with much clearer separations between honey types. Ward Linkage and Euclidean distances were used to cluster observations.


**Figure 4.4: Dendrogram showing clusters of observations**

A dendrogram of all honey samples (**Figure 4.4**) showed that cluster one was the most poorly defined, with less than half of the samples consisting of rewarewa honey, and the rest a mixture. This was because with the probability plot method, rewarewa was the 'default' category; that is, the category that samples were assigned to if they did not fit the criteria for any of the other floral types. Rewarewa honey did not have any defining compounds; only the absence of others.

The other nine clusters were well defined. Cluster two contained mostly tawari samples, cluster three mostly clover samples, and cluster four consisted of mainly kamahi samples. Cluster five was nearly all rata samples, and cluster six was completely made up of viper's bugloss honeys. Clusters seven and eight also consisted of only one floral source; thyme and honeydew respectively. Cluster nine was mainly manuka samples and cluster ten consisted only of pohutukawa samples.

### 4.3.2 Principal Component Analysis

Like HCA on this data set, PCA also gave better results than the original data. The first two components described 40% of variation, and only ten components were needed to explain 90% of variance (compared with 26 previously).

2-Methylbenzofuran and 1-(2-methoxyphenyl)ethanol, compounds used in describing manuka honey, gave large positive contributions to the first component. Dimethyl sulfide, dimethyl sulfoxide and (*E*)-cinnamaldehyde, compounds used in identifying pohutukawa and rata honeys, gave large negative weightings to the first component.

3-Methylbut-2-enal, linalool, 2-methylbenzofuran, (*E*)-cinnamaldehyde and *ortho*-methoxyacetophenone all gave large positive contributions to the second component.

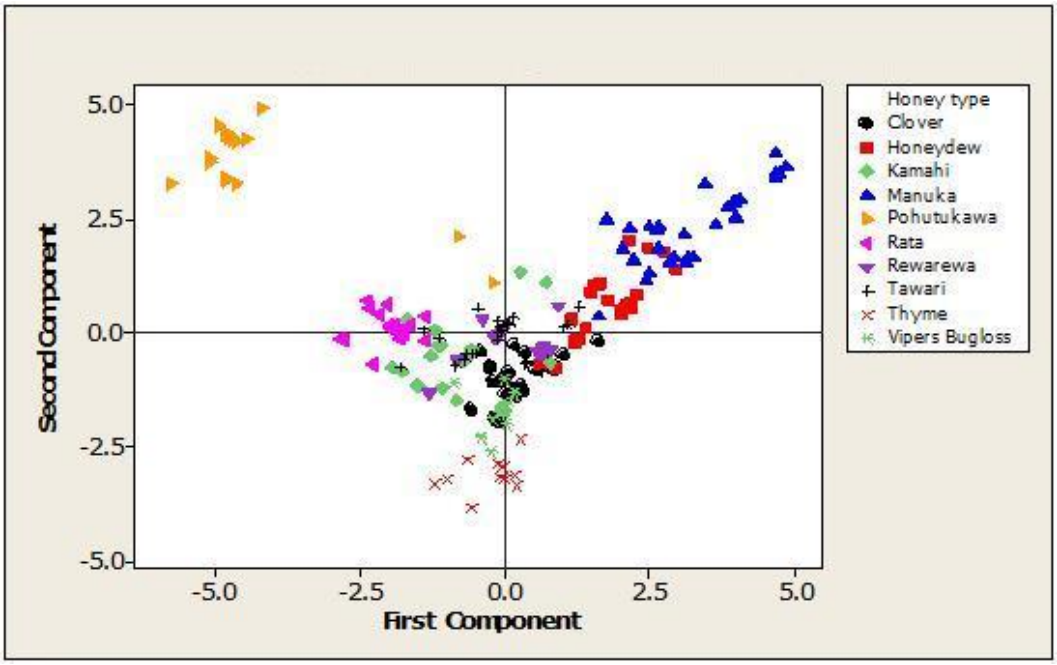



**Figure 4.5: Score plot showing honey types plotted against first two principal components**

A score plot (**Figure 4.5**) showed pohutukawa, thyme, manuka and rata to be well separated. There was considerable overlap between viper's bugloss, kamahi, clover, rewarewa and tawari honeys.

## 4.4 Discussion

Hierarchical cluster analysis and principal component analysis showed that it should be possible to satisfactorily distinguish between honey types using multivariate statistical techniques, because natural groupings of honey types could be observed; both in the dendrograms and score plots. This meant that it was appropriate to use supervised pattern recognition to build a model with which to classify honey samples.

Supervised learning techniques are explained in **Chapter Five**, and reveal to what extent discrimination between floral sources is possible.

### 4.4.1 Comment on Similarity Between Manuka and Honeydew Honeys

It was observed that groups of manuka and honeydew samples were next to each other on PCA score plots, and had a common compound used to identify them in the probability plot method: 1-(2-methoxyphenyl)ethanol. A possible explanation for such similarities is that scale insects that produce honeydew live on manuka plants, and a black sooty mould is often seen on plants inhabited by such insects. It is common for fungal particles of sooty mould to be found in manuka honey.[2, 3]

## 4.5 References


1. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., *Chemometrics A Practical Guide*. John Wiley & Sons, Inc.: New York, 1998.
2. Tan, S. T.; Holland, P. T.; Wilkins, A. L.; Molan, P. C., Extractives from New Zealand honeys. 1. White clover, manuka and kanuka unifloral honeys. *Journal of Agricultural and Food Chemistry* **1988,** 36, (3), 453-460.
3. Airborne's New Zealand Honey Collections. http://www.airborne.co.nz/manuka.htm (06/08/08),

# Chapter Five: Supervised Learning

The aim of using supervised learning techniques was to build a model that could successfully discriminate between honey samples, and be used for classification of future samples.

## 5.1 Introduction

The unsupervised learning techniques used in **Chapter Four** showed that honey data naturally falls into groups based on floral source; this meant that it was appropriate to use supervised learning techniques (those in which class membership is known) to discriminate between floral sources. Supervised methods have the potential to build a model that can be used to classify the floral source of future honey samples.[1]

This process is also known as data mining; the process of finding patterns in data automatically or semi-automatically.[2] In the present study, the software used for this application was Weka (Waikato Environment for Knowledge Analysis). Weka is a collection of machine learning algorithms and is capable of regression, classification, clustering, association rule mining and attribute selection.[2]

## 5.2 Weka Terminology

### 5.2.1 Input

Weka describes data sets in terms of 'instances' and 'attributes.' Instances are individual honey samples; attributes are features of instances, in this case the compounds that were looked for in honey samples. Spreadsheets of instances and attributes are loaded into Weka. A variety of filters are available to pre-process the data, although filters were not used in the present study.

### 5.2.2 Evaluation

To build models, a classifier (learning algorithm) is selected. Some classifiers can be used in combination with each other. The error rate (percentage of correct classifications) for the learning scheme is evaluated. However, if all data is used for building a model (training), the error will be overly optimistic. It is ideal to use some data for testing the model, and gaining a more realistic error estimate. Weka has several methods for dealing with this.

The method used in the present study was stratified 10-fold cross-validation. In this procedure, the data set is split into ten parts; proportions of classes are approximately the same as in the original data set.[2] One-tenth of the data is held back for testing, while the remaining nine-tenths are used for training. This process is performed ten times, with a different tenth of data held back for testing each time. The error rate is calculated for the test set each time, and at the end of the process the average error rate is calculated for the ten test sets. This procedure enables as much data to be used for training as possible, while still allowing a reasonable error estimate to be obtained.[2]

### 5.2.3 Output

Depending on the algorithm used, the output may take the form of a decision tree, or a set of equations that can be used to calculate the probability of each sample belonging to a particular class. A confusion matrix is always presented; an example of one is shown in **Figure 5.1**.

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
40  0  0  0  0  0  0  1  0  2 |  a = Clover
 0 26  0  0  0  0  0  0  0  0 |  b = Honeydew
 0  0 16  0  0  1  1  0  0  0 |  c = Kamahi
 0  0  0 28  0  0  0  0  0  0 |  d = Manuka
 0  0  0  0 16  1  0  0  0  0 |  e = Pohutukawa
 0  0  0  0  0 20  0  2  0  1 |  f = Rata
 0  0  0  0  0  1 14  1  0  0 |  g = Rewarewa
 2  2  0  1  0  1  0 21  0  1 |  h = Tawari
 0  0  0  0  0  0  0  1 11  0 |  i = Thyme
 6  0  0  0  0  0  0  0  0 17 |  j = Viper's Bugloss
```

**Figure 5.1: Example of a confusion matrix**

Numbers in the diagonal starting at the top left-hand corner of **Figure 5.1** moving towards the bottom right-hand corner are samples that have been assigned to the correct class. Row 'a' corresponds to clover honey. It can be seen that 40 samples

have been correctly classified as clover, one sample has been incorrectly classified as tawari (column 'h') and two samples have been incorrectly classified as viper's bugloss (column 'j'). Looking down column 'a', it can be seen that two tawari samples have been incorrectly classified as clover, and six viper's bugloss samples have been incorrectly classified as clover.

Weka also reports a number of statistics that give information about correct and false classifications that result after a model has been built. These statistics and their meanings are summarised in the **Table 5.1**.

**Table 5.1: Statistics Used to Evaluate the Performance of Learning Algorithms**[2]

| **Statistic** | **Description** |
|---|---|
| TP rate | True positive rate (should be close to one) |
| FP rate | False positive rate (should be close to zero) |
| Precision | = (Number of samples of a particular class correctly classified)/(Number of samples classified as that class) |
| Recall | = (Number of samples of a particular class correctly classified)/(Number of samples in that class) |
| F-measure | = (2 × recall × precision)/(recall + precision)<br>This is a single measure of performance. |
| ROC area | Receiver Operating Characteristic. This characterises the trade-off between true positive rate and false negative rate (should be close to one but not one exactly: this is an indication of data over-fitting). |

## 5.3 Method

Three data sets were tested; these consisted of data for all honey samples and differed in the number of attributes present. The first set contained all 55 compounds that were looked for in honey samples ('All data'). There were problems with identification for some compounds, so a second data set was obtained that excluded these compounds ('Manually selected data'). This set contained 37 attributes. The third data set ('Probability plot data') contained the 18 compounds that were used in the probability plots method (**Chapter Three**). See **Appendix 12** for a list compounds in the 'all data' set, and **Table 4.1** (section 4.3) for a list of compounds in the 'probability plot data' set. A list of compounds in the 'manually selected data' set is given below:

Dimethyl sulfide
3-Methylbut-2-enal
Dimethyl sulfoxide
1,3,5,7-Cyclooctatetraene
*p*-Benzoquinone
3-Methylpentanoic acid
Hexanoic acid
5-Methyloxolan-2-one
Phenol
Benzaldehyde
1,1'-Bicyclopentyl
2,2,4,4-Tetramethylcyclobutane-1,3-dione
Benzyl alcohol
Pantoyl lactone
Benzeneacetaldehyde
1-Phenylethanol
4-Methyl-5*H*-furan-2-one
*cis*-Linaloloxide
Acetophenone
2-Ethylhexanoic acid
Linalool oxide
Hortrienol
Linalool
Nonanal
2-Phenylethanol
2-Methylbenzofuran
2,6-Dimethyl-1,3,5,7-octatetraene
Benzoic acid
2,2,6-Trimethylcyclohexane-1,4-dione
2,6-Dimethyl-3,7-octadiene-2,6-diol
Terpineol
Myrtenal
(*E*)-Cinnamaldehyde
Thymol
1-(2-Methoxyphenyl)ethanol
*ortho*-Methoxyacetophenone
2,6-Dimethyl-2,7-octadiene-1,6-diol

Weka contains a multitude of learning algorithms that can be used in various combinations. Seven algorithms were selected and trialled on the three data sets. The algorithms were:

- J48
- LMT (Logistic model tree)
- Logistic
- Classification via Regression
- IBk (K-Nearest Neighbours)
- Attribute Selected Classifier
- AdaBoostM1

These are described in further detail below, grouped under the heading below which they are found in Weka.

### 5.3.1 Decision Trees

Two decision tree algorithms were trialled: J48 and LMT. J48 uses a divide-and-conquer approach to classify data. LMT builds decision trees with linear logistic regression models at nodes.[2]

### 5.3.2 Functions

The logistic function builds linear logistic regression models. Large coefficients are penalized; therefore over-fitting of data is discouraged.[2]

### 5.3.3 Lazy Classifiers

Lazy classifiers store all training instances then assign samples depending on what class their nearest neighbours belong to.[2] The IBk classifier was used, which is the algorithm for K-nearest neighbours (KNN). In KNN, Euclidean distances (**Equation 4.1, section 4.2.1**) are used to measure how close samples are to each other. To classify unknown samples, the distance is calculated between it and the training samples. The closest K samples are used for classification.[1] If the classes are well separated, K = 1 can be used. For more confidence, more nearest neighbours should be used (K should be set to a higher number).

### 5.3.4 Metalearning Algorithms

Metalearners turn simple classifiers into more powerful algorithms. This is often done by using combinations of algorithms.[2]

- Classification via regression binarizes the class and builds a regression model for each value.[2]
- The attribute selected classifier selects attributes before classification, so that dimensionality of the data is reduced.[2] It can be used in combination with any of the algorithms described above.
- AdaBoostM1 is a boosting method. Multiple algorithms are used and new models are encouraged to focus on instances that were incorrectly classified by earlier models. This is done by giving such instances a higher weight.[2] It can be used in combination with any of the algorithms described above except logistic, with which it is not compatible.

## 5.4 Results

**Table 5.2** gives a summary of classification rates for the various combinations of algorithms tested on the three data sets. Up to three algorithms could be used at one time.

**Table 5.2: Summary of Algorithm Performance on Three Data Sets**

| | | | **Percentage of correctly classified instances** | | |
|---|---|---|---|---|---|
| **Algorithm 1** | **Algorithm 2** | **Algorithm 3** | **All data** | **Manually selected data** | **Probability plot data** |
| J48 | | | 84.2 | 81.6 | 82.9 |
| LMT | | | 90.6 | **89.8** | **88.5** |
| Logistic | | | 87.6 | 80.8 | 80.8 |
| IBk* | | | 82.0 (K = 10) | 86.8 (K = 5) | 88.5 (K = 5) |
| Classification via regression | | | 83.8 | 83.8 | **88.0** |
| Attribute selected classifier | J48 | | 82.0 | 80.3 | 77.8 |
| Attribute selected classifier | LMT | | 85.9 | 85.9 | 82.9 |
| Attribute selected classifier | Logistic | | 76.5 | 68.8 | 79.5 |
| Attribute selected classifier | IBk* | | 81.6 (K = 5) | 85.4 (K = 5) | 82.9 (K = 5) |
| Attribute selected classifier | Classification via regression | | 82.9 | 82.9 | 83.8 |
| AdaBoostM1 | J48 | | 87.1 | **88.0** | **88.5** |
| AdaBoostM1 | LMT | | 90.6 | **88.0** | **89.7** |
| AdaBoostM1 | IBk* | | 81.6 (K = 10) | 84.6 (K = 5) | 86.8 (K = 10) |
| AdaBoostM1 | Classification via regression | | 87.1 | **89.3** | 83.8 |
| Attribute selected classifier | AdaBoostM1 | J48 | 85.0 | 84.6 | 82.5 |
| Attribute selected classifier | AdaBoostM1 | LMT | 85.4 | 84.6 | 82.5 |
| Attribute selected classifier | AdaBoostM1 | IBk* | 79.5 (K = 10) | 82.5 (K = 10) | 82.9 (K = 5) |
| Attribute selected classifier | AdaBoostM1 | Classification via regression | 86.3 | 85.0 | 83.8 |

* IBk classifier was tested with K = 5 and K = 10. The higher classification rate of the two was reported.

Important points relating to **Table 5.2**:

- The highest classification rate (90.6%) was obtained using the LMT algorithm on the training set 'all data.' It was likely that over-fitting occurred due to the large number of attributes used in building the model. Also, there were identification problems for some compounds in this data set. Therefore it would be better to use the manually selected data or probability plot data for training.
- Performance of models decreased when the attribute selected classifier was employed.
- In general, an increase in classification rate was seen when the AdaBoostM1 classifier was used; however it was possible that data over-fitting occurred due to the use of multiple algorithms.
- Outputs for the values highlighted in bold were studied in detail. It was decided that the LMT algorithm on the manually selected data gave the best results in terms of true and false classification. Statistics relating to this model (**Table 5.3**) and the confusion matrix (**Figure 5.2**) are presented below. Statistics and confusion matrices for the other models highlighted in bold are displayed in **Appendix 13**.

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
37  1  0  0  0  0  0  4  0  1 |  a = Clover
 1 25  0  0  0  0  0  0  0  0 |  b = Honeydew
 0  0 17  0  0  0  1  0  0  0 |  c = Kamahi
 1  0  1 26  0  0  0  0  0  0 |  d = Manuka
 0  0  0  0 15  2  0  0  0  0 |  e = Pohutukawa
 0  0  1  0  0 21  0  1  0  0 |  f = Rata
 1  0  0  0  0  0 15  0  0  0 |  g = Rewarewa
 1  1  0  1  0  0  1 24  0  0 |  h = Tawari
 0  0  0  0  0  0  0  1 11  0 |  i = Thyme
 4  0  0  0  0  0  0  0  0 19 |  j = Viper's Bugloss
```

**Figure 5.2: Confusion matrix for LMT algorithm on manually selected data**

**Table 5.3: Evaluation Statistics for LMT Algorithm on Manually Selected Data**

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.86 | 0.042 | 0.822 | 0.86 | 0.841 | 0.957 | Clover |
| 0.962 | 0.01 | 0.926 | 0.962 | 0.943 | 0.998 | Honeydew |
| 0.944 | 0.009 | 0.895 | 0.944 | 0.919 | 0.971 | Kamahi |
| 0.929 | 0.005 | 0.963 | 0.929 | 0.945 | 0.988 | Manuka |
| 0.882 | 0 | 1 | 0.882 | 0.938 | 0.998 | Pohutukawa |
| 0.913 | 0.009 | 0.913 | 0.913 | 0.913 | 0.996 | Rata |
| 0.938 | 0.009 | 0.882 | 0.938 | 0.909 | 0.942 | Rewarewa |
| 0.857 | 0.029 | 0.8 | 0.857 | 0.828 | 0.965 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.982 | Thyme |
| 0.826 | 0.005 | 0.95 | 0.826 | 0.884 | 0.958 | Viper's Bugloss |

The LMT model was relatively successful in classifying tawari and viper's bugloss honeys, which were problematic when other algorithms were used. 24/28 tawari samples and 19/23 viper's bugloss samples were correctly classified when the LMT algorithm was used; a better result than was obtained by the majority of algorithms. As shown in **Table 5.3**, the rate of true positives for all honey types was generally high and the rate of false positives low – zero for pohutukawa and thyme. High rates of correct classification were reflected by the high precision, recall, F-measure and ROC area values.

## 5.5 Conclusions

Weka has great potential as a tool for future classification of honey samples. It is a rapid technique, and once models are obtained they are easy to apply to new data sets. Because of the tendency to over-fit data it is recommended that a larger number of honey samples are tested before an algorithm is selected for future classification. The LMT algorithm gave a high classification rate and relatively few false classifications.

# 5.6 References

1. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., *Chemometrics A Practical Guide*. John Wiley & Sons, Inc.: New York, 1998.

2. Witten, I. H.; Frank, E., *Data Mining Practical Learning Tools and Techniques*. Elsevier Inc.: San Francisco, 2005; Vol. 2.

# Chapter Six: Discussion and Conclusions

The aim of this research was to investigate the use of headspace solid-phase microextraction (SPME) coupled with GC-MS together with statistical analysis of the resulting profiles of volatile compounds as a potential tool for the rapid analysis of the floral origin of New Zealand unifloral honeys. This resulted in two methods of classification that discriminated between unifloral clover, honeydew, kamahi, manuka, pohutukawa, rata, rewarewa, tawari, thyme and viper's bugloss honeys with approximately 90% success.

## 6.1 Summary of Findings

**Chapter One:** A review of publications covering SPME-GC-MS analyses of honey and any statistical methods used in these studies. Other methods for determining the botanical origin of honey were discussed.

**Chapter Two:** A SPME-GC-MS method was developed to extract and analyse the volatile compounds from honey. This proved to be a useful and rapid tool for fingerprinting the volatiles in honey. Once 340 SPME-GC-MS chromatograms had been obtained from ten types of unifloral honeys, peak areas of compounds were integrated and this data was subjected to a number of statistical analyses.

**Chapter Three:** Probability plots were used in discriminating between floral sources. This method enabled compounds important in determining floral source to be identified. 91% of samples were classified correctly; however because honey classes were eliminated once they had been classified, one honey type (rewarewa) could not be classified because there were no other honey classes left to compare it with.

Honeys were also classified by applying learning algorithms to the data extracted from SPME-GC-MS chromatograms. This constituted supervised learning (statistical analysis in which class membership is known), so unsupervised

learning techniques (those in which class membership is not known) had to be applied first to ensure that the results obtained from supervised learning were reliable.

**Chapter Four:** The unsupervised techniques hierarchical cluster analysis and principal component analysis were used. It was found that a natural structure (based on floral source) existed within the data obtained from the honey samples; therefore the use of supervised learning techniques was appropriate.

**Chapter Five:** Learning algorithms offered by Weka software were used to discriminate between honey types. It was found that the logistic model tree algorithm afforded the best classification overall, with 89.8% of samples classified correctly.

## 6.2 Discussion

Probability plots and the logistic model tree algorithm classified samples with approximately the same overall success rate: 91% and 89.8% respectively. Correct classifications for individual honey types are shown in **Table 6.1**.

**Table 6.1: Comparison of Samples Correctly Classified by the Probability Plots Method and LMT Algorithm. The number of correctly classified samples for each honey type is displayed. Numbers in brackets next to the honey class are the total number of samples in that category.**

| Honey Type | Probability Plots | Logistic Model Tree |
|---|---|---|
| Clover (43) | 35 | 37 |
| Honeydew (26) | 26 | 25 |
| Kamahi (18) | 17 | 17 |
| Manuka (28) | 27 | 26 |
| Pohutukawa (17) | 15 | 15 |
| Rata (23) | 23 | 21 |
| Rewarewa (16) | - | 15 |
| Tawari (28) | 22 | 24 |
| Thyme (12) | 11 | 11 |
| Viper's Bugloss (23) | 19 | 19 |

The same classification rate was obtained for kamahi, pohutukawa, thyme and viper's bugloss using the probability plots method and LMT algorithm. Rewarewa was unable to be classified by the probability plots method. Better classifications were obtained with the probability plots method for honeydew, manuka and rata (although differences were not significant). However LMT achieved better classifications for clover and tawari; honey types that were poorly classified using probability plots.

The probability plots method gave a set of compounds that can identify floral source; this method provided a simple description of honey types. The major limitation of this method was that it could be over-fitted; more samples should be tested to check the error rate.

The LMT algorithm when applied by Weka gave a rapid classification of floral source. It had greater scope for future work since it could classify all honey types, unlike the probability plot method which could not classify rewarewa samples. As with the probability plot method, a large number of honey samples should be tested to check model performance before it can be used routinely.

This study focused on unifloral honeys (classified by pollen counting), therefore it is not known what results would be obtained for polyfloral honeys. Further study could attempt mixing unifloral honeys to determine if the mixed floral source could be classified. Results could be compared with polyfloral honey samples with the same composition as defined by pollen analysis.

## 6.3 Previous Studies

This is the first study on New Zealand honeys using analysis of volatile compounds and statistical classification, and the first to use SPME-GC-MS, probability plots and learning algorithms such as LMT to classify honeys. Several studies have analysed the volatile composition of honeys (using SPME-GC-MS) with different statistical tests. A study of five Spanish honey types used canonical discriminant analysis and stepwise discriminant analysis to classify honeys with 100% success.[1] Principal component analysis and discriminating factor analysis were used to classify six Swiss honey types with 98% of samples classified correctly.[2] An Argentinean study of five honey types used hierarchical cluster analysis, stepwise discriminant analysis and K-nearest neighbours, resulting in 93% correct classification.[3]

These three studies all resulted in a higher correct classification rate than those obtained in the present study; however fewer honey types were studied than the ten used in this study so classification would have been less complicated.

## 6.4 Conclusions

The floral source of New Zealand unifloral honeys could be determined using headspace SPME-GC-MS and multivariate statistical analysis. Probability plots and the logistic model tree algorithm provided by Weka software both enabled honeys to be classified with approximately 90% success.

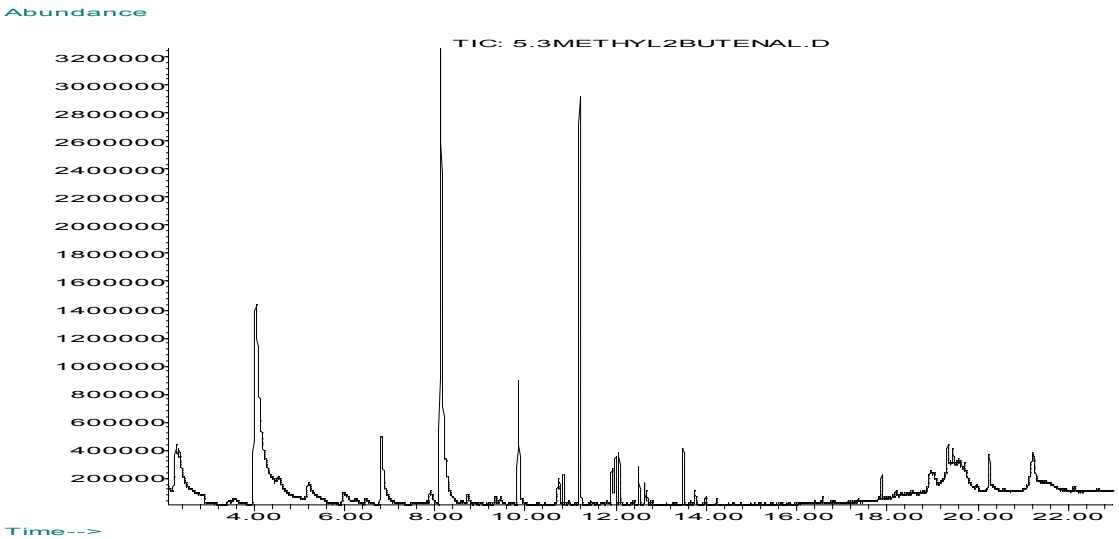
## 6.5 Suggestions for Future Work

Results of this study suggest that routine commercial testing of floral source is a future possibility; results may be very useful in combination with near infrared spectroscopy, pollen counting or analysis of non-volatile compounds such as phenolics. However due to the potential problem of model over-fitting it is suggested that at least a further ten samples from each floral source are tested, and the resulting error rate from the probability plot and LMT methods calculated. Additionally, polyfloral samples should be tested and their interference in results evaluated.

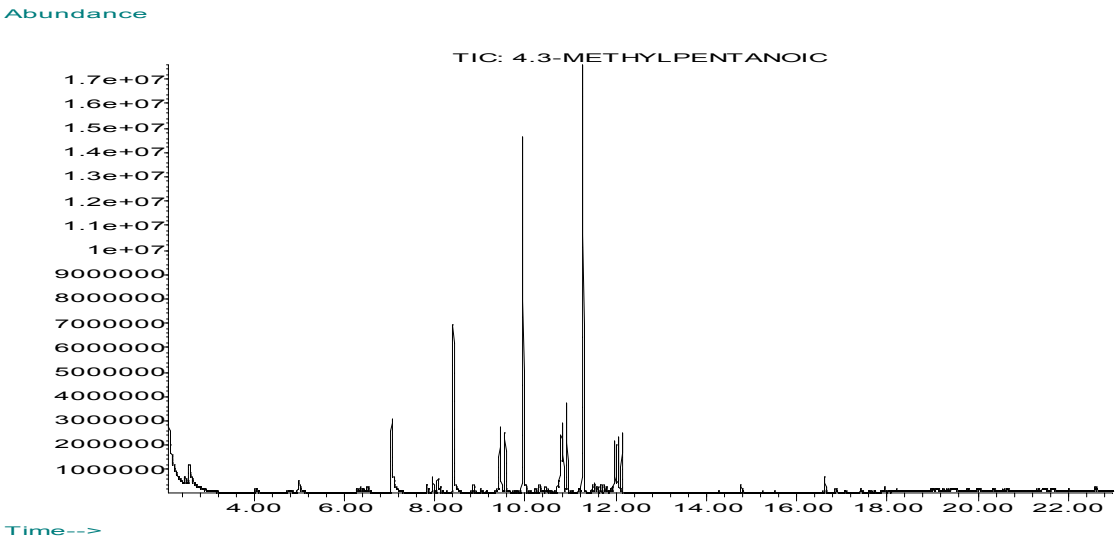
## 6.6 References


1. Perez, R. A.; Sanchez-Brunete, C.; Calvo, R. M.; Tadeo, J. L., Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography-mass spectrometry. *Journal of Agricultural and Food Chemistry* **2002,** 50, (9), 2633-2637.
2. Ampuero, S.; Bogdanov, S.; Bosset, J.-O., Classification of unifloral honeys with an MS-based electronic nose using different sampling modes: SHS, SPME and INDEX. *European Food Research and Technology* **2004,** 218, 198-207.
3. Baroni, M. V.; Nores, M. L.; Diaz, M. D. P.; Chiabrando, G. A.; Fassano, J. P.; Costa, C.; Wunderlin, D. A., Determination of volatile organic compound patterns characteristic of five unifloral honey by solid-phase microextraction-gas chromatography-mass spectrometry coupled to chemometrics. *Journal of Agricultural and Food Chemistry* **2006,** 54, (19), 7235-7241.

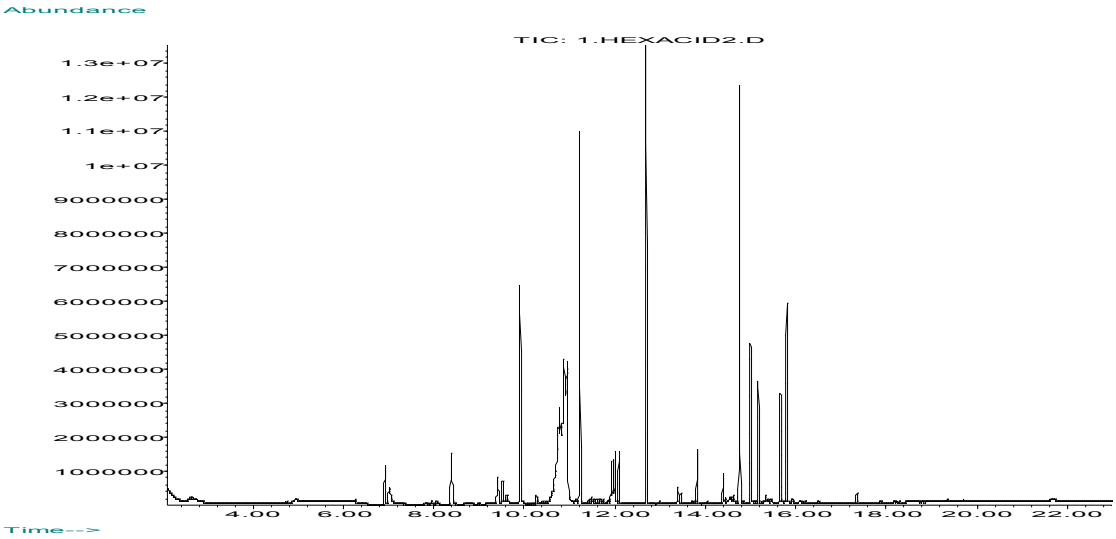
# Appendix 1: Chromatograms of Standards



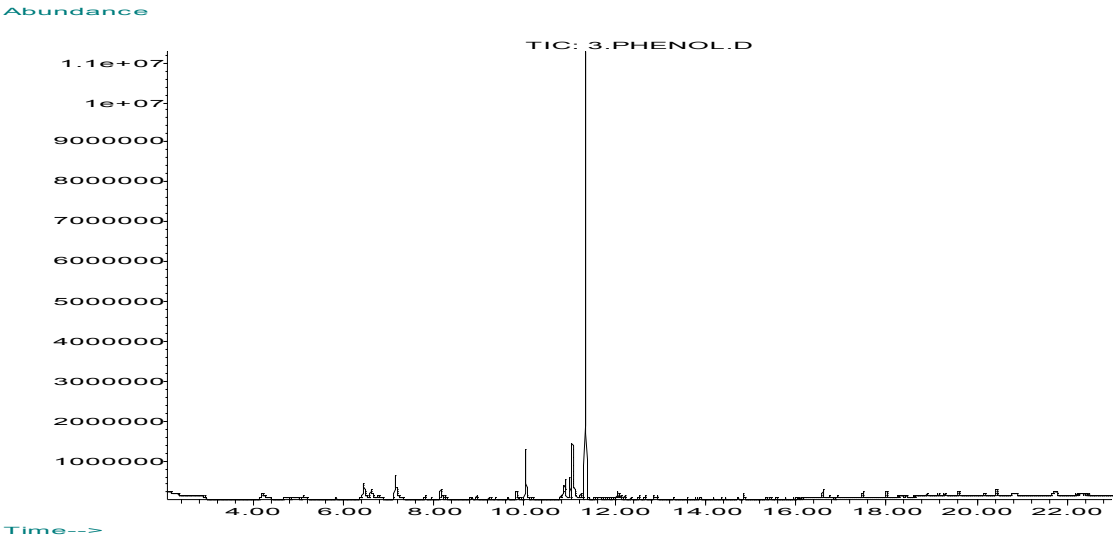

**3-Methylbut-2-enal**



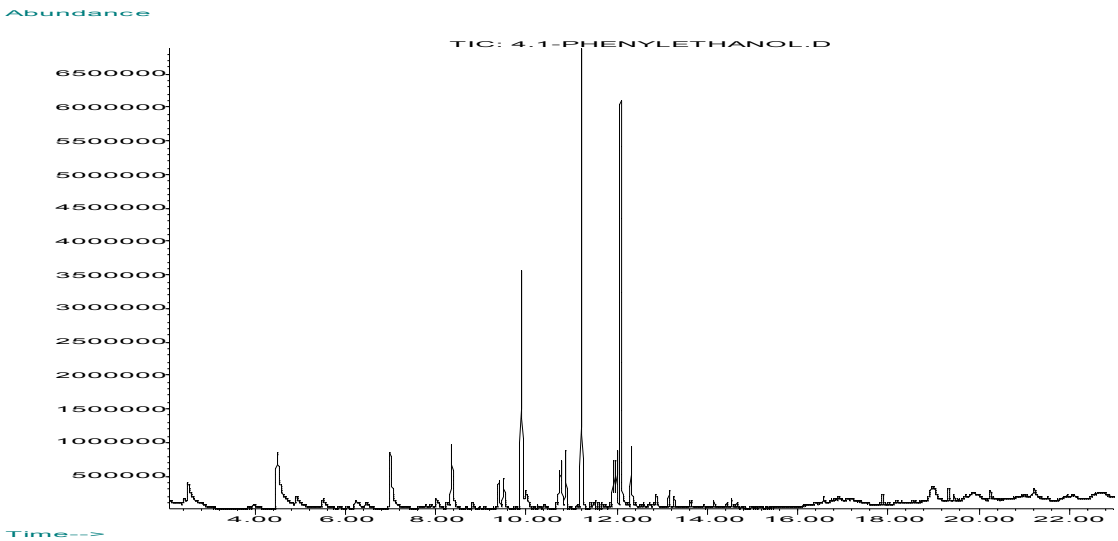

**3-Methylpentanoic acid**



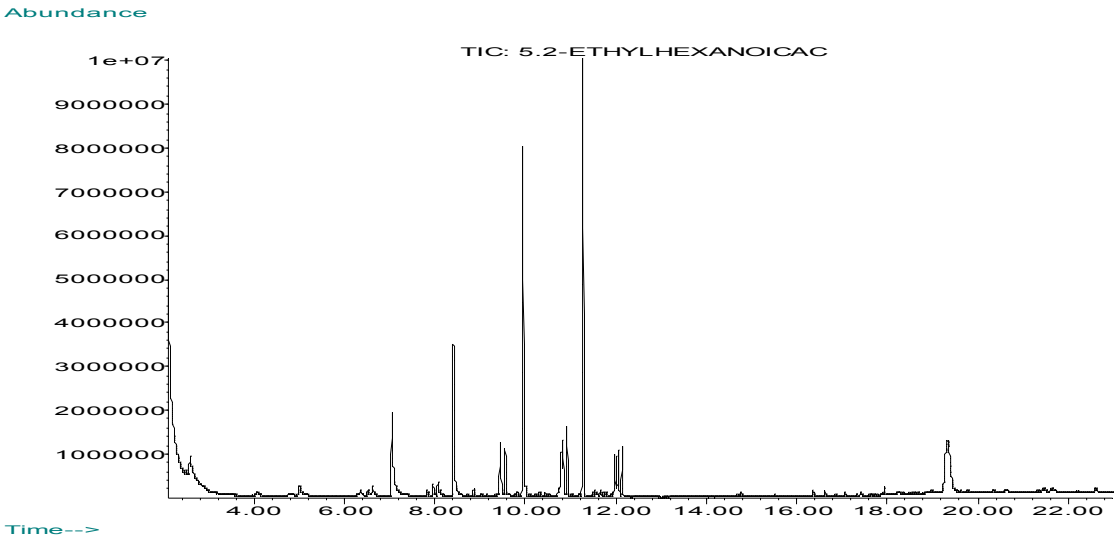

**Hexanoic acid**



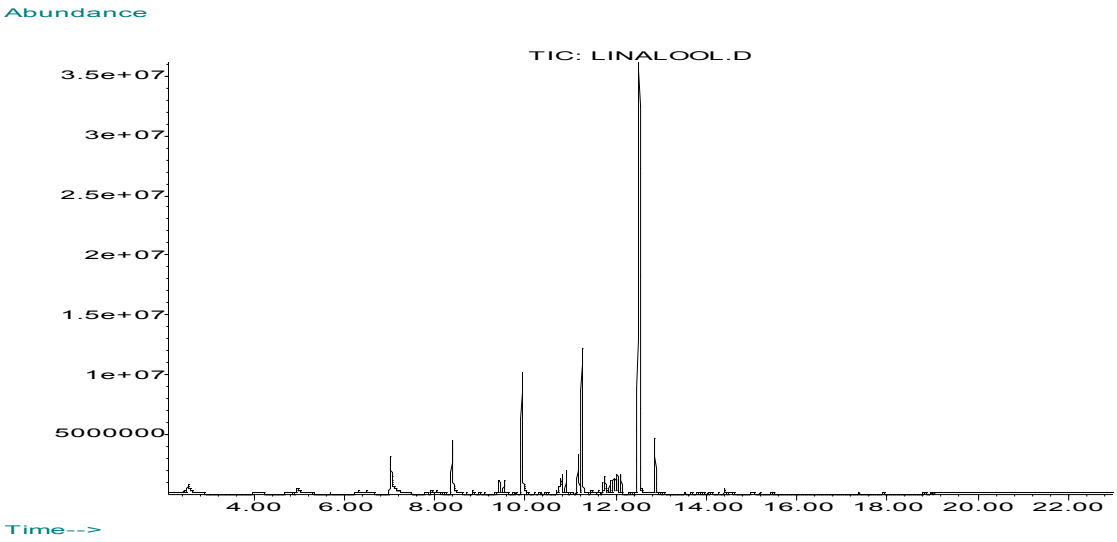

**Phenol**



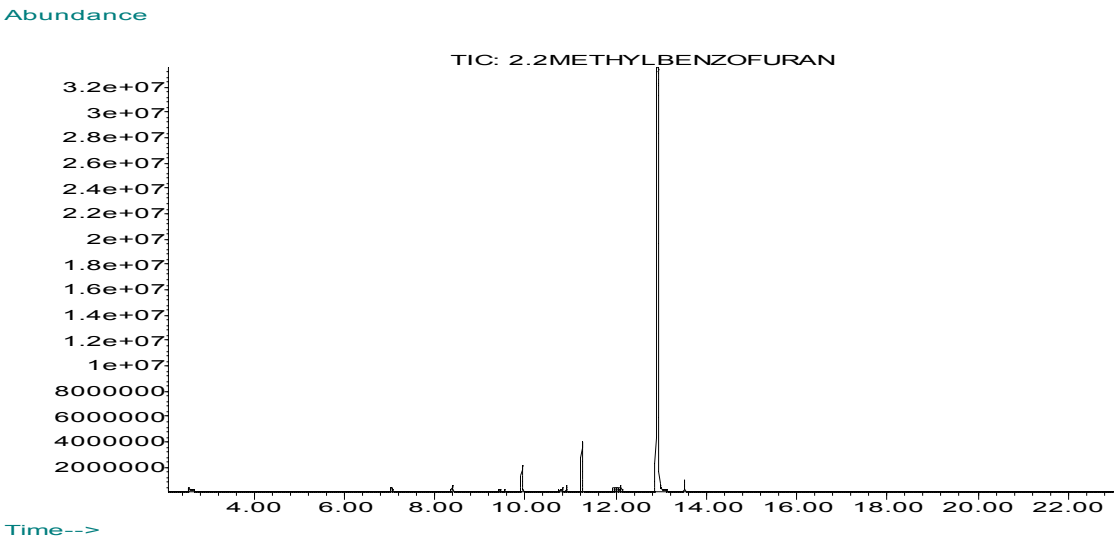

**1-Phenylethanol**



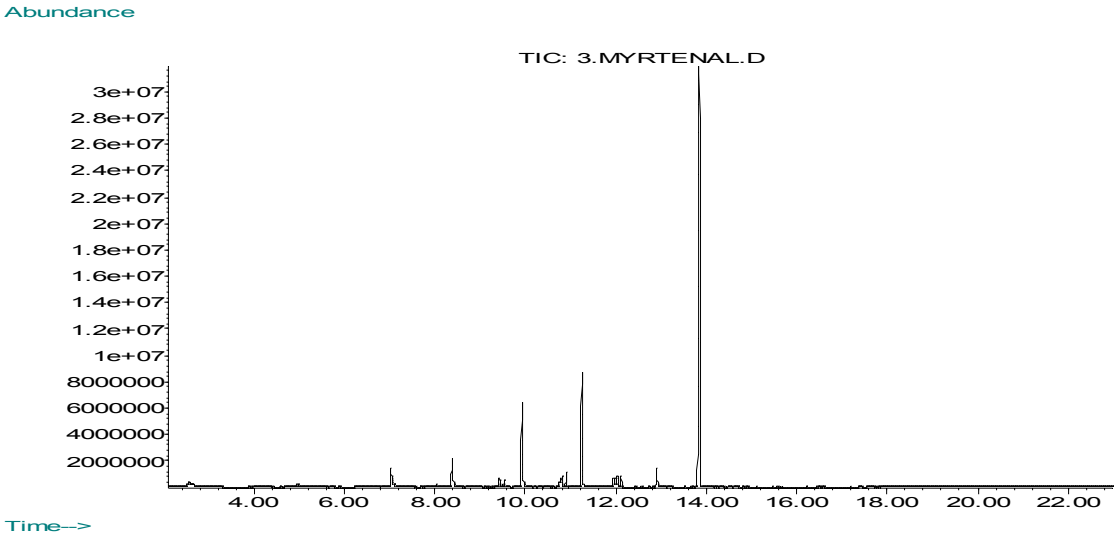

**2-Ethylhexanoic acid**



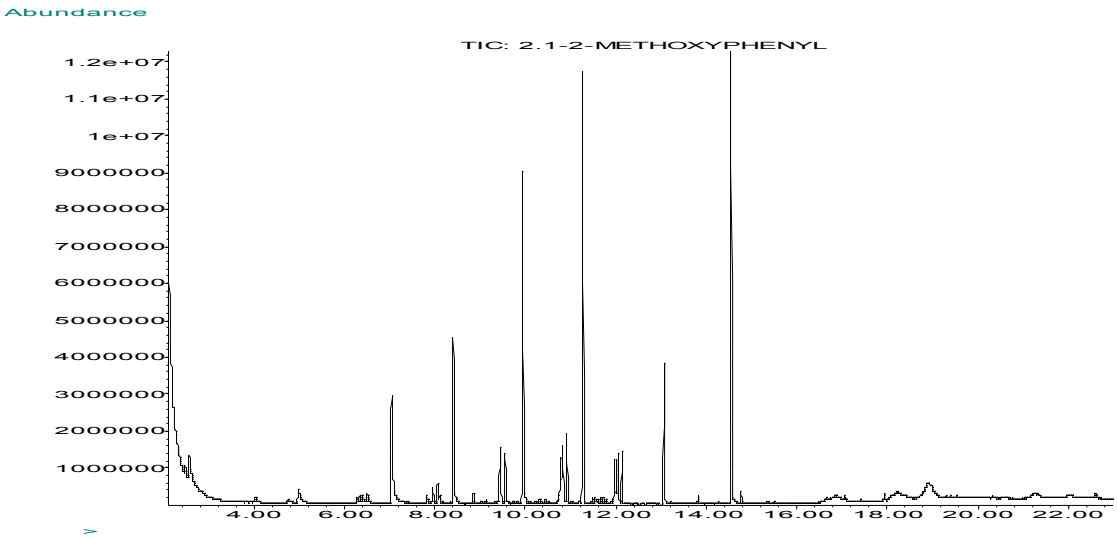

**Linalool**



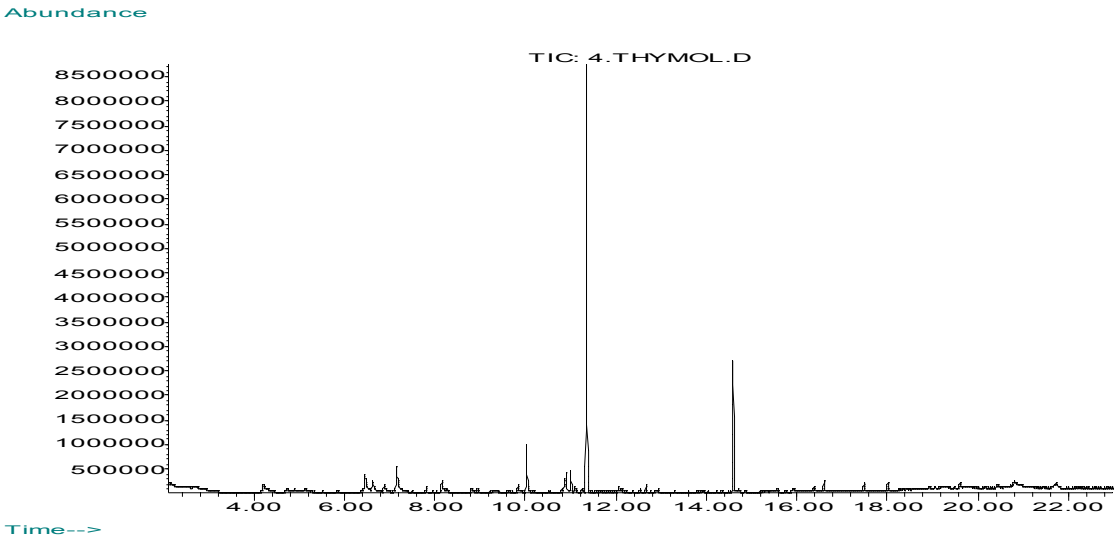

**2-Methylbenzofuran**



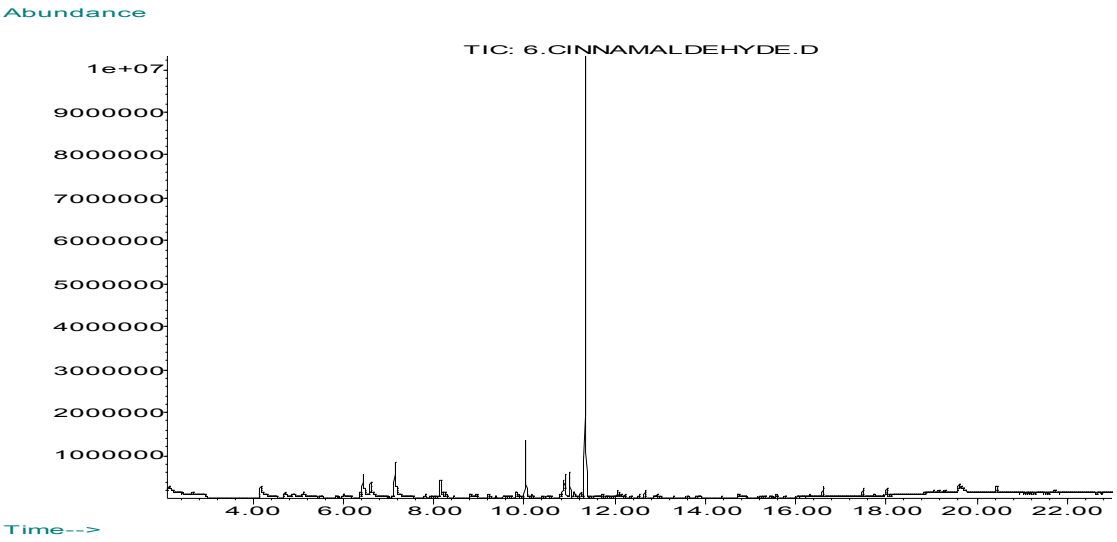

**Myrtenal**



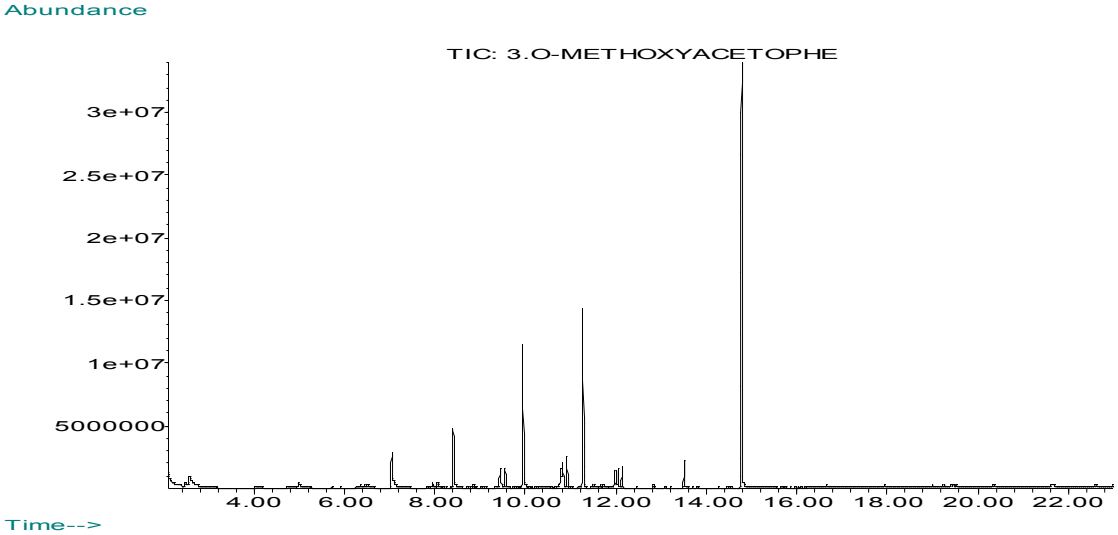

**1-(2-Methoxyphenyl)ethanol**



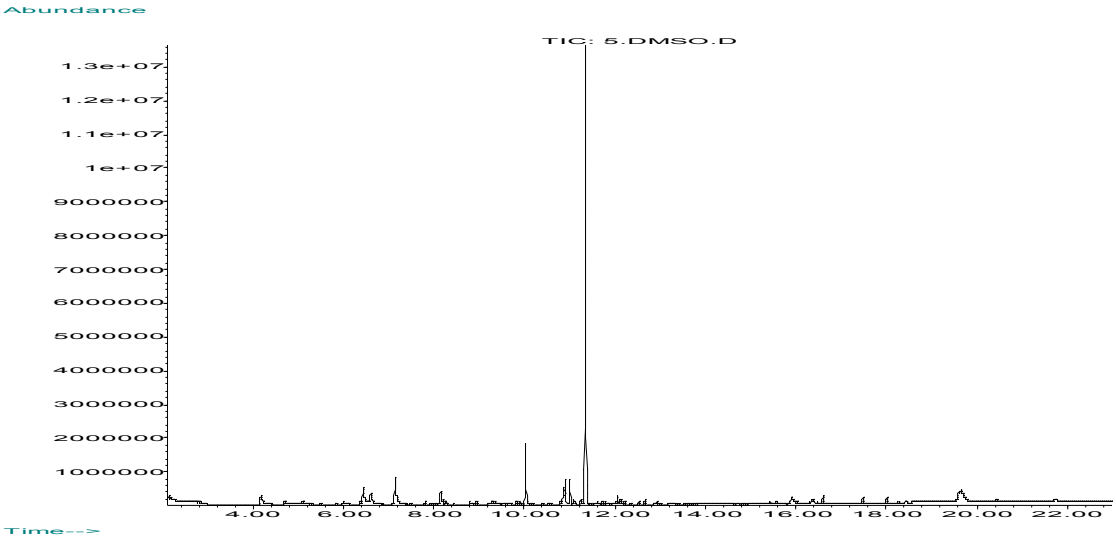

**Thymol**



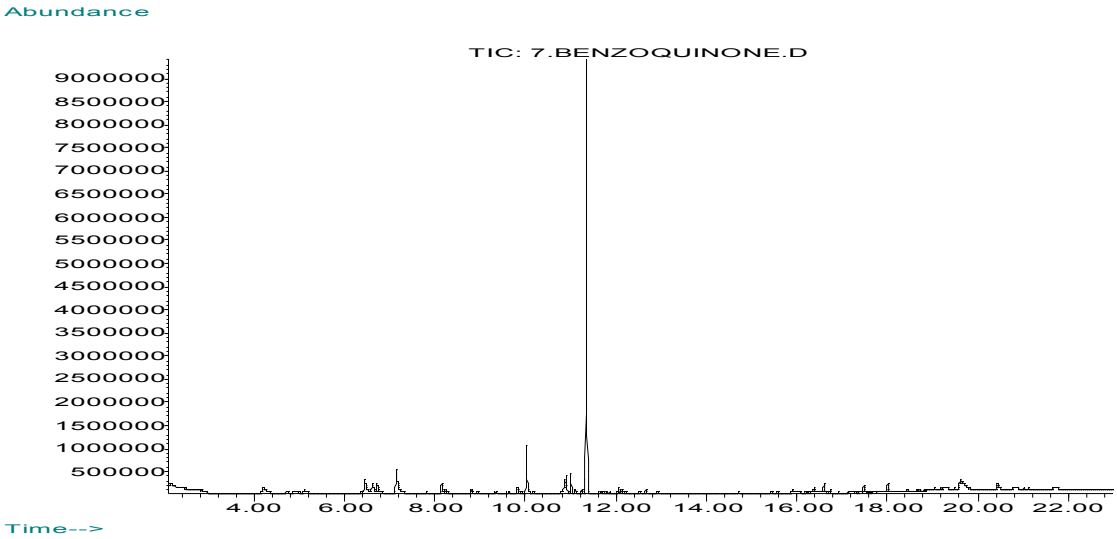

**(*E*)-Cinnamaldehyde**




***ortho*-Methoxyacetophenone**

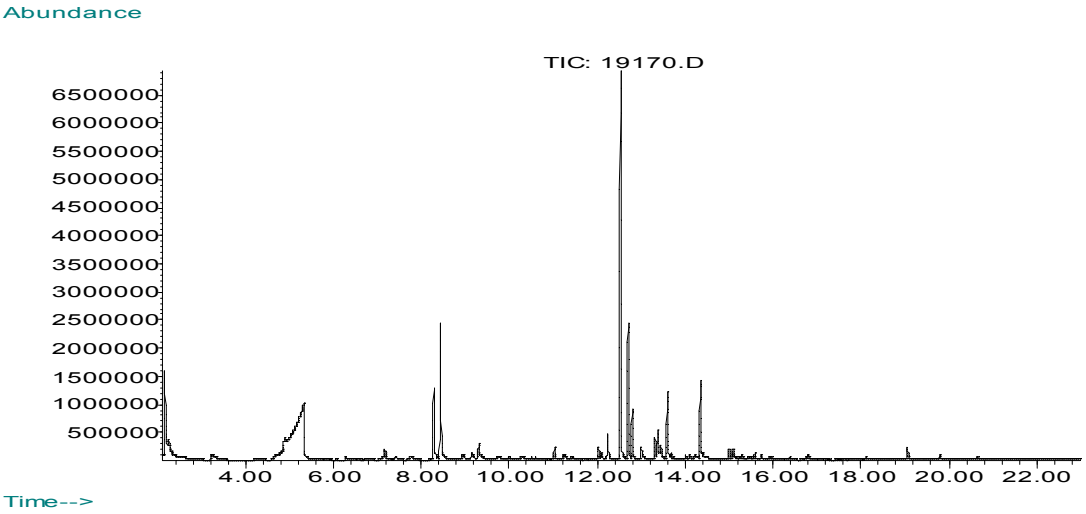



**Dimethyl sulfoxide**

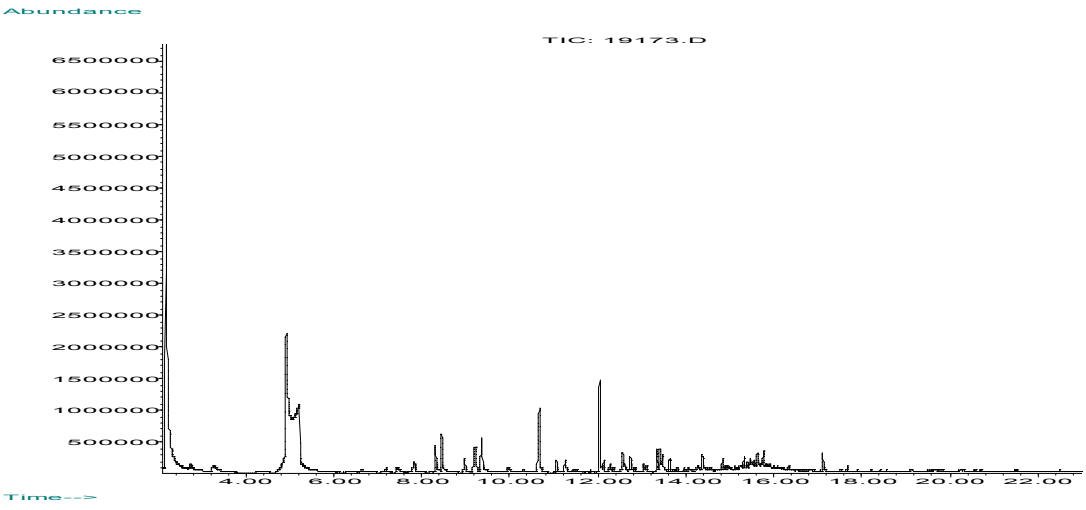



***p*-Benzoquinone**

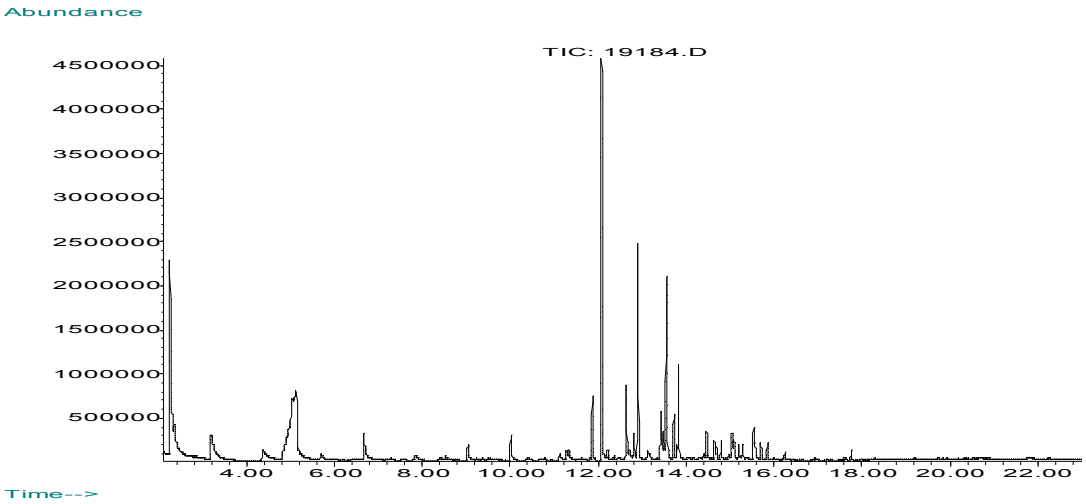
# Appendix 2: Chromatograms of Clover Honey Samples



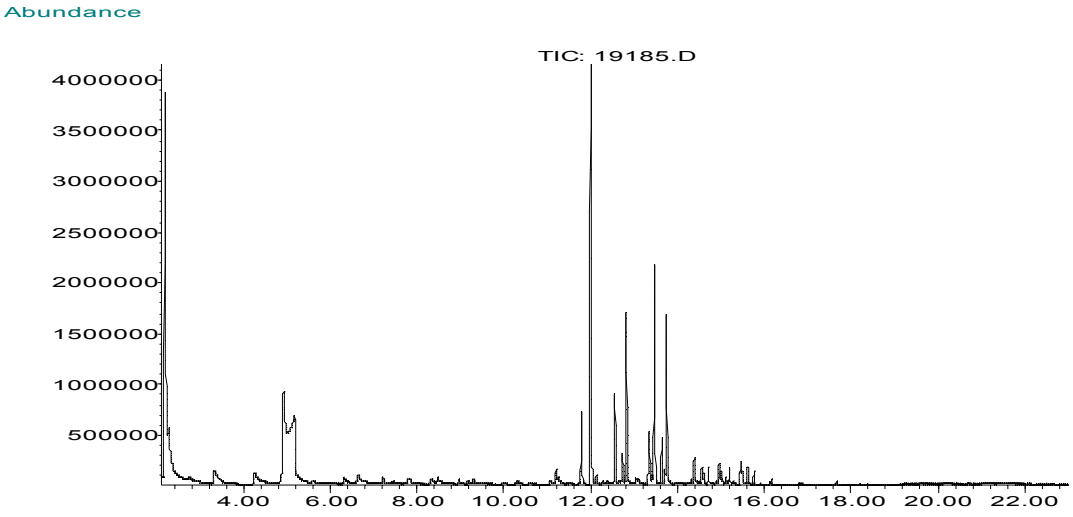

**19170**



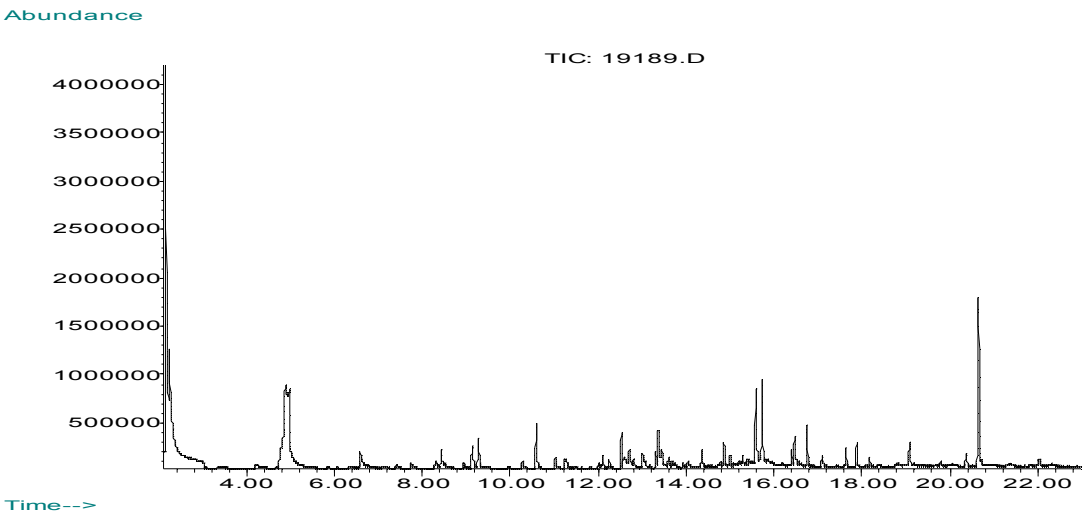

**19173**



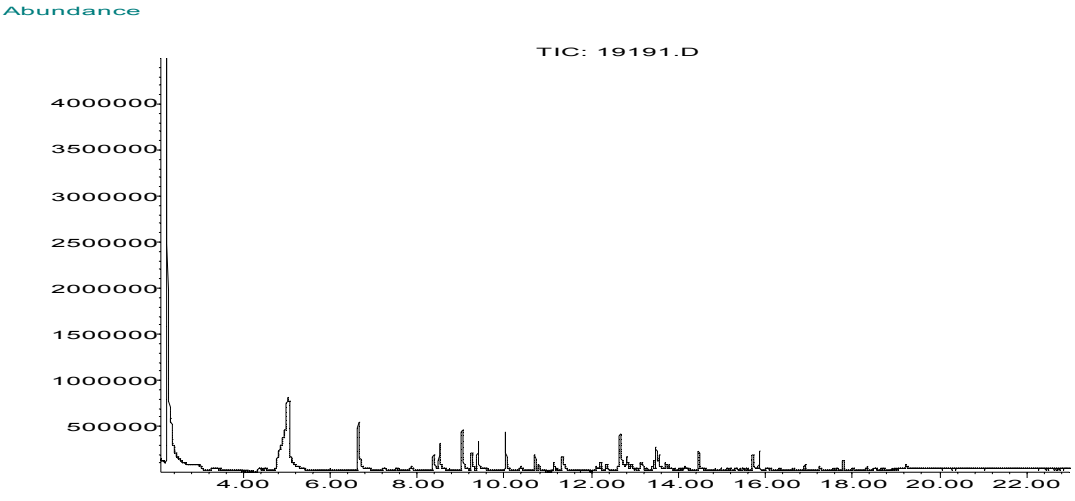

**19184**



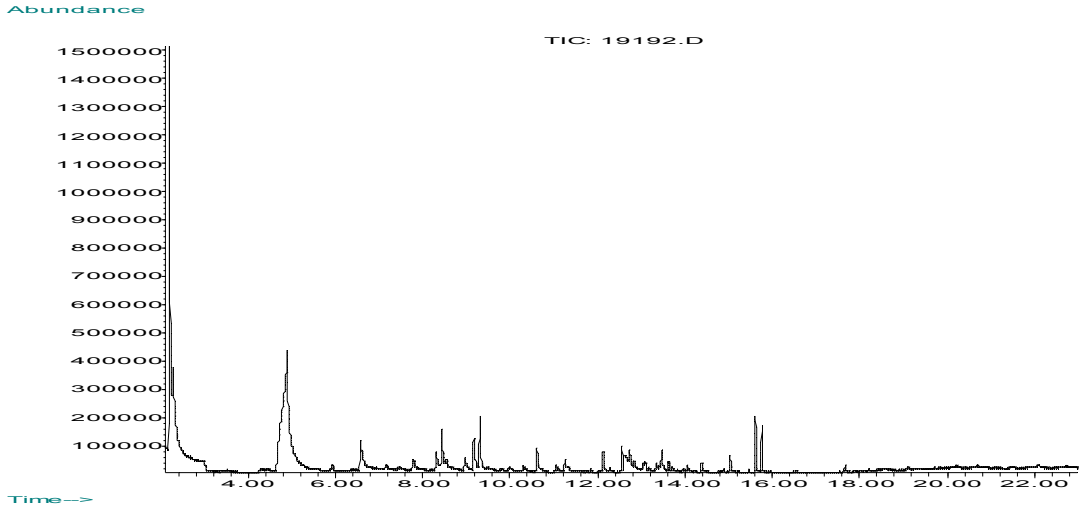

**19185**



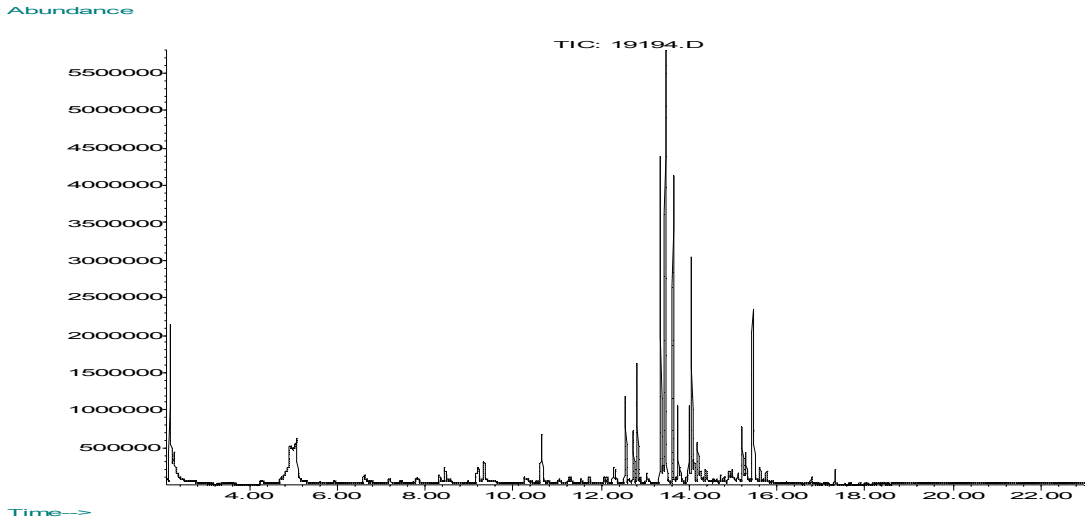

**19189**



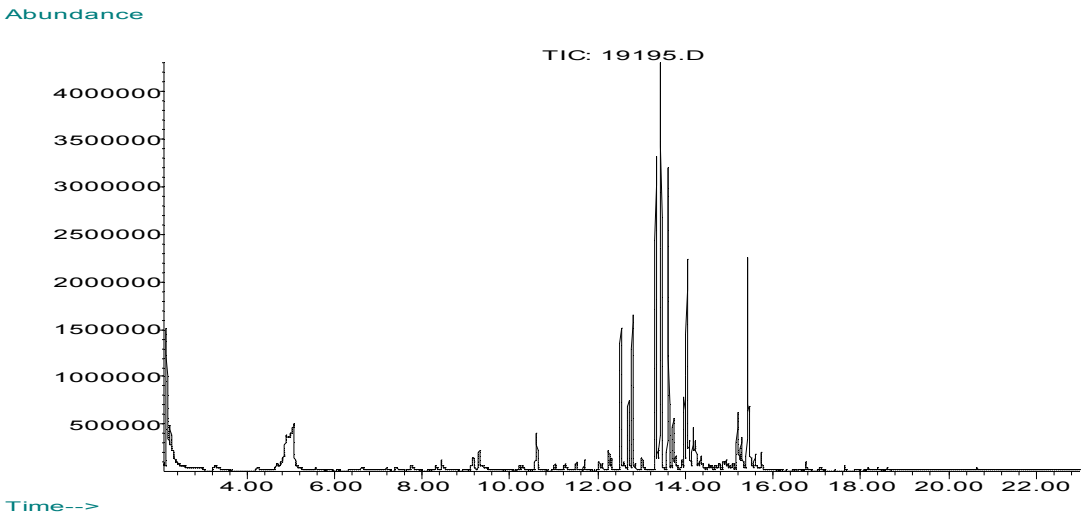

**19191**



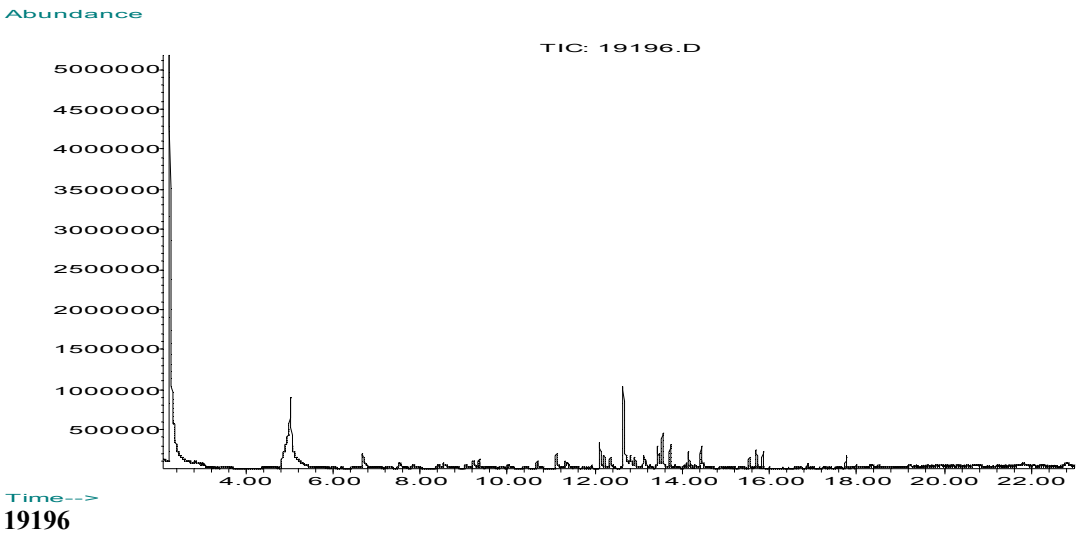

**19192**



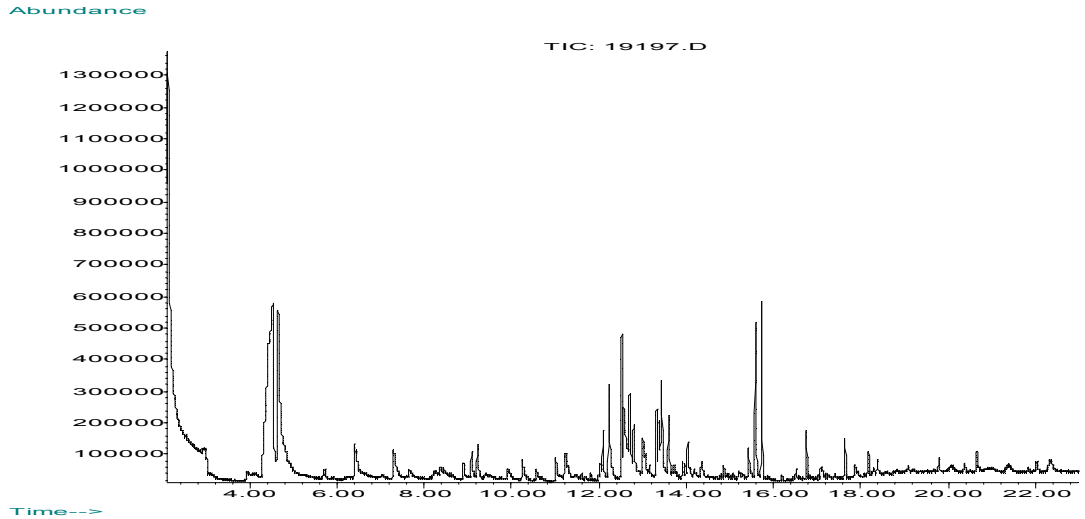

**19194**



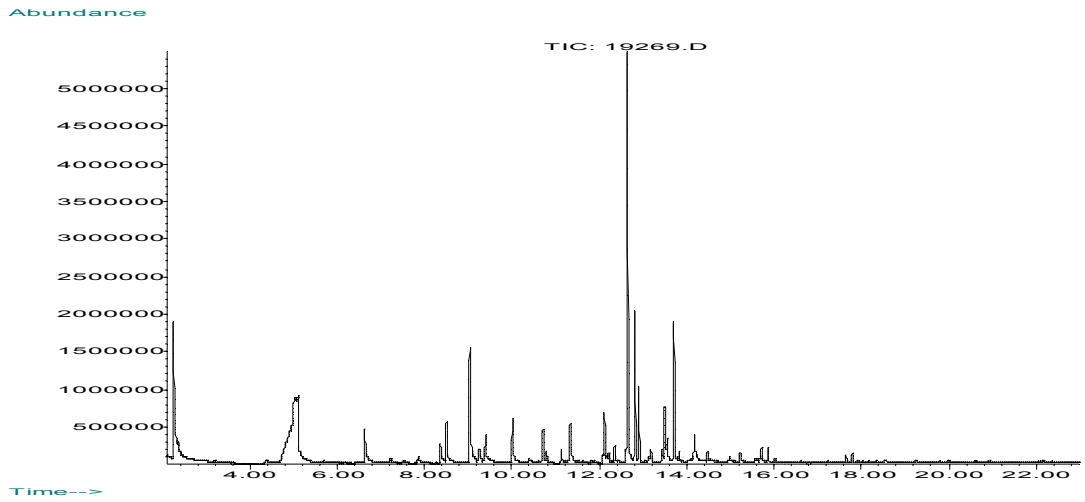

**19195**



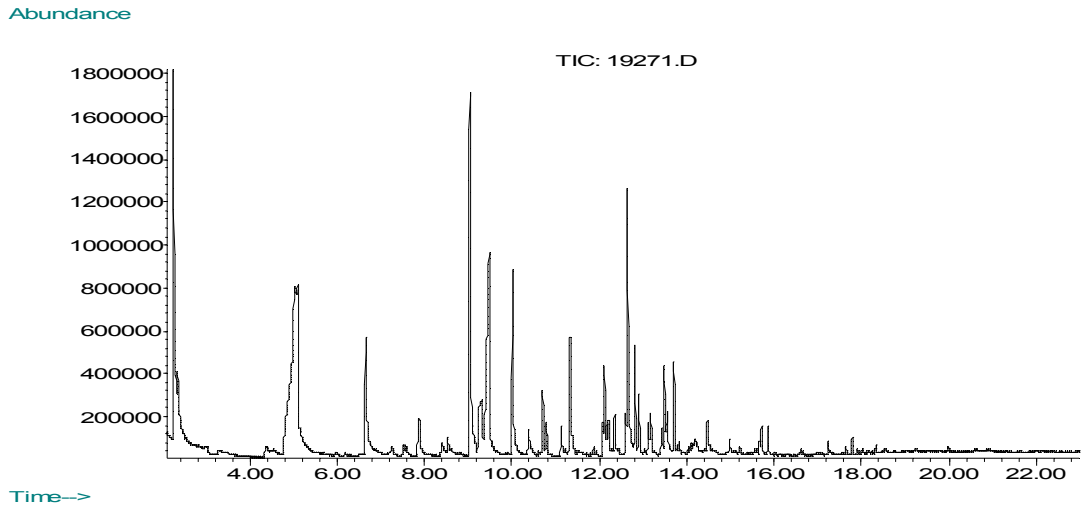

**19196**



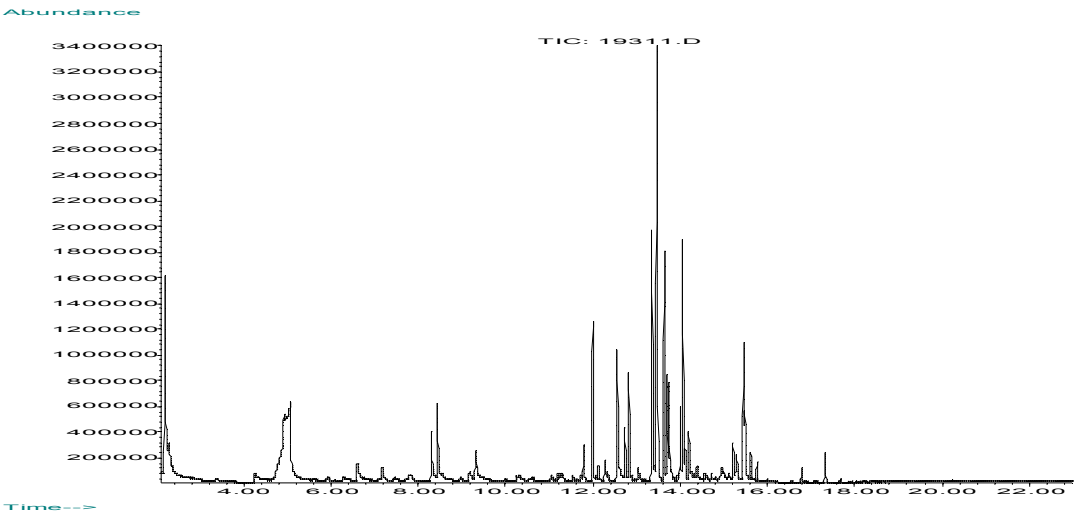

**19197**


Abundance
TIC: 19269.D
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

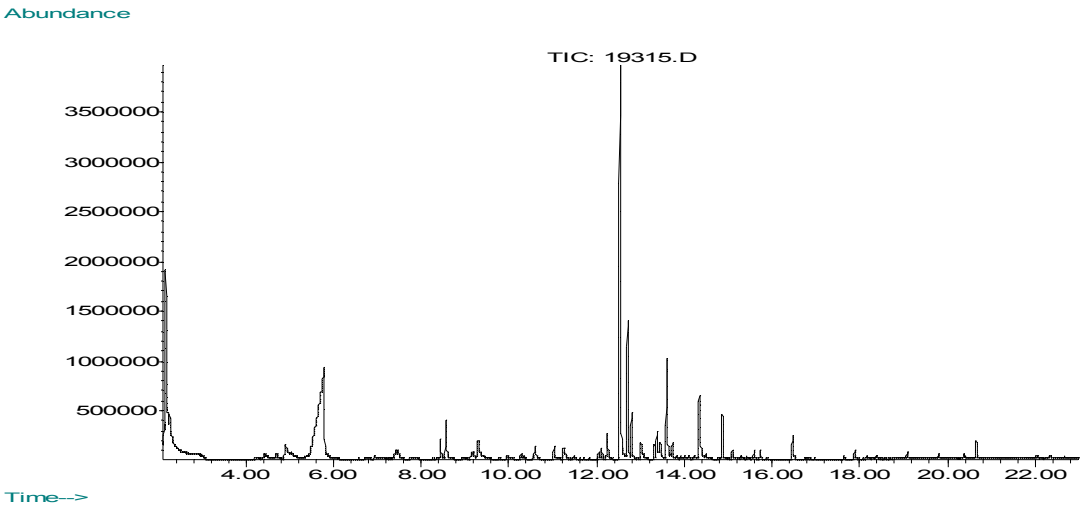
**19269**


Abundance
TIC: 19271.D
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

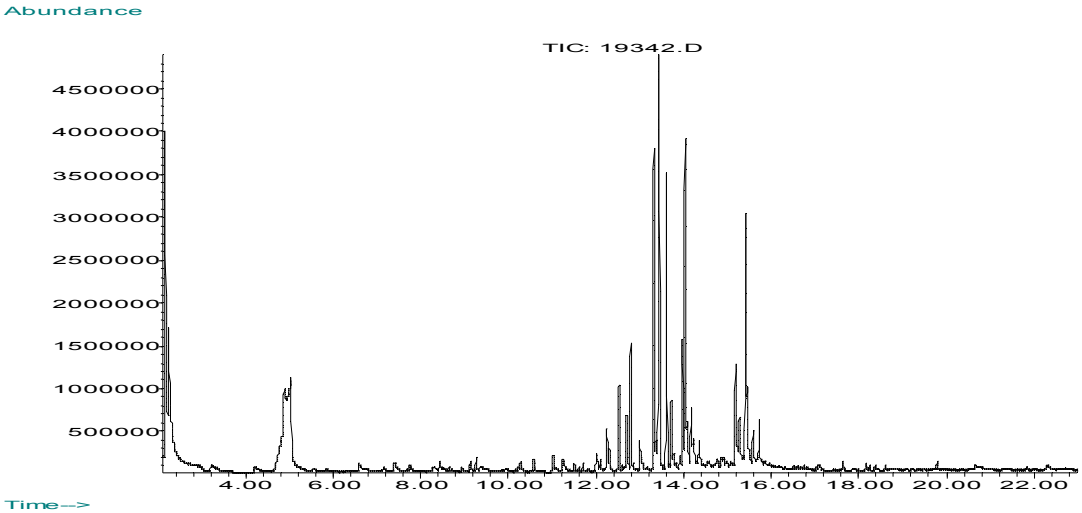
**19271**



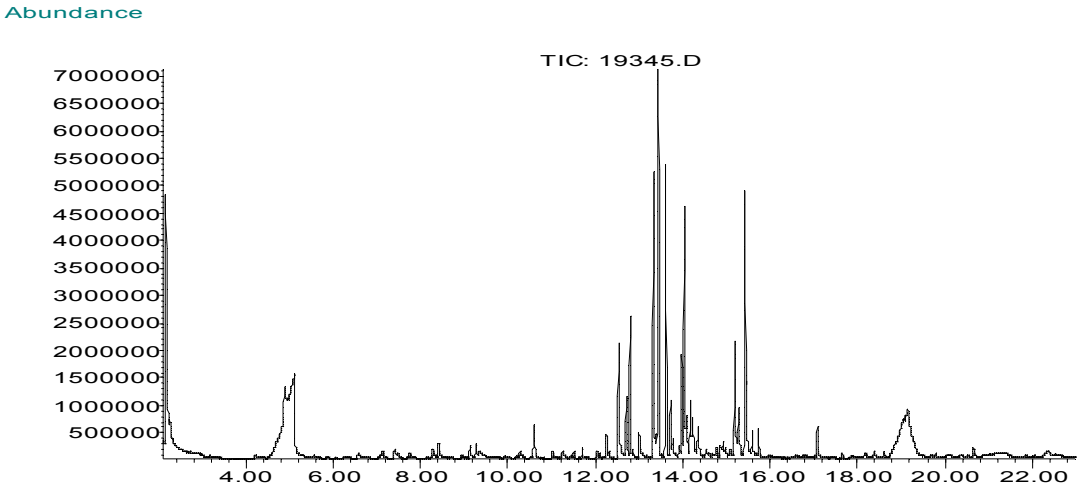

**19311**



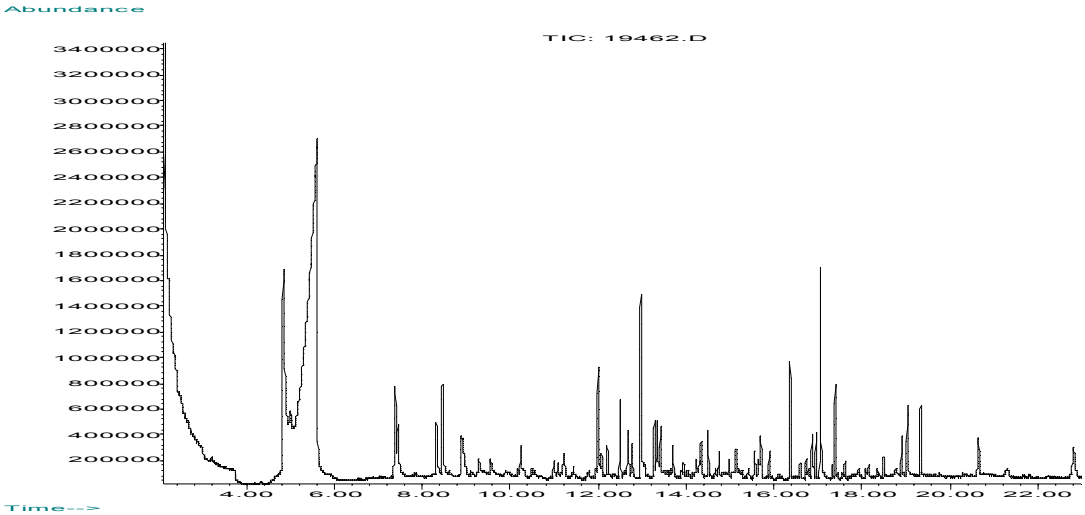

**19315**



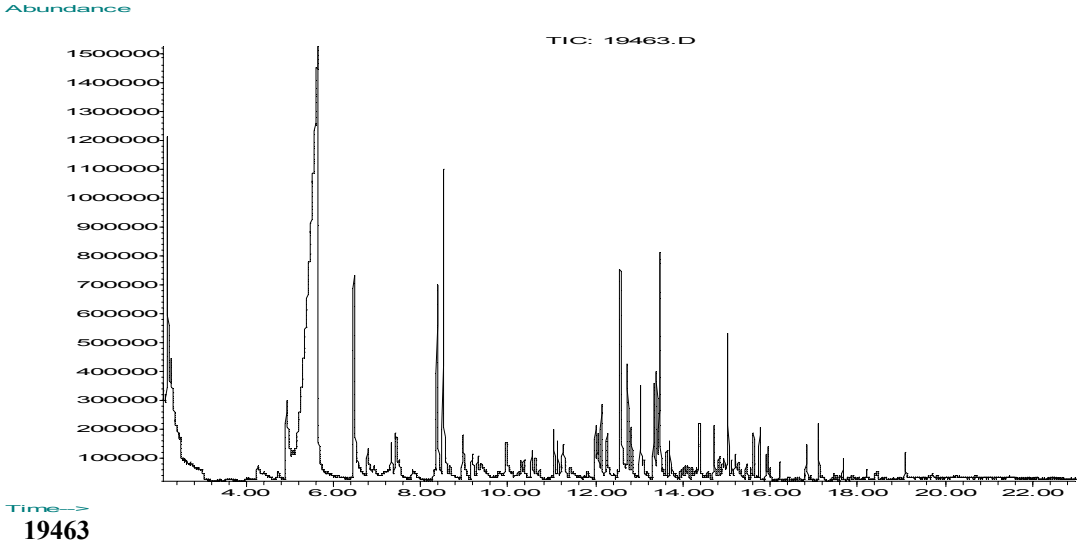

**19342**



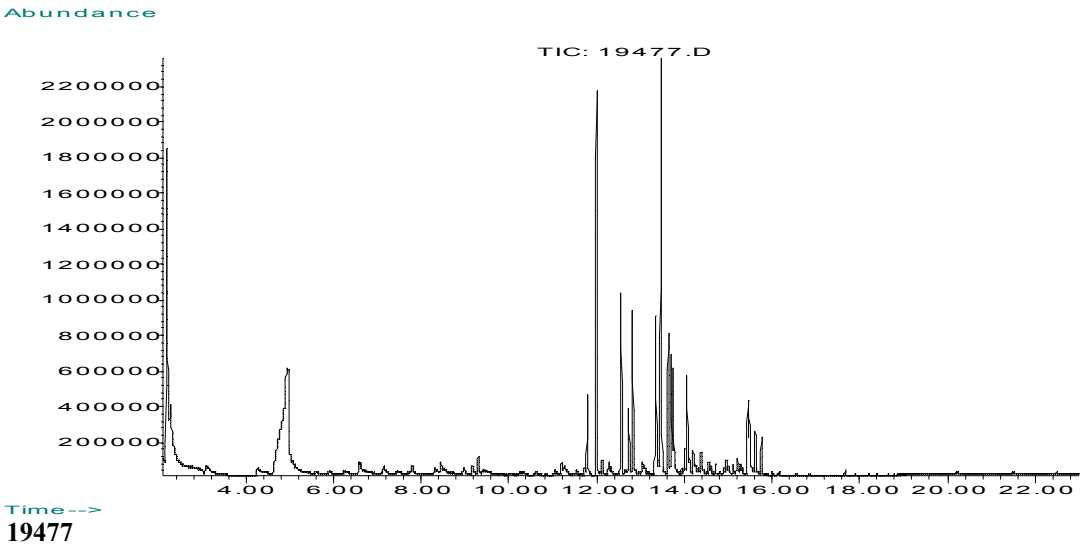

**19345**

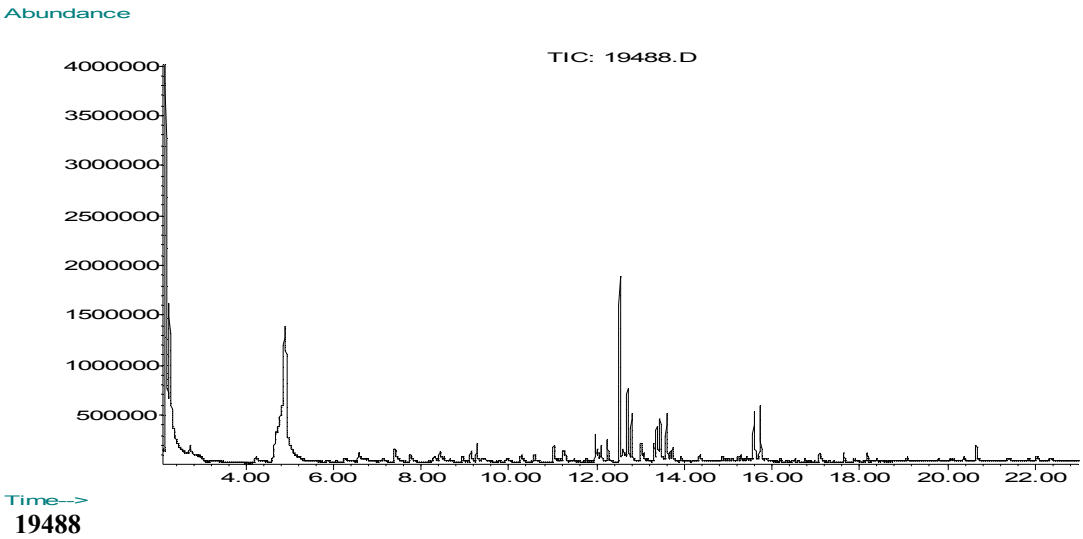


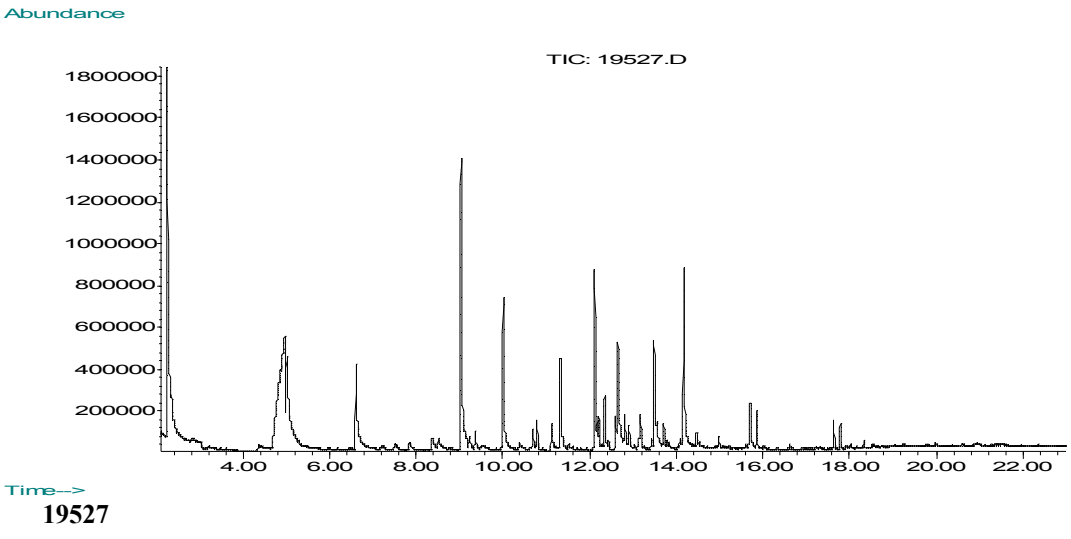

**19462**

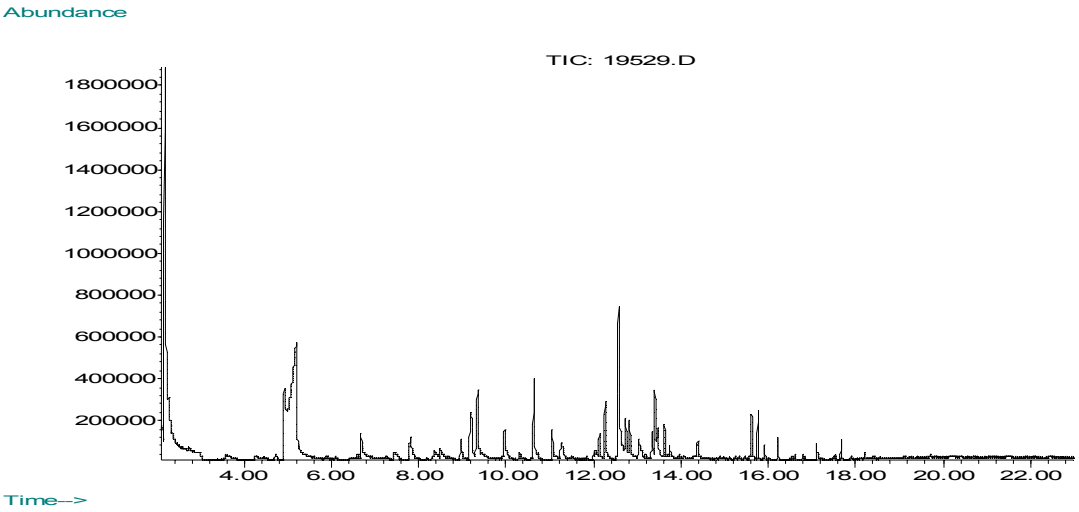



**19463**

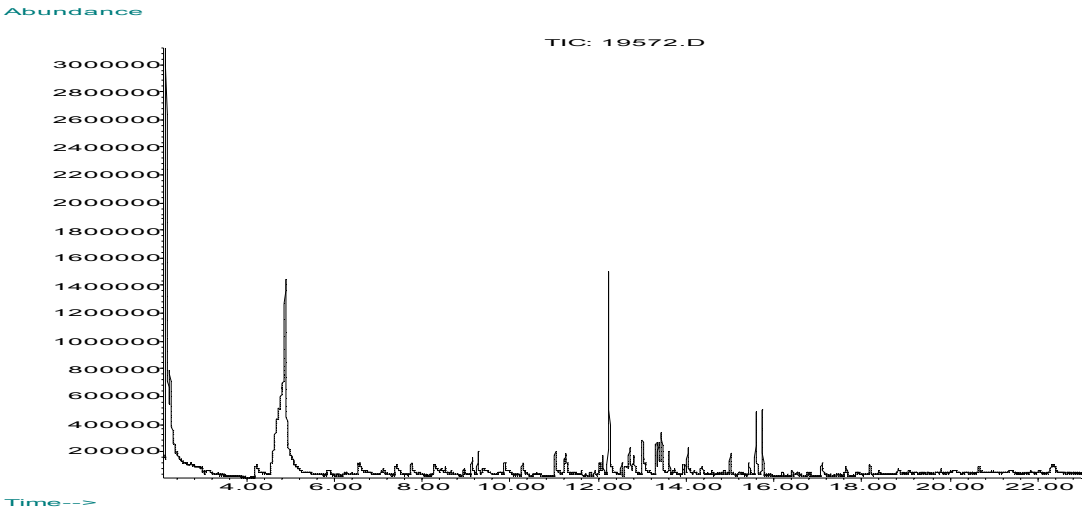

Abundance
TIC: 19477.D
2200000
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->
19477


Abundance
TIC: 19488.D
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->
19488

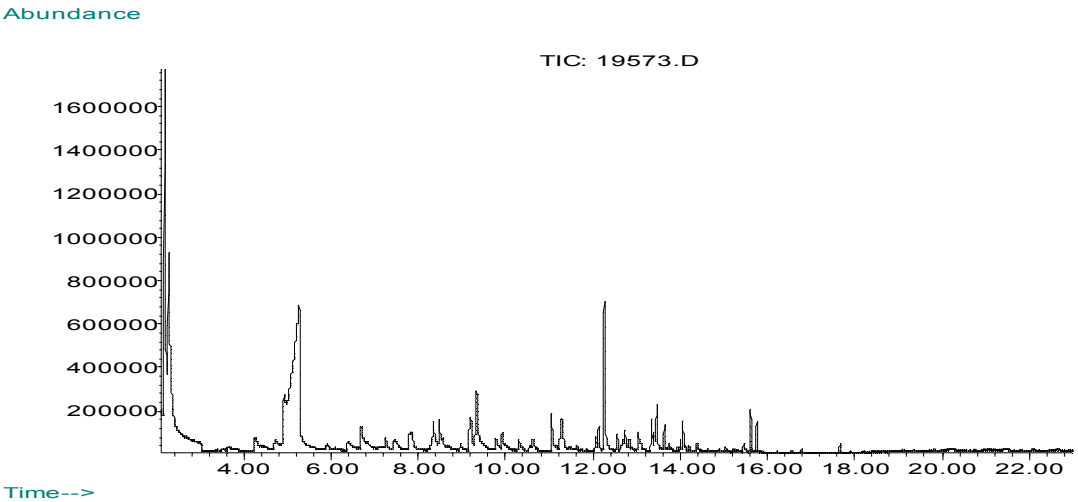

Abundance
TIC: 19527.D
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->
19527



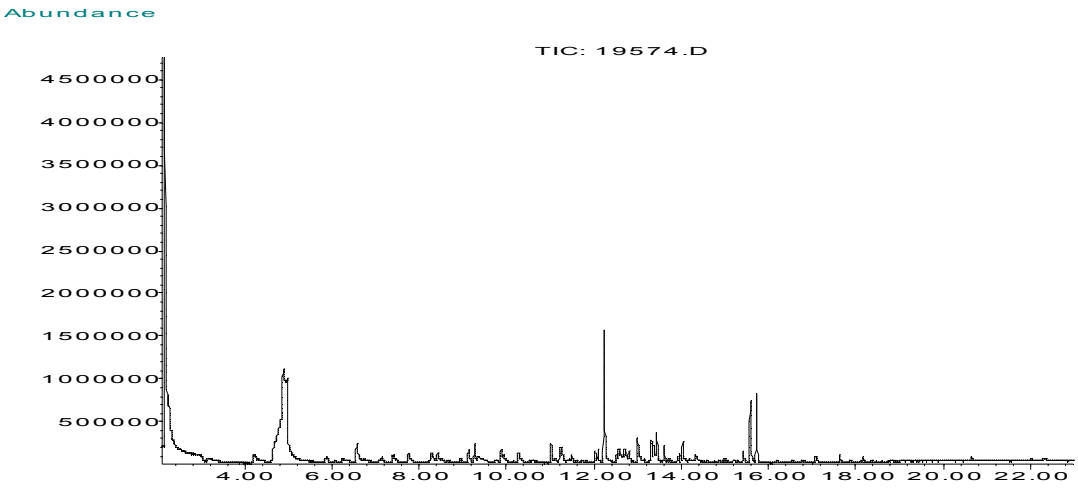

**19529**



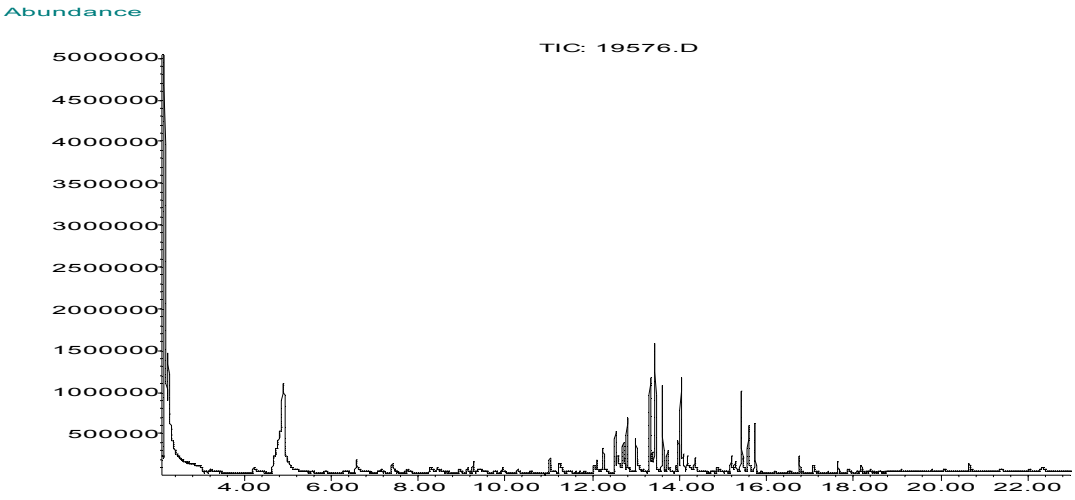

**19572**



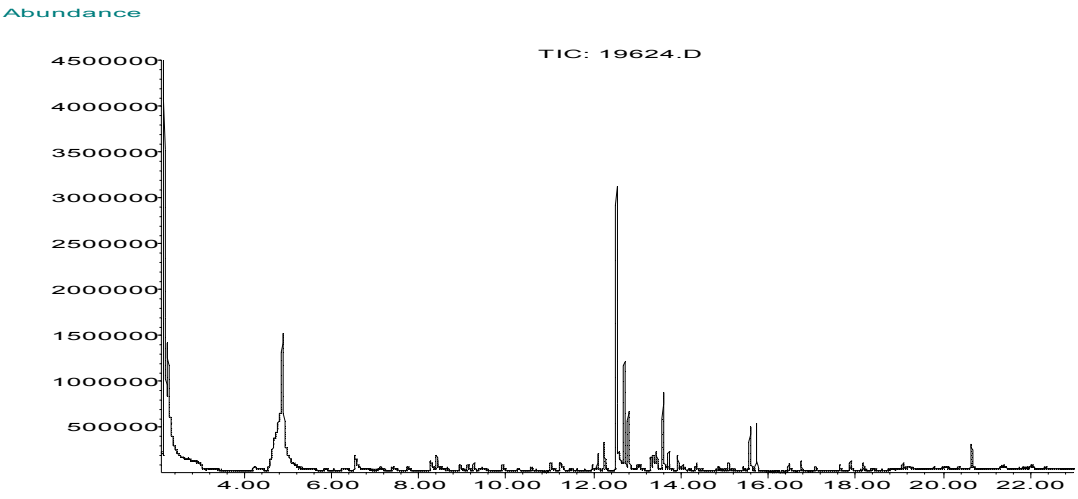

**19573**



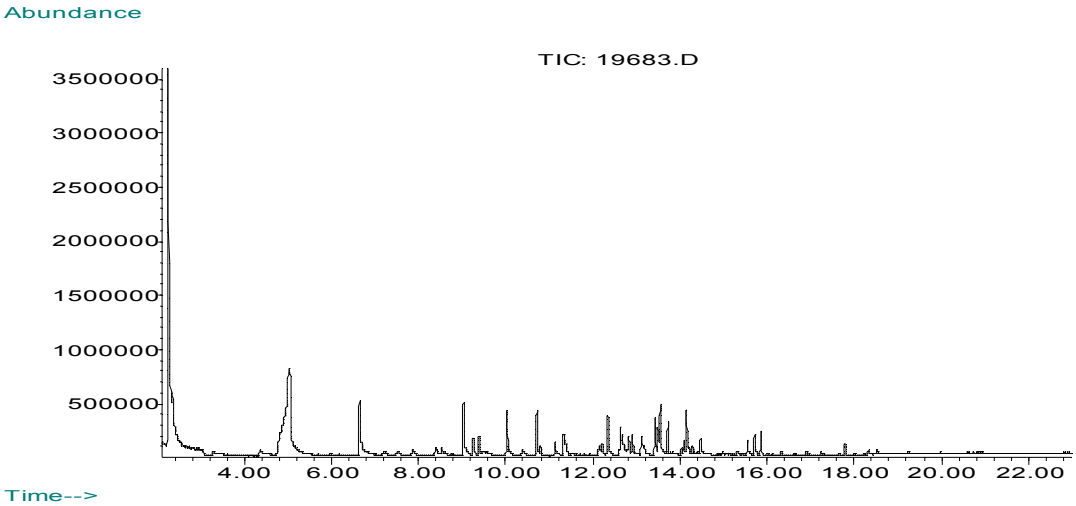

**19574**



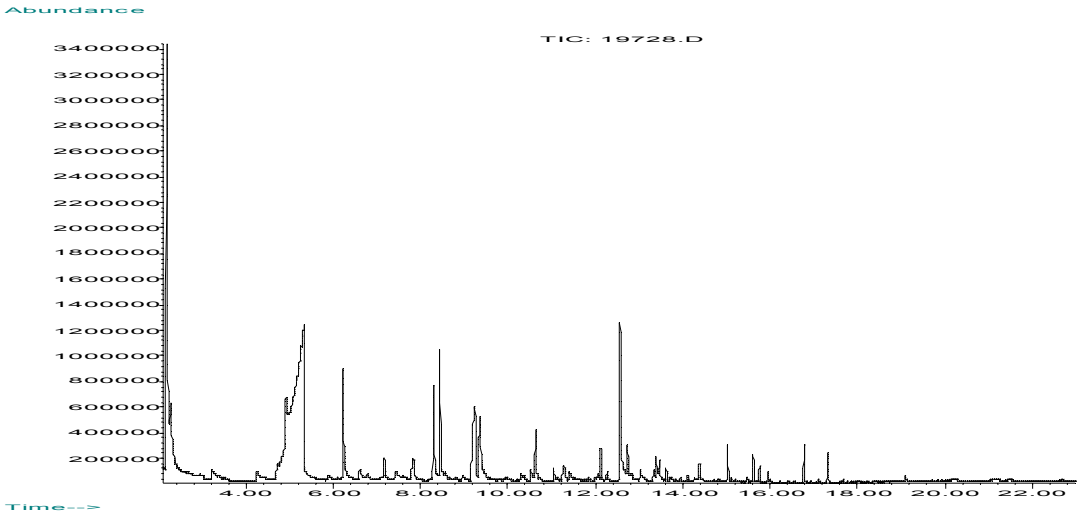

**19576**



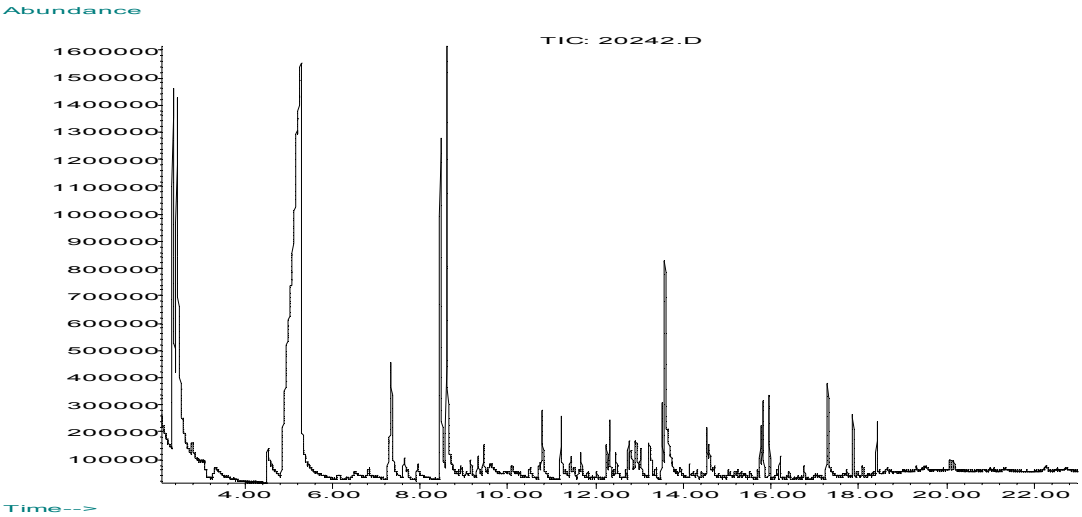

**19624**



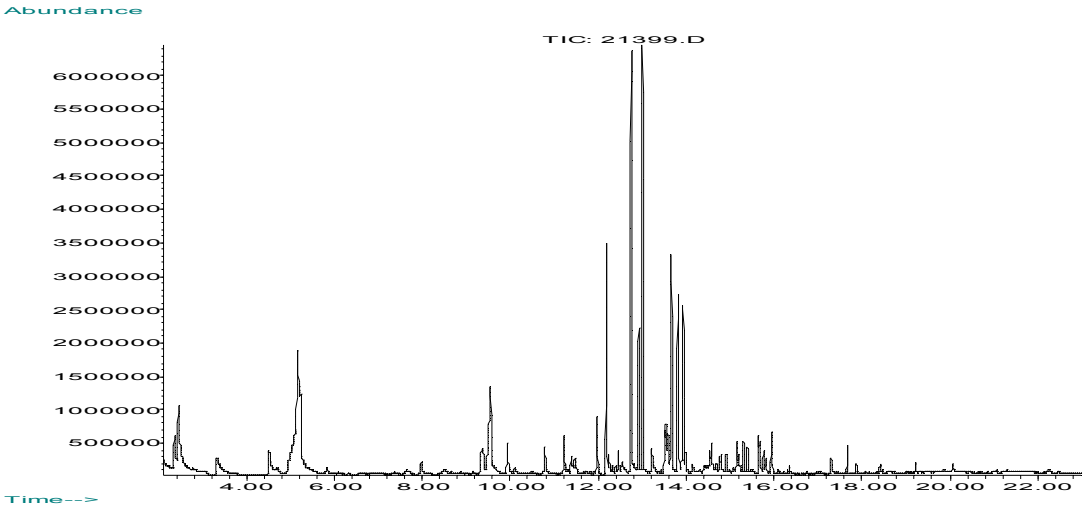

**19683**



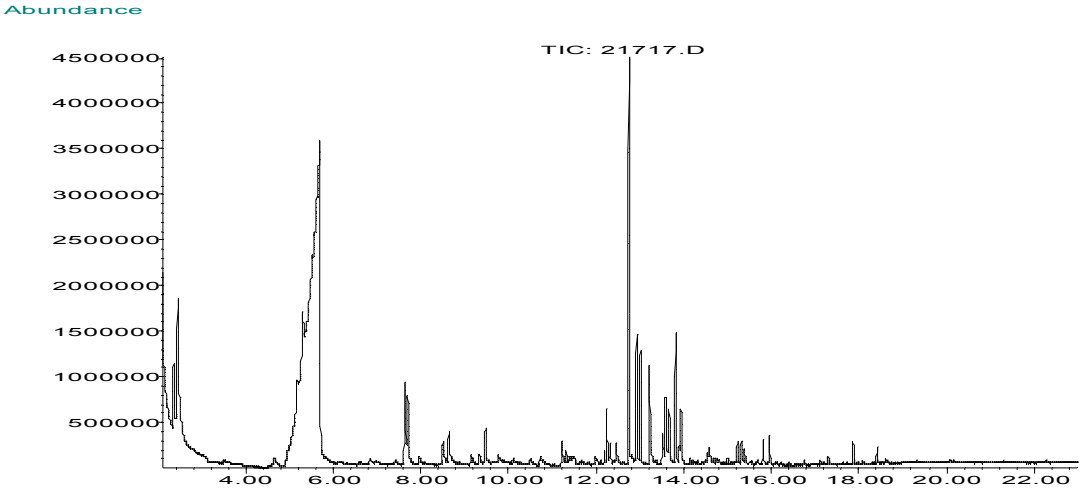

**19728**



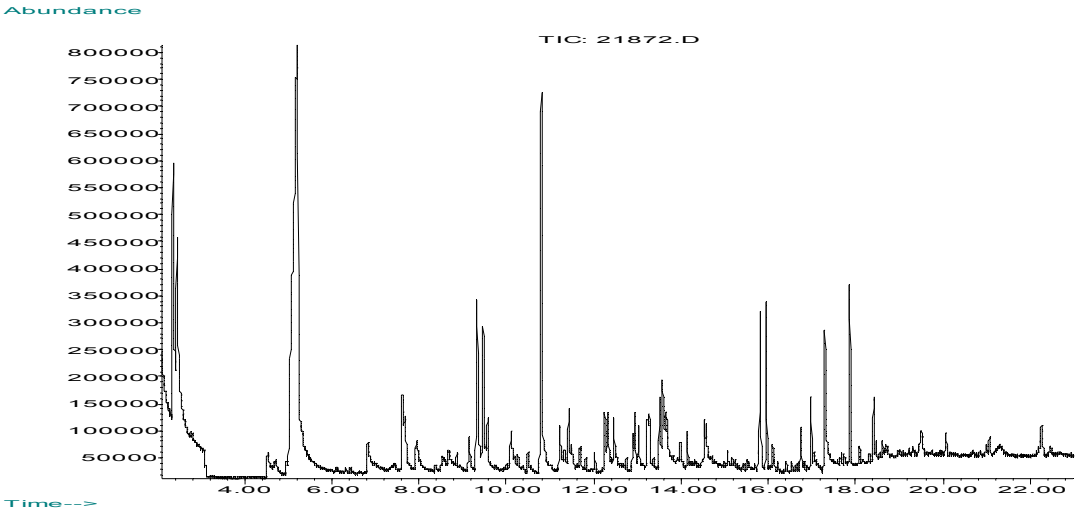

**20242**



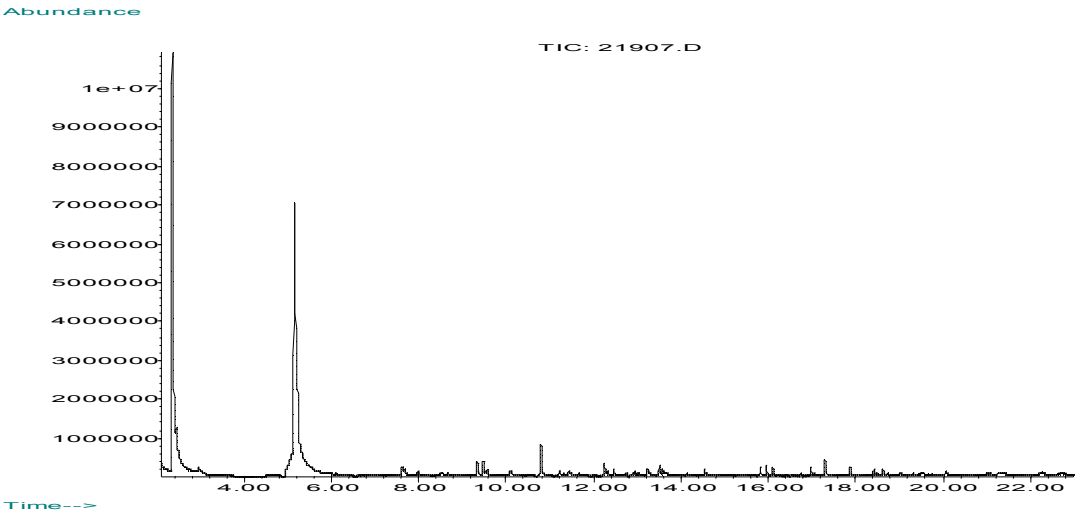

**21399**



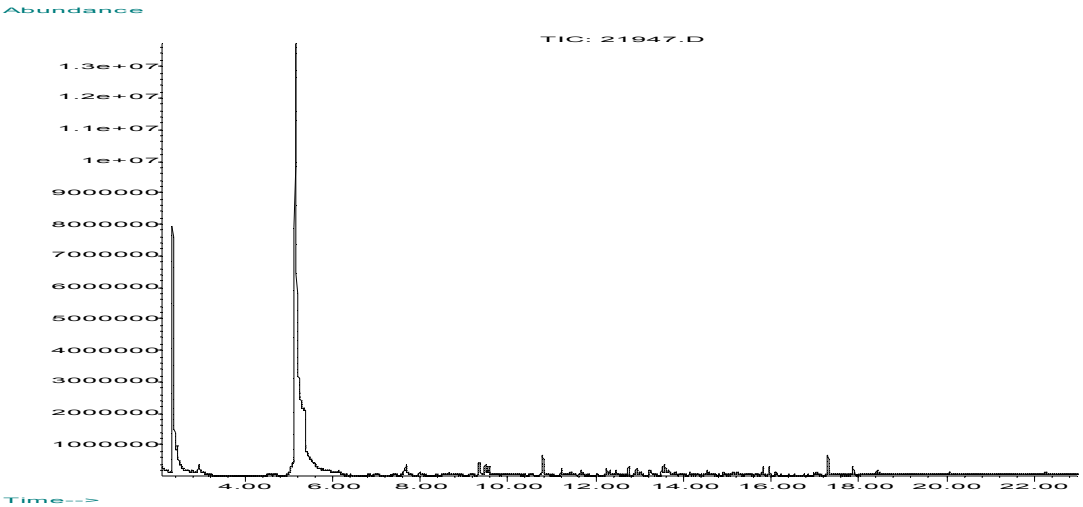

**21717**



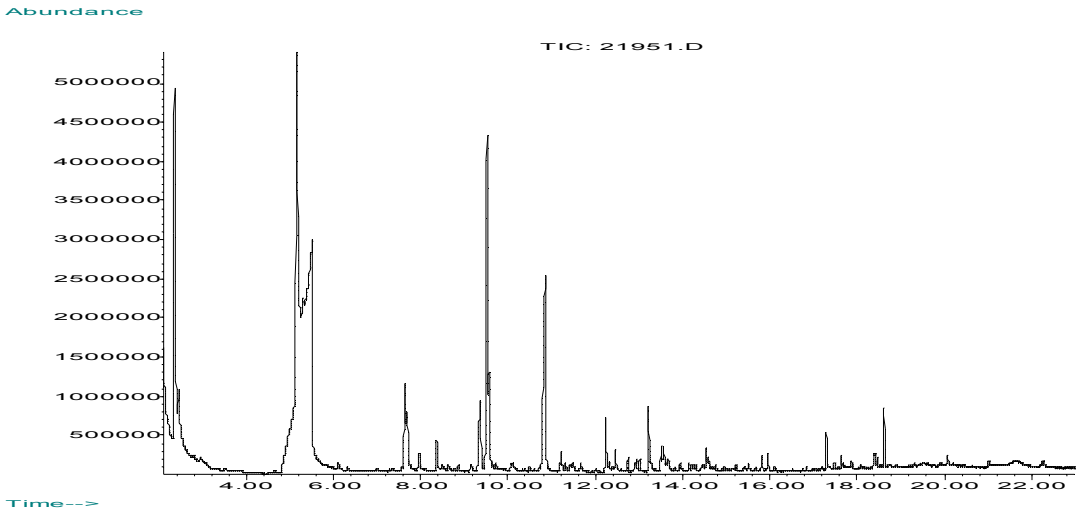

**21872**



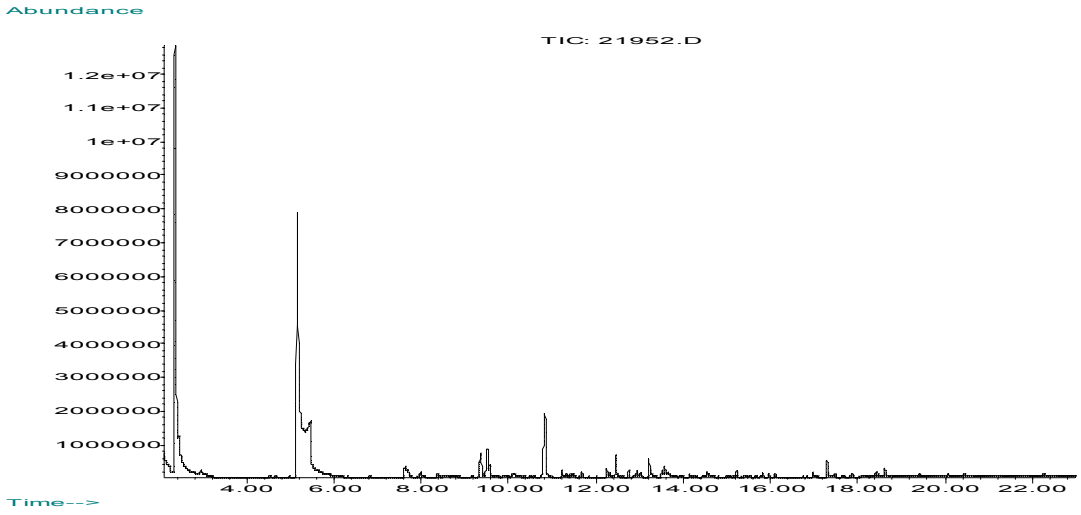

**21907**



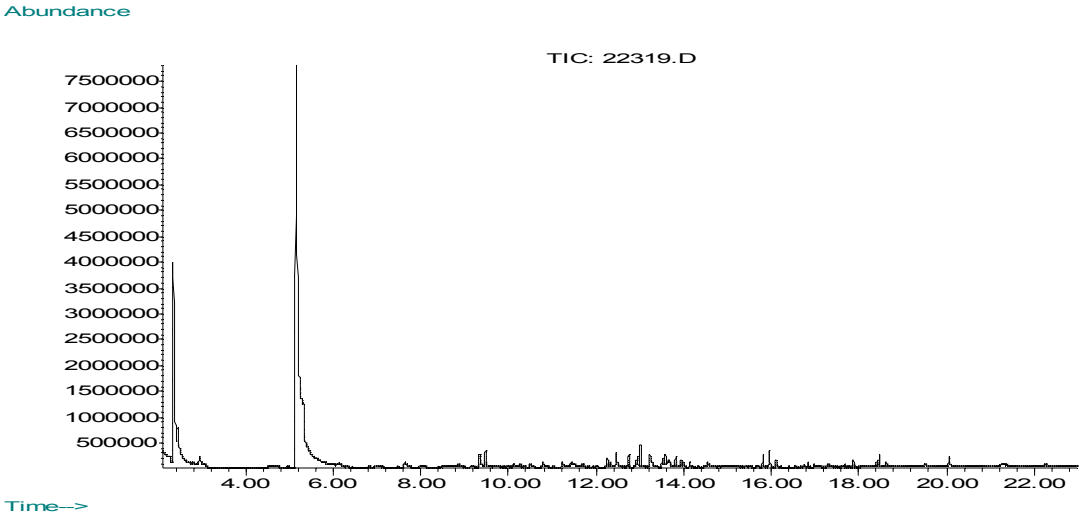

**21947**



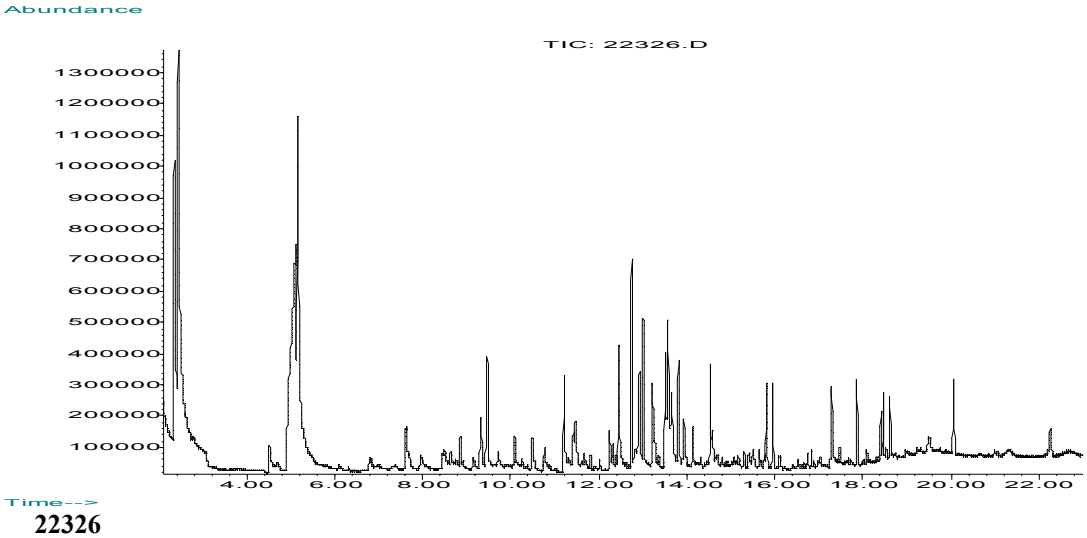

**21951**



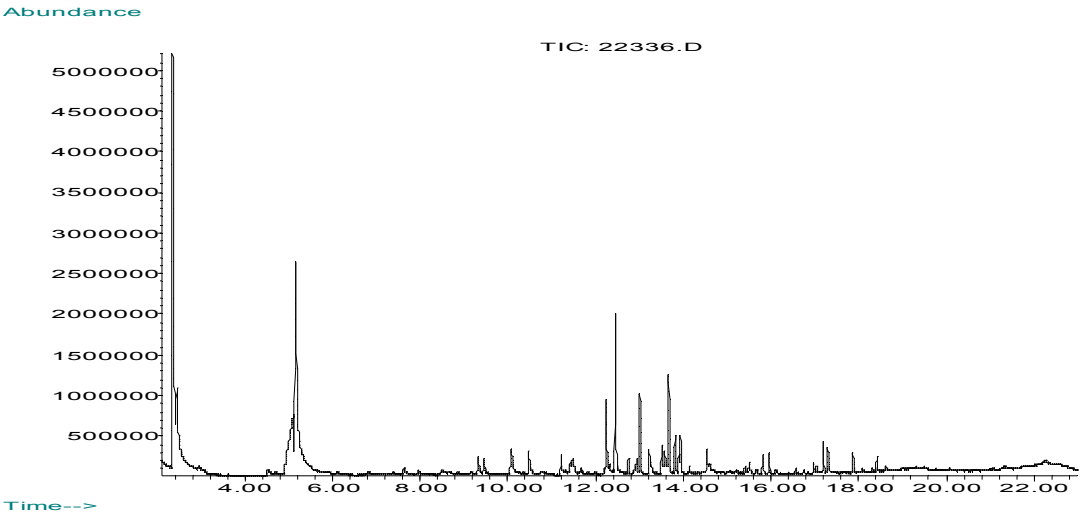

**21952**



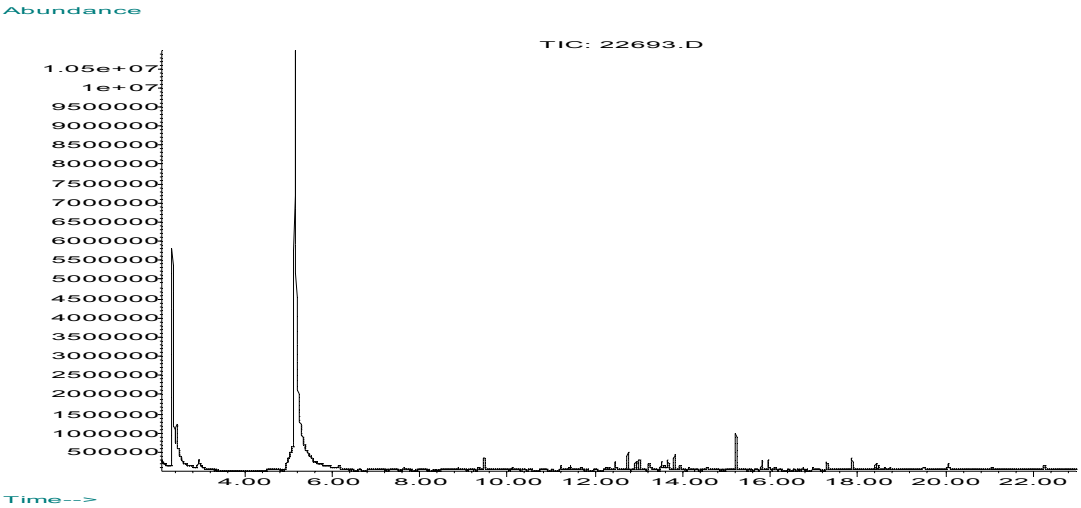

**22319**



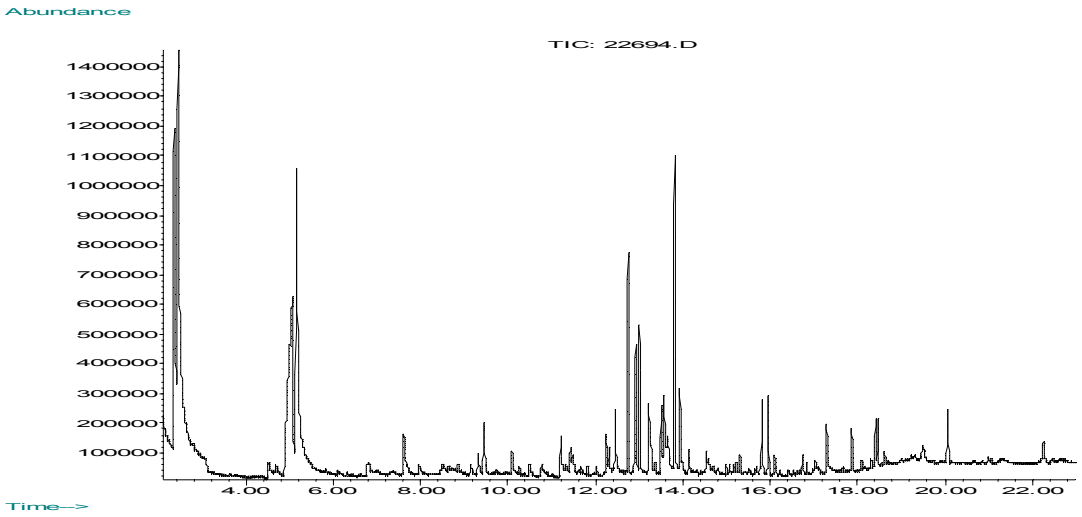

**22326**



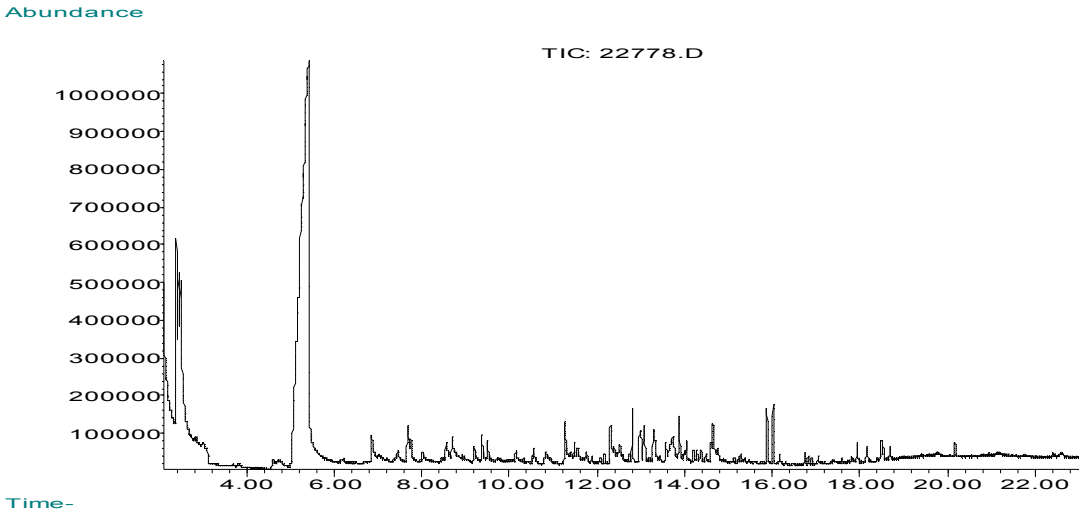

**22336**




**22693**

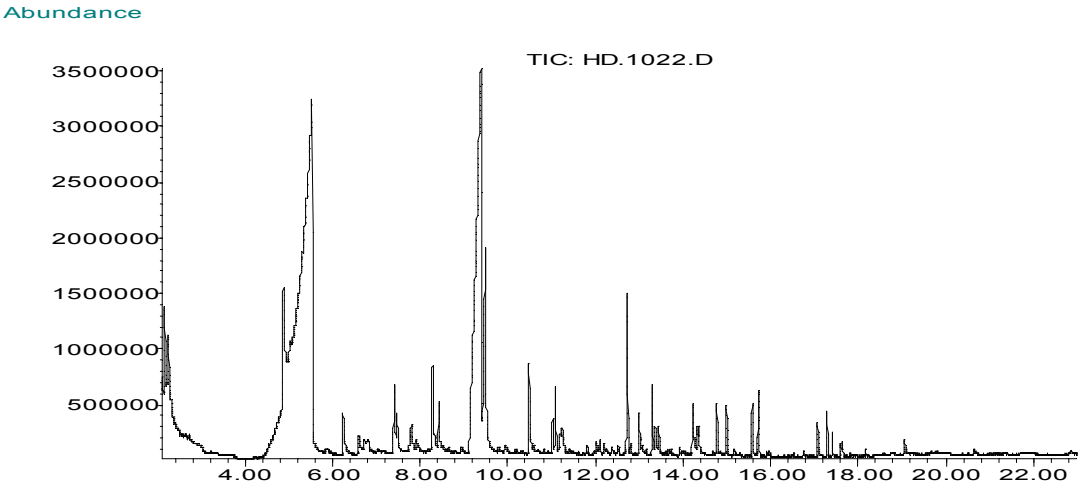



**22694**

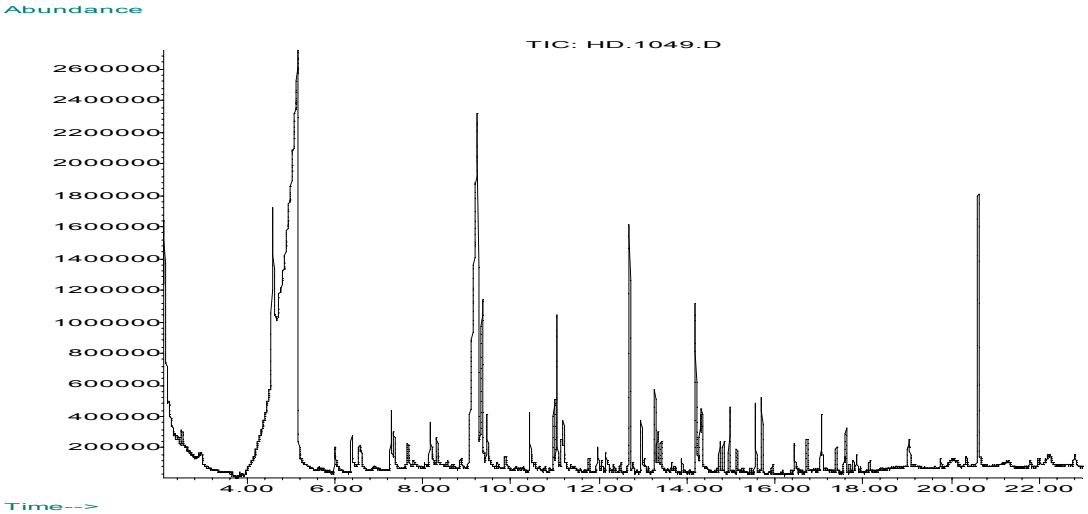

Abundance
TIC: 22778.D
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-

**22778**

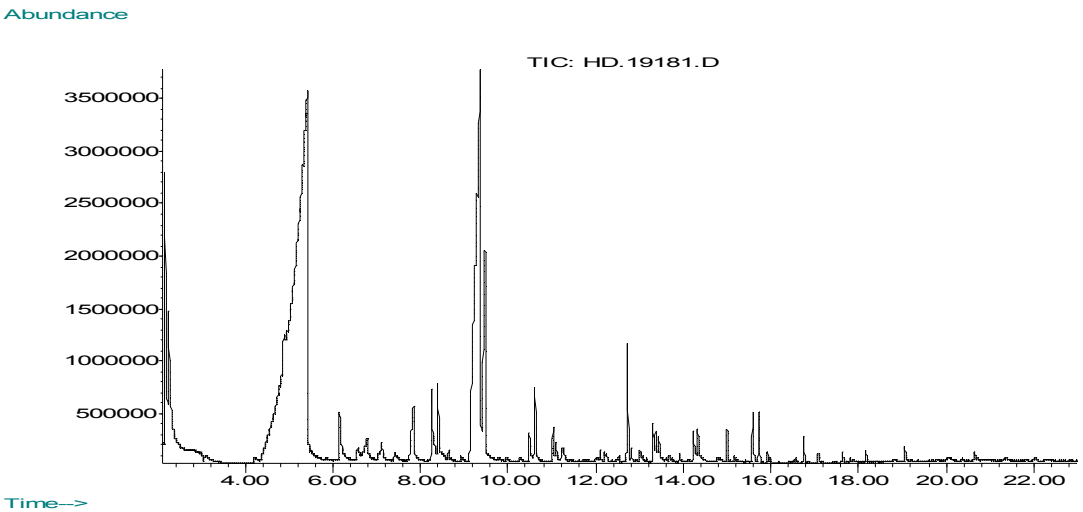
# Appendix 3: Chromatograms of Honeydew Honey Samples



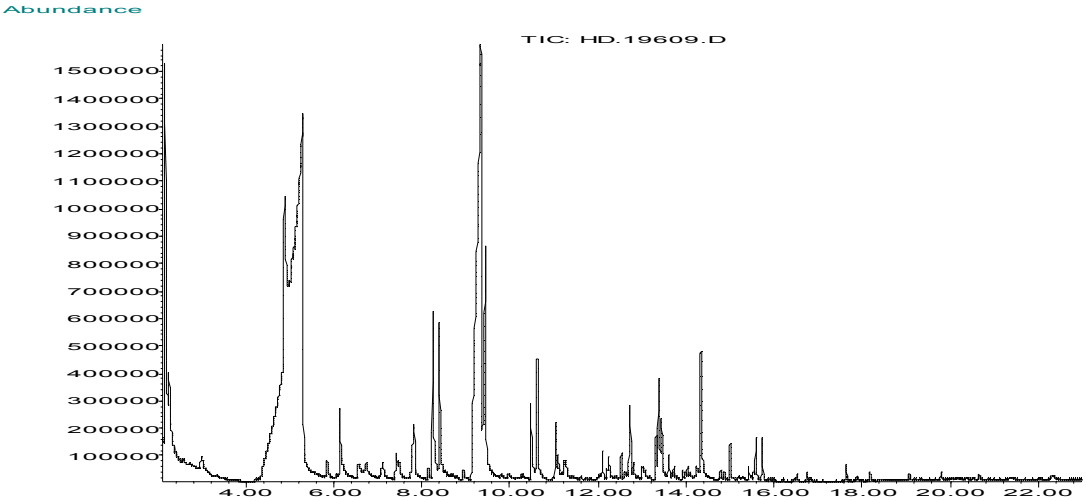

**1022**



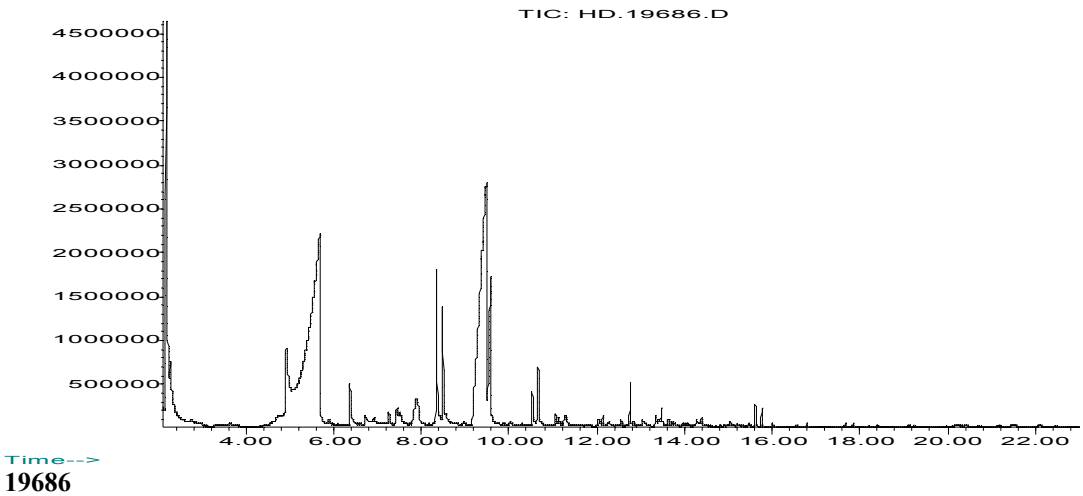

**1049**



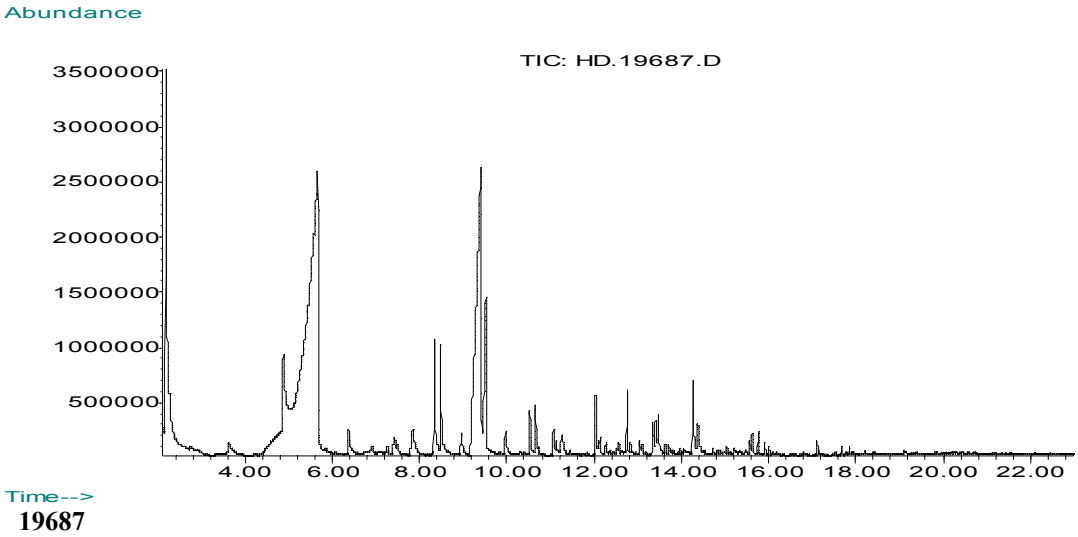

**19181**



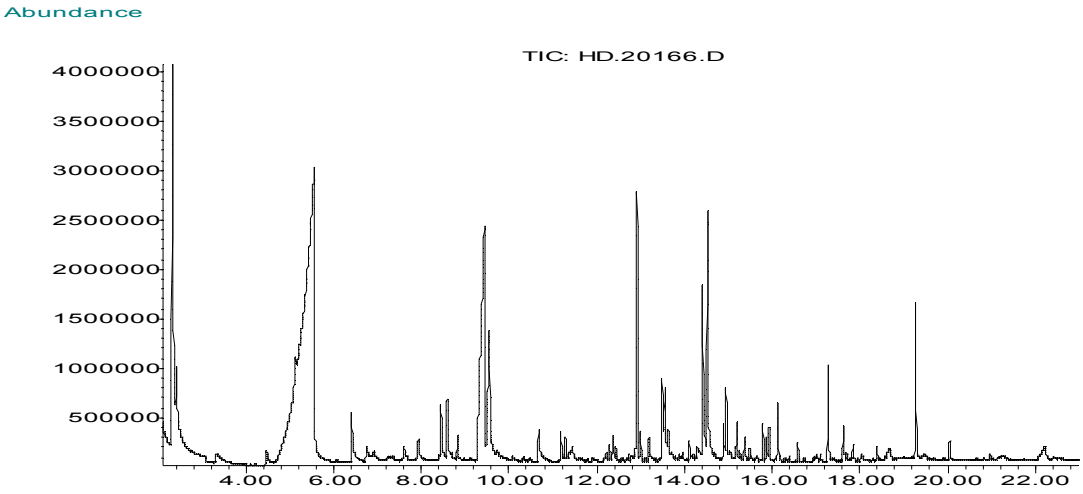

**19609**



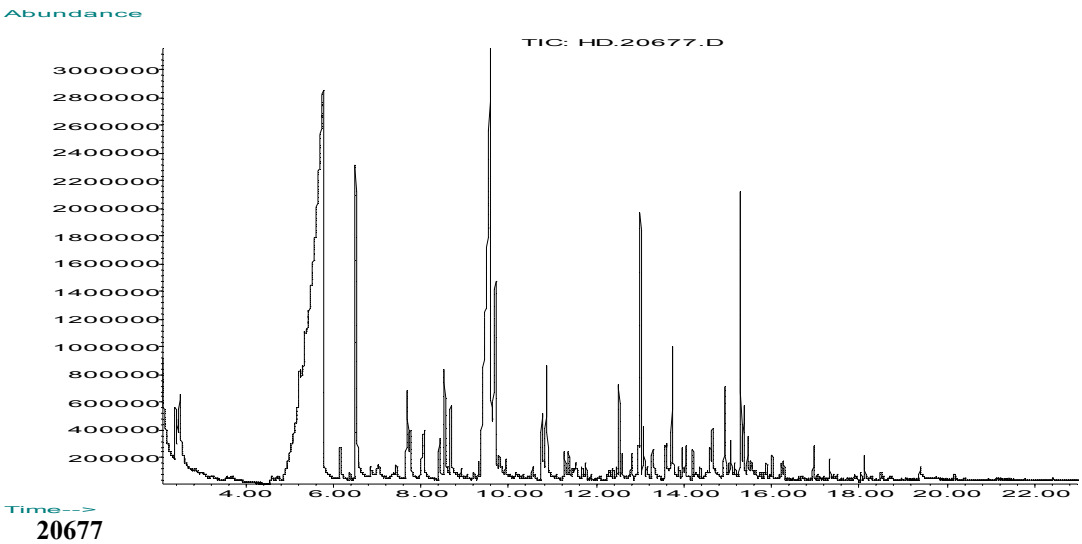

**19686**



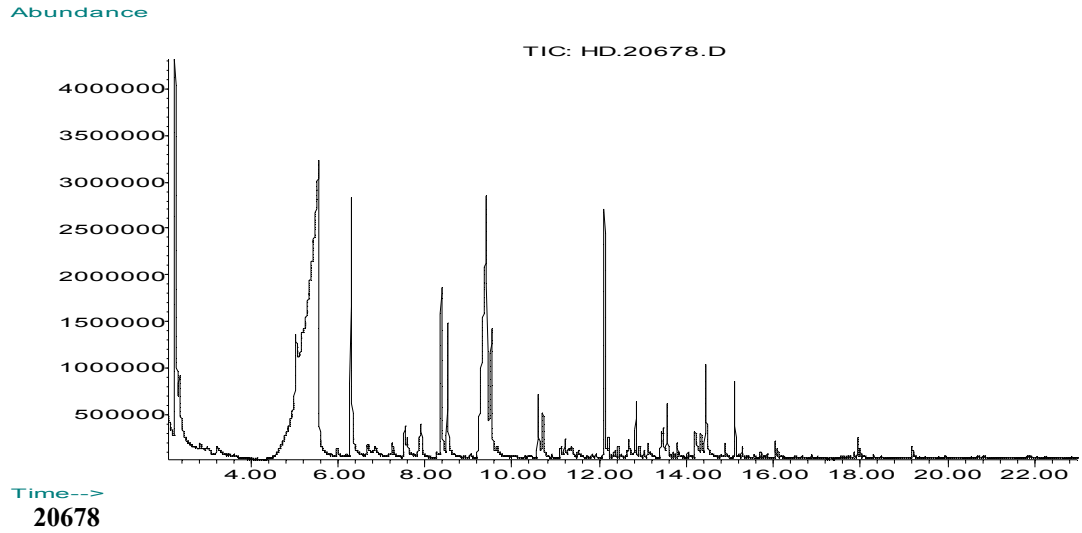

**19687**



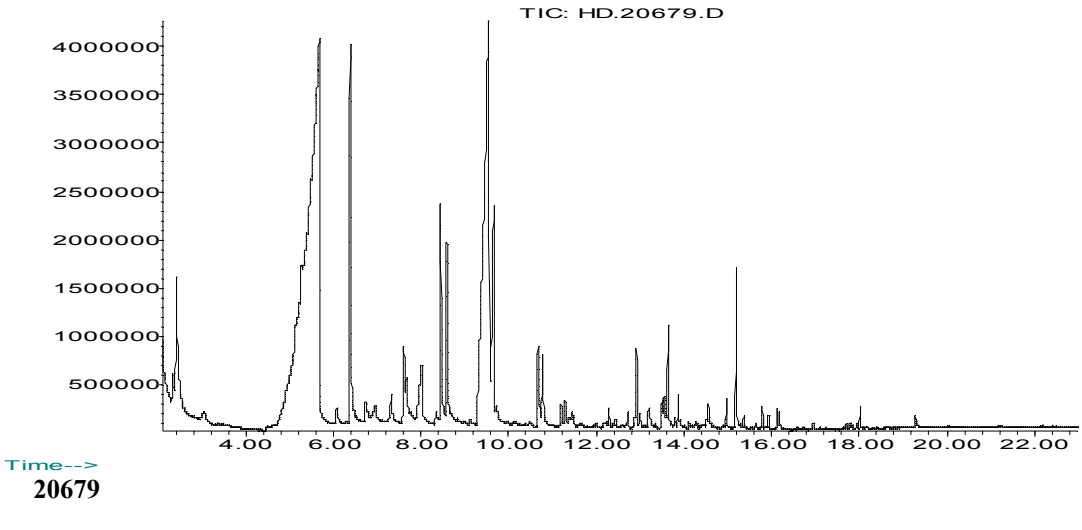

**20166**



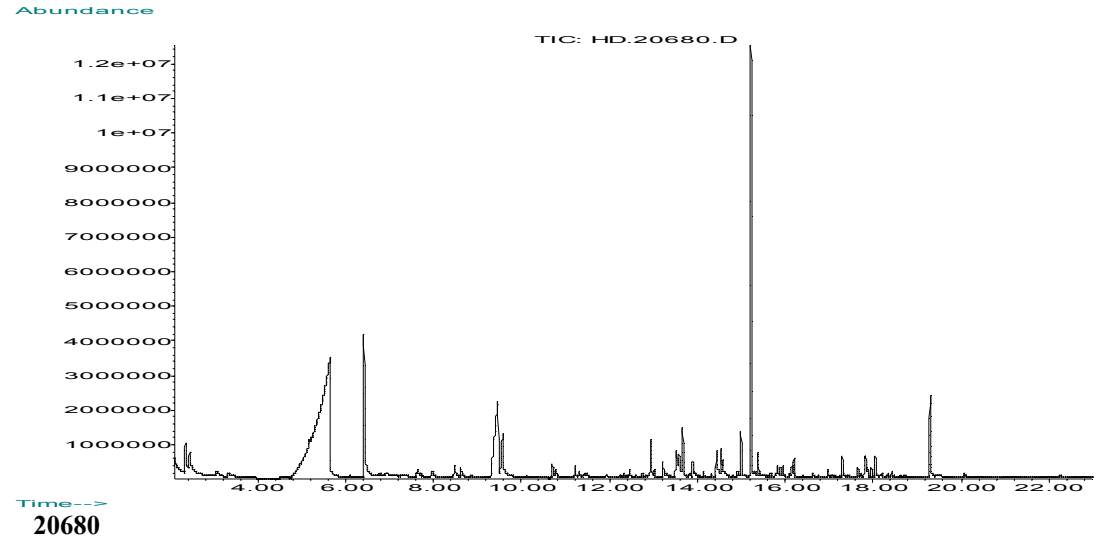

**20677**



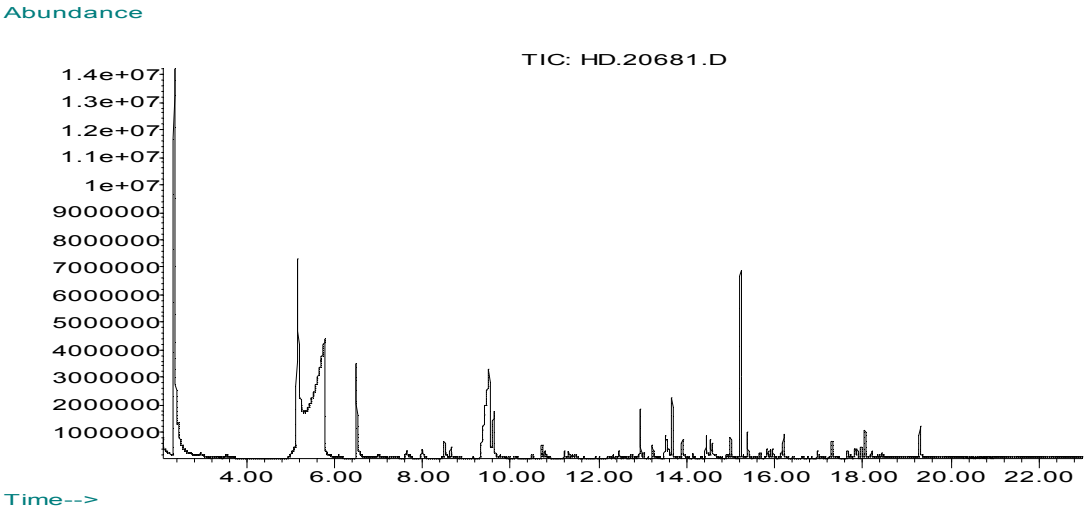

**20678**



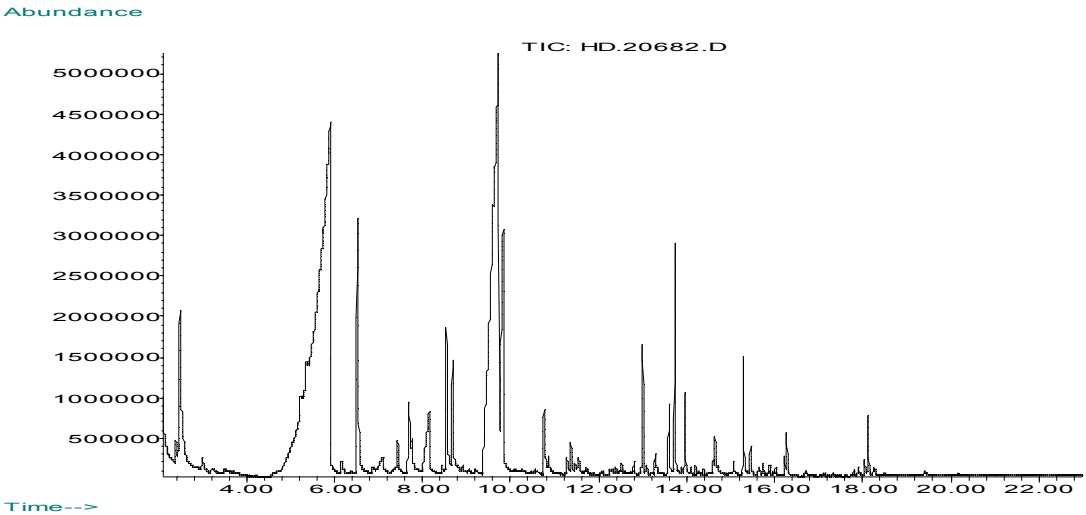

**20679**



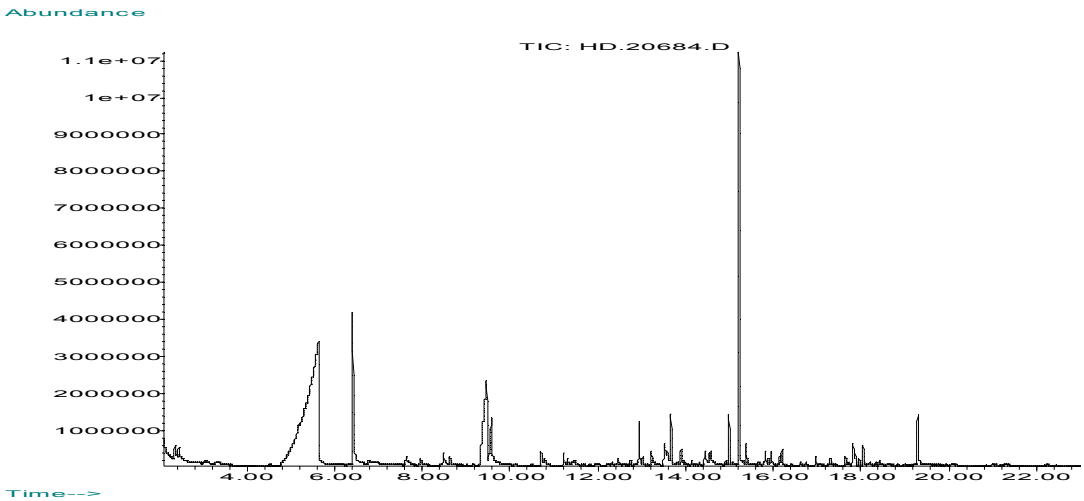

**20680**



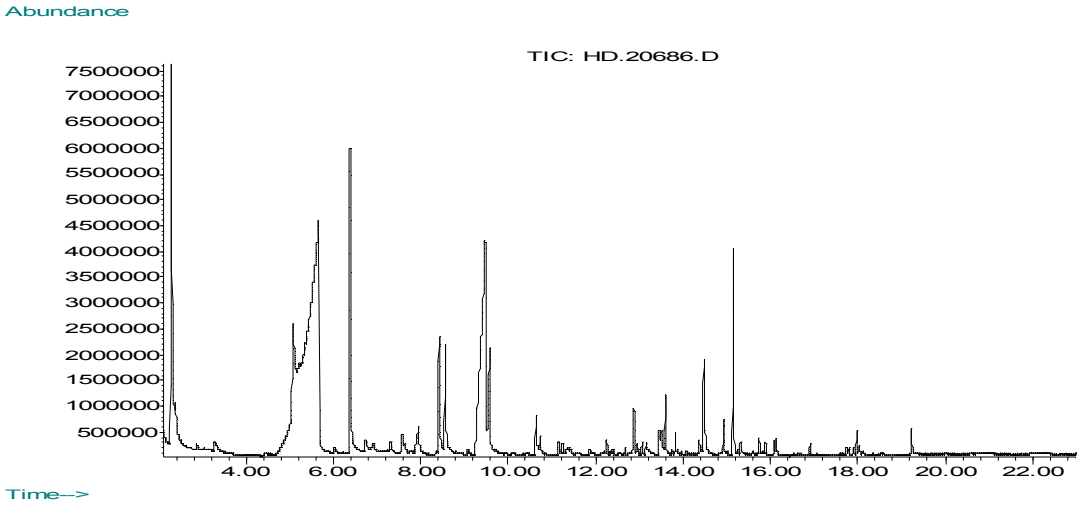

**20681**



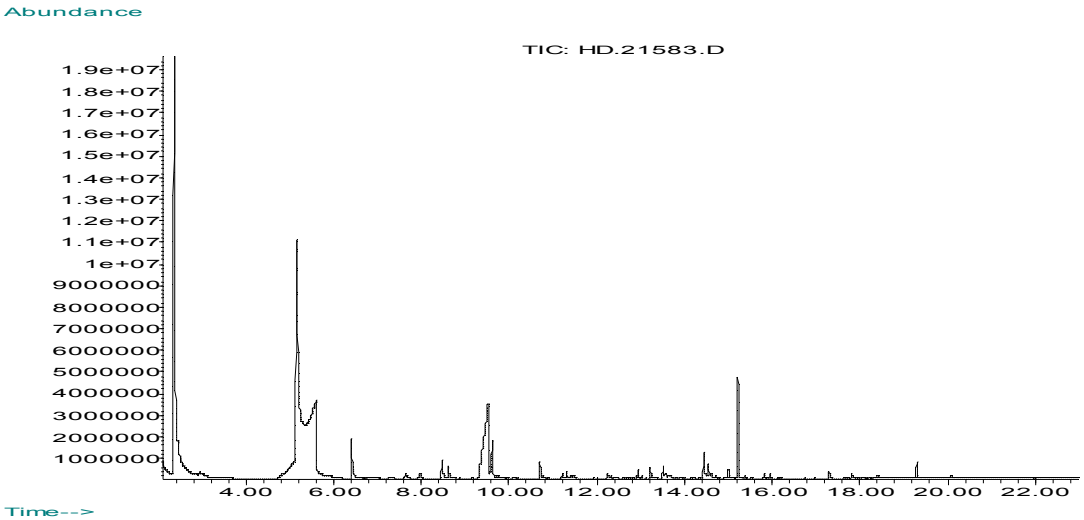

**20682**



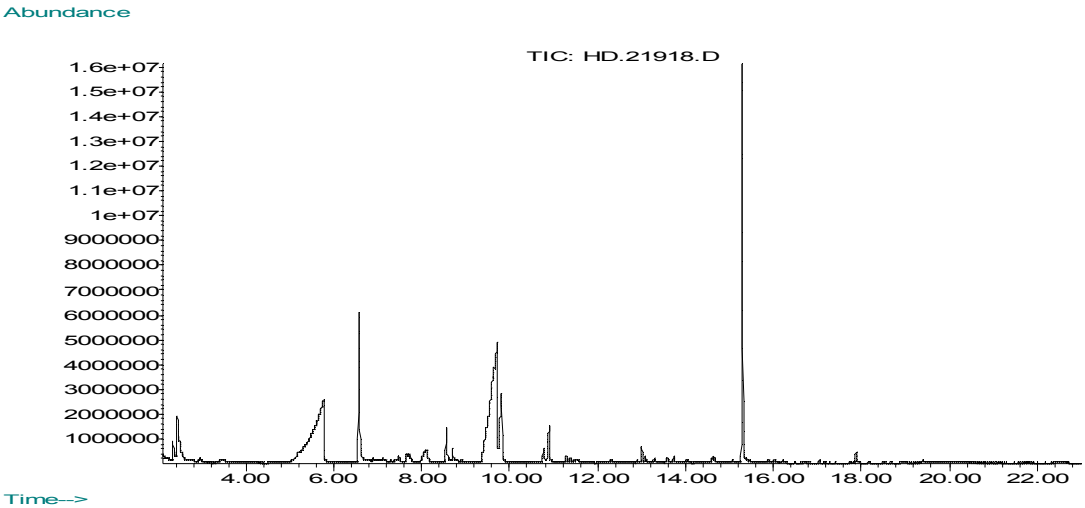

**20684**



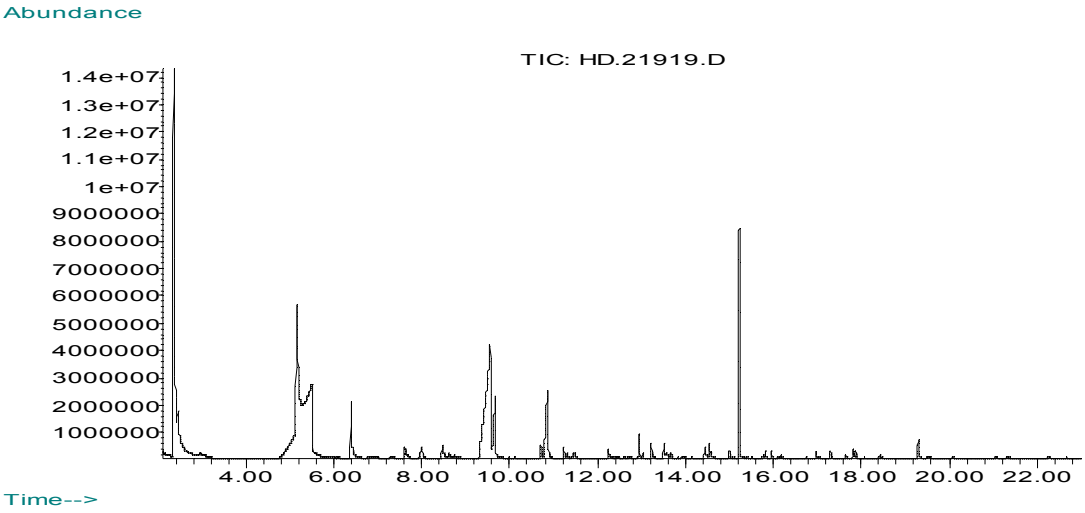

**20686**



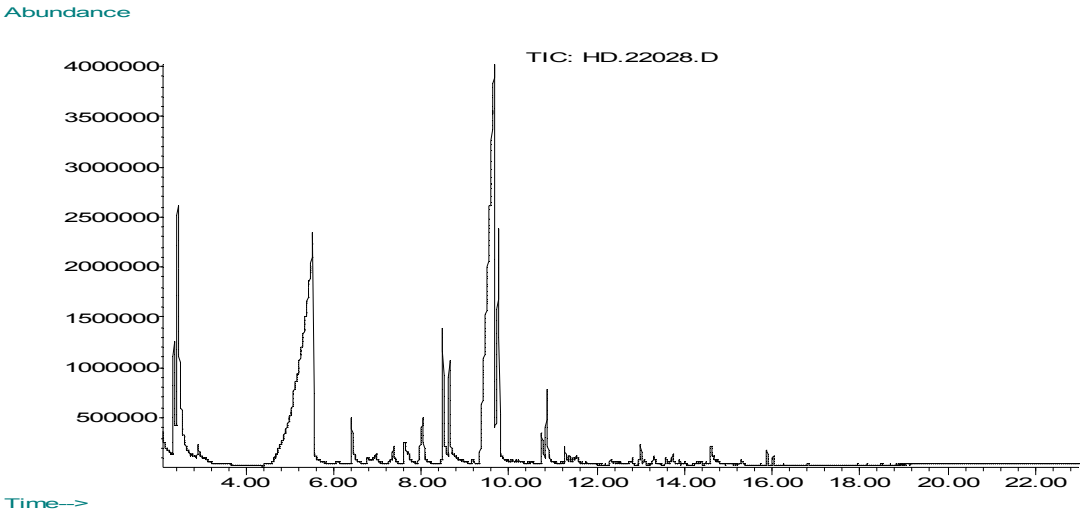

**21583**



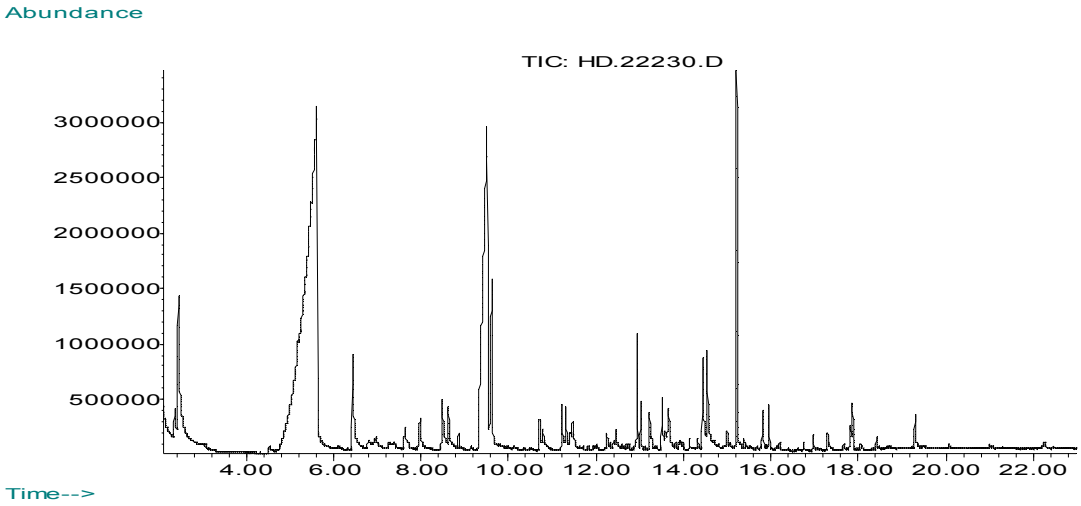

**21918**



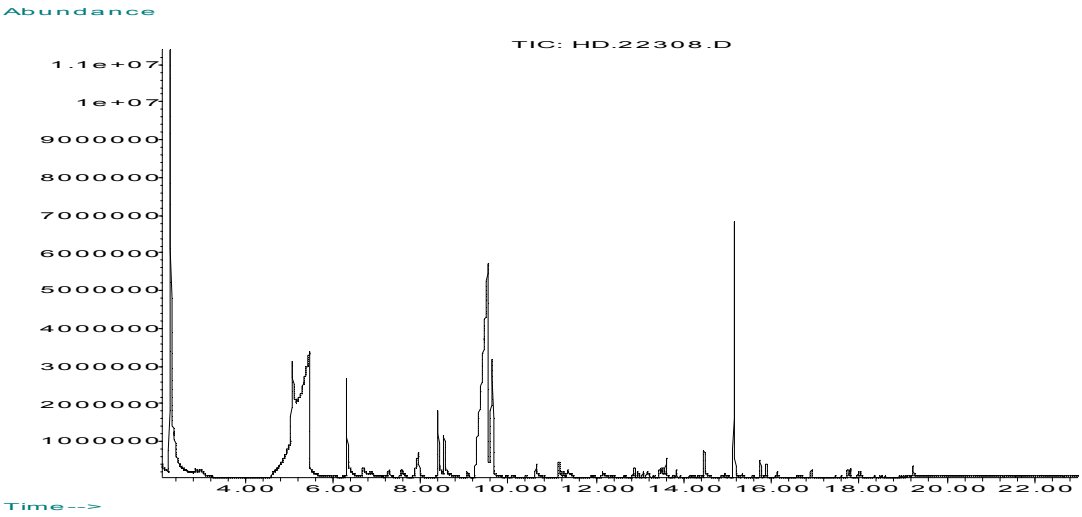

**21919**



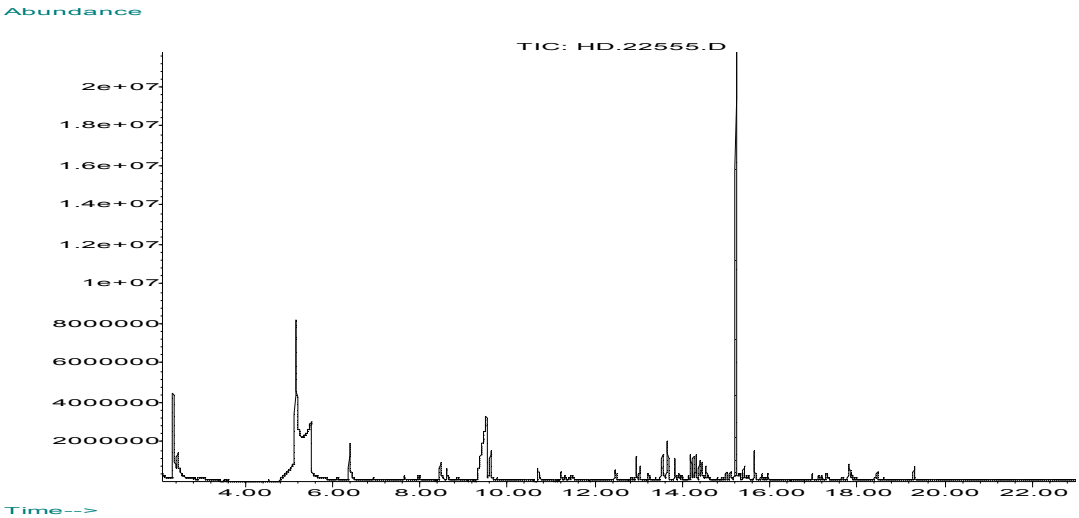

**22028**



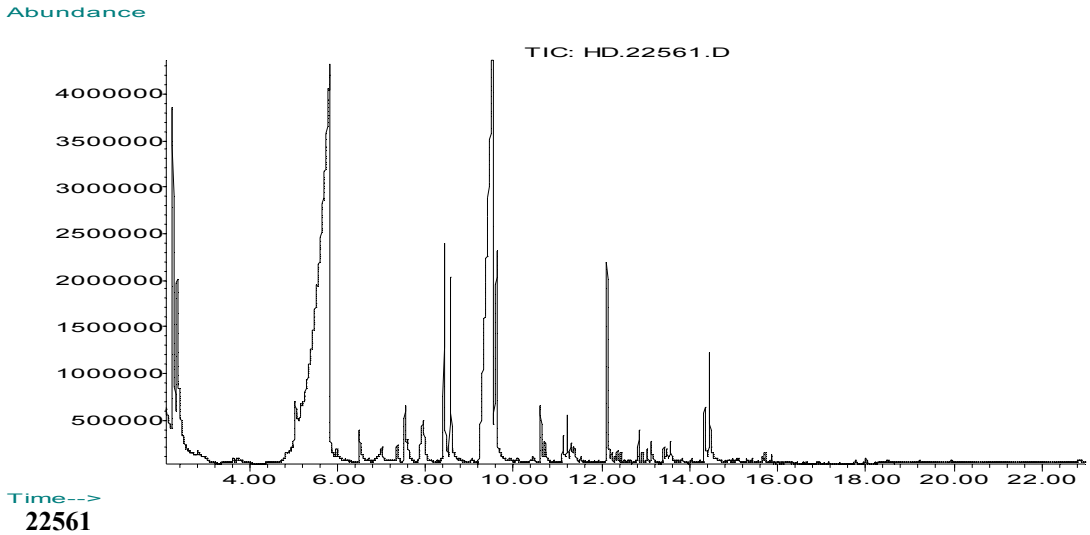

**22230**



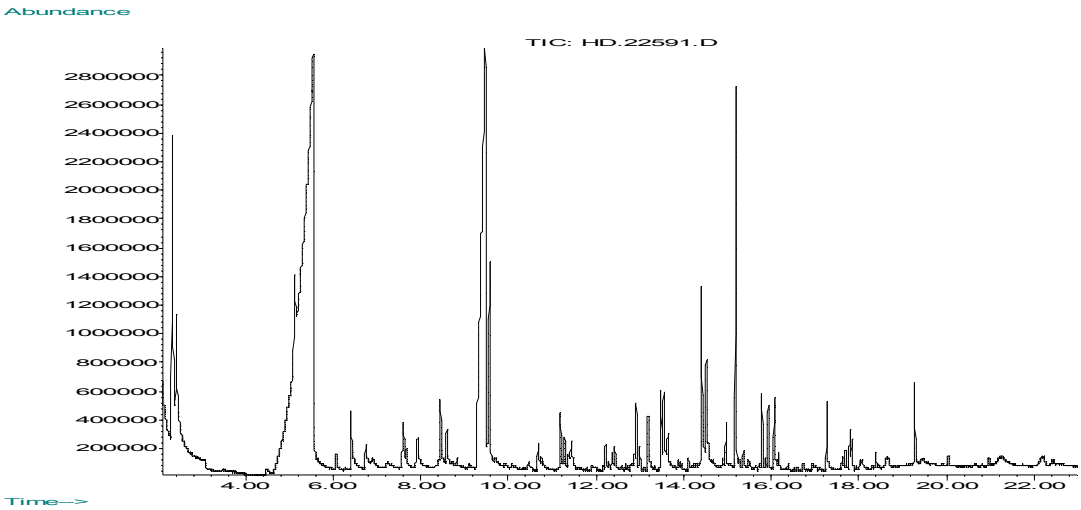

**22308**



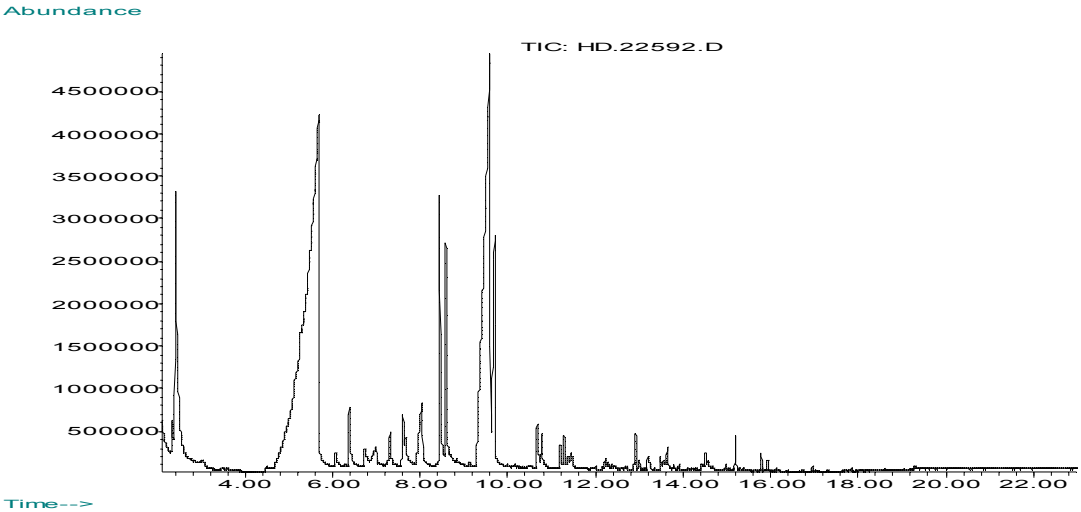

**22555**



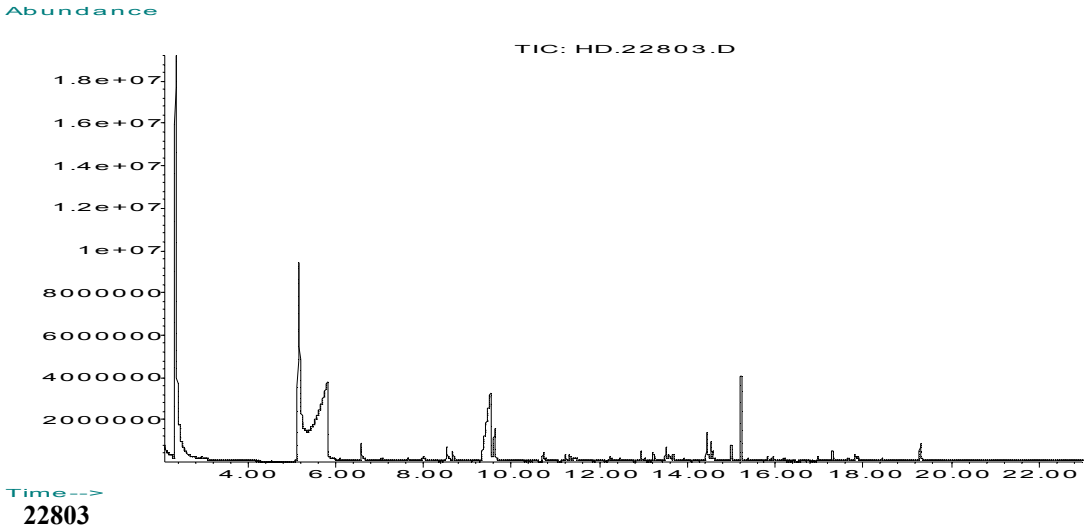

**22561**




**22591**

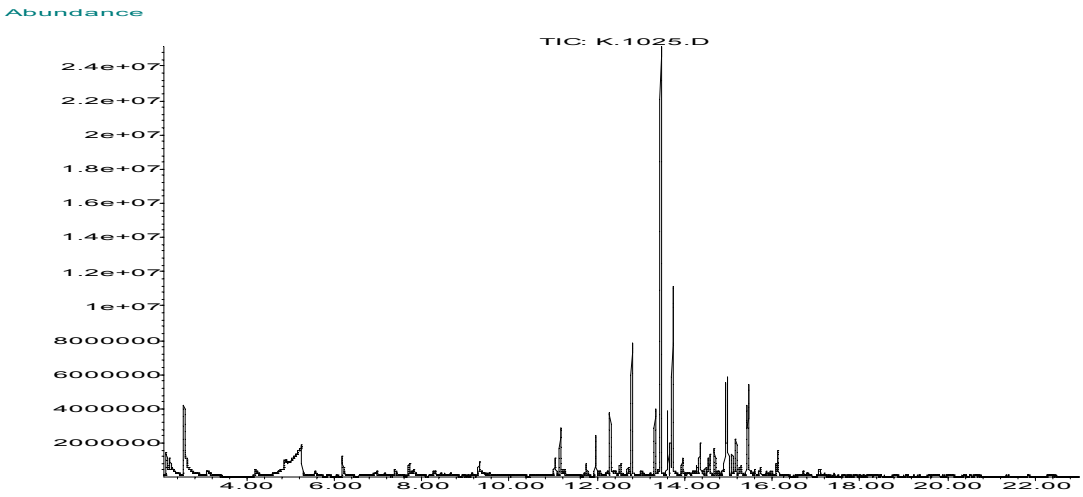



**22592**

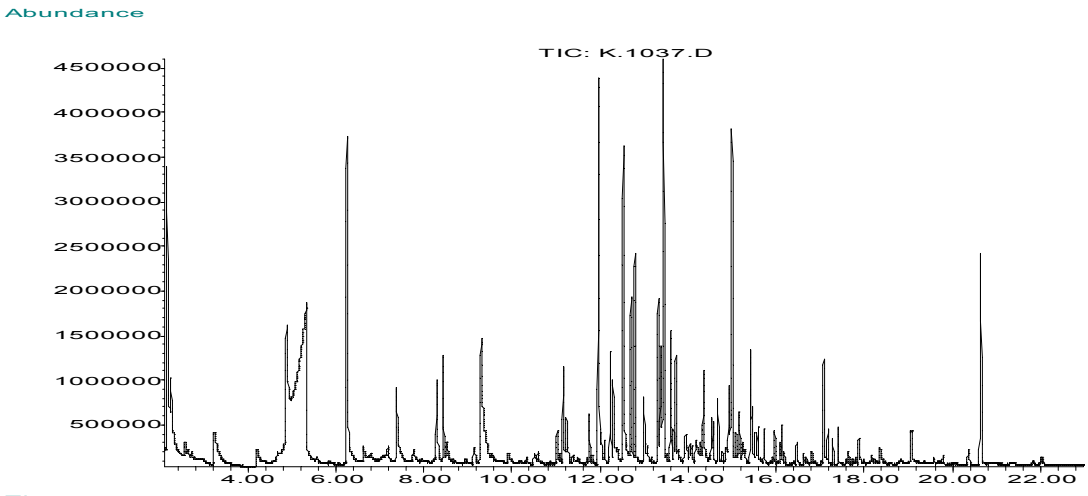



**22803**

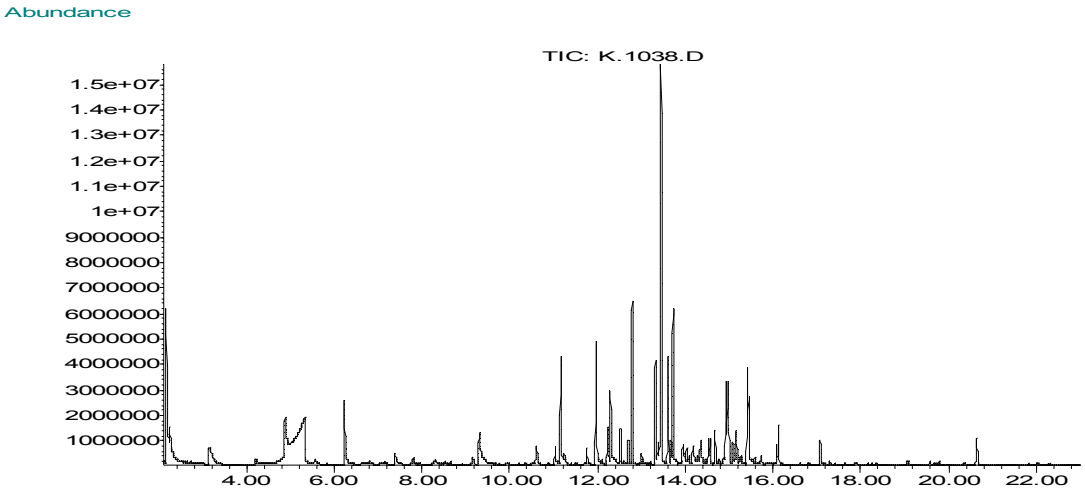
# Appendix 4: Chromatograms of Kamahi Honey Samples



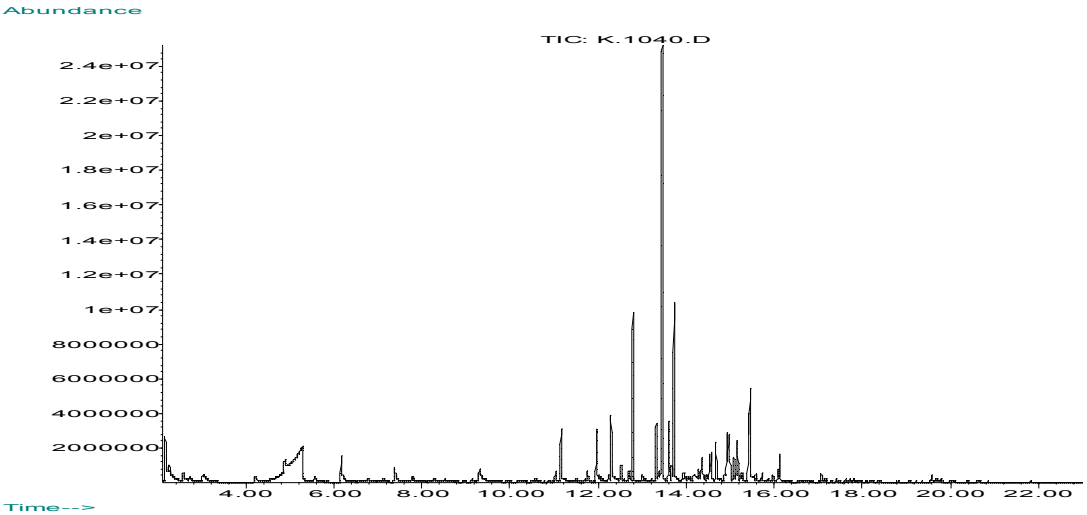

**1025**



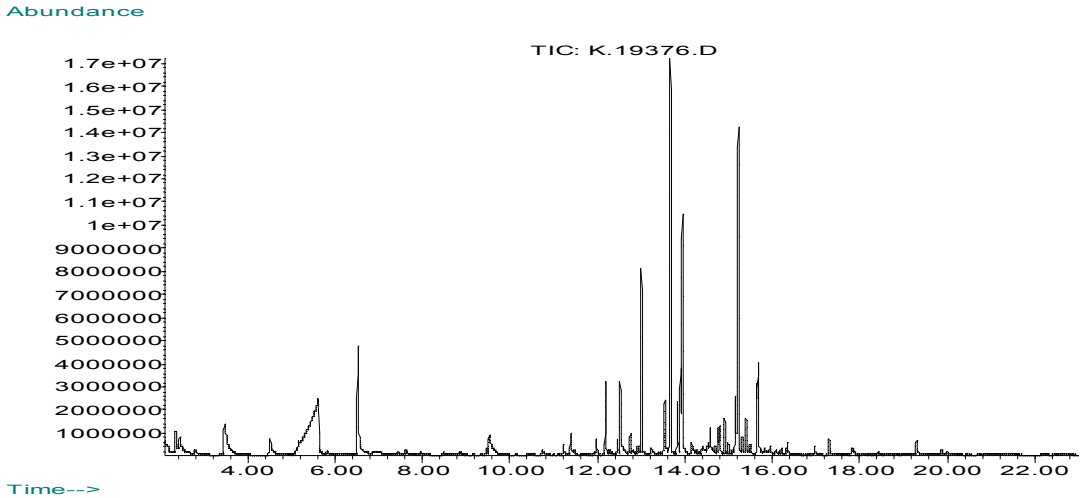

**1037**



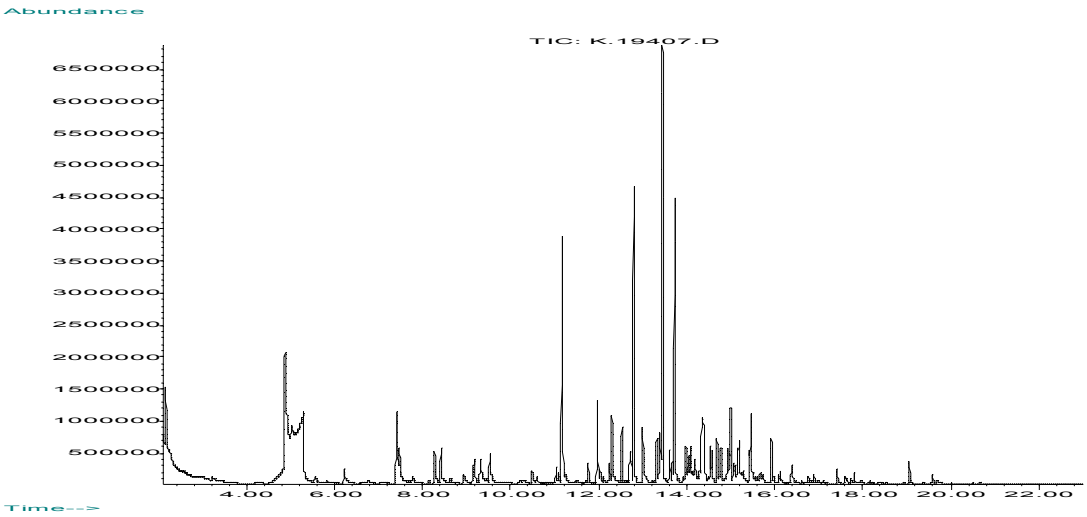

**1038**



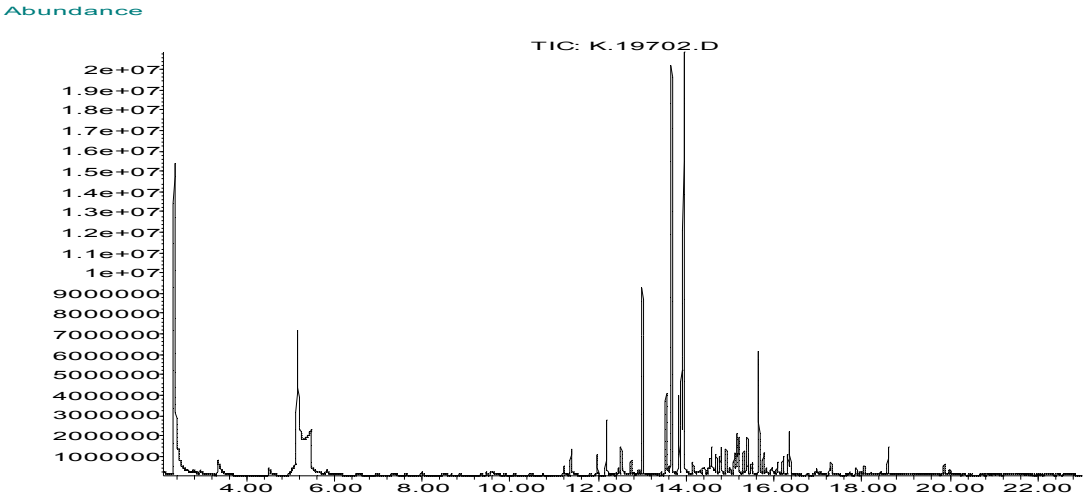

**1040**



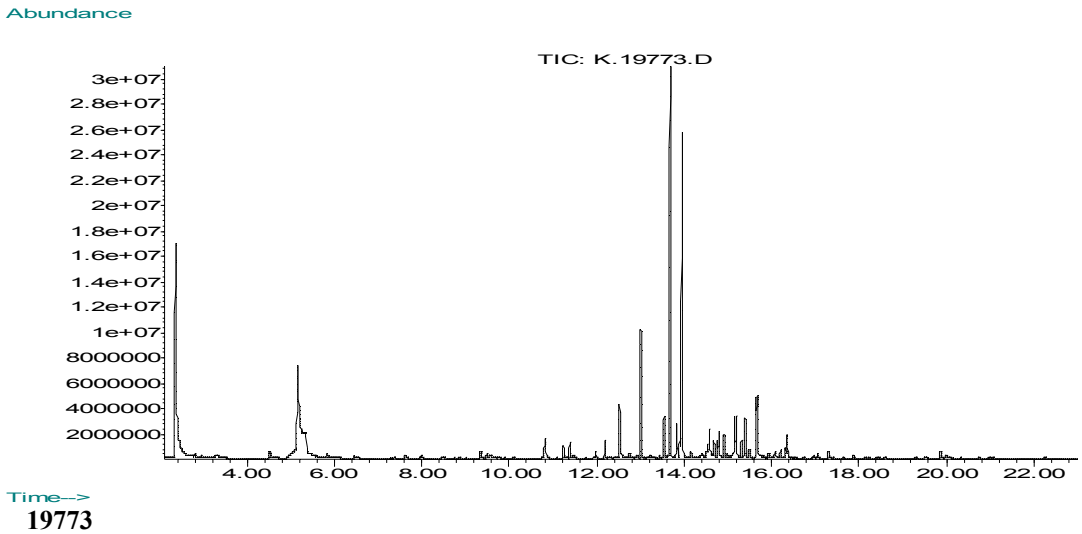

**19376**



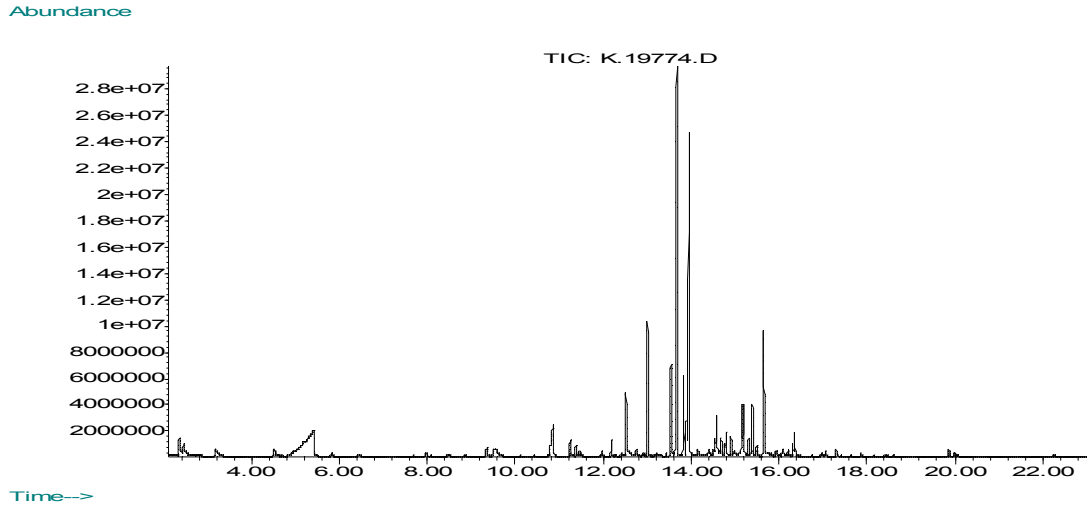

**19407**



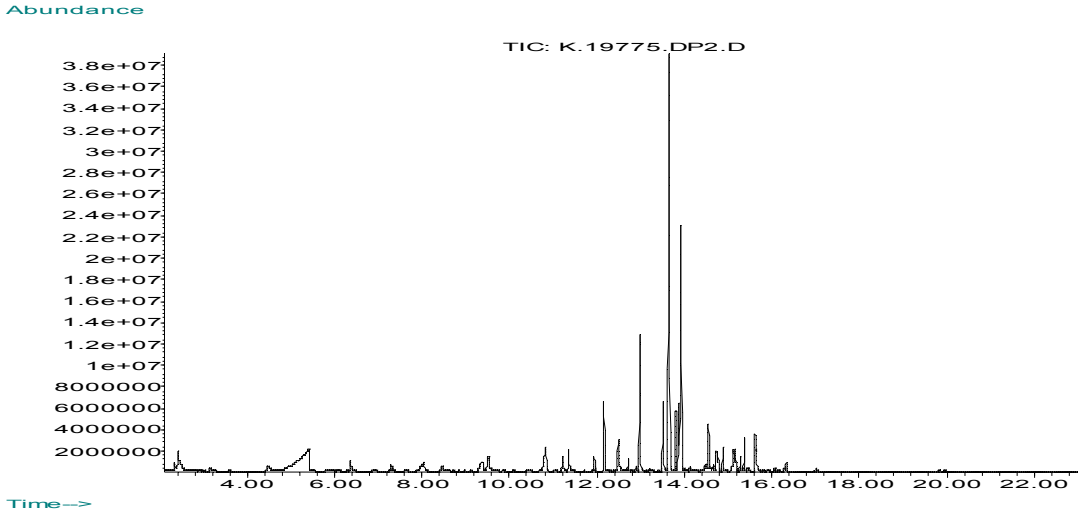

**19702**



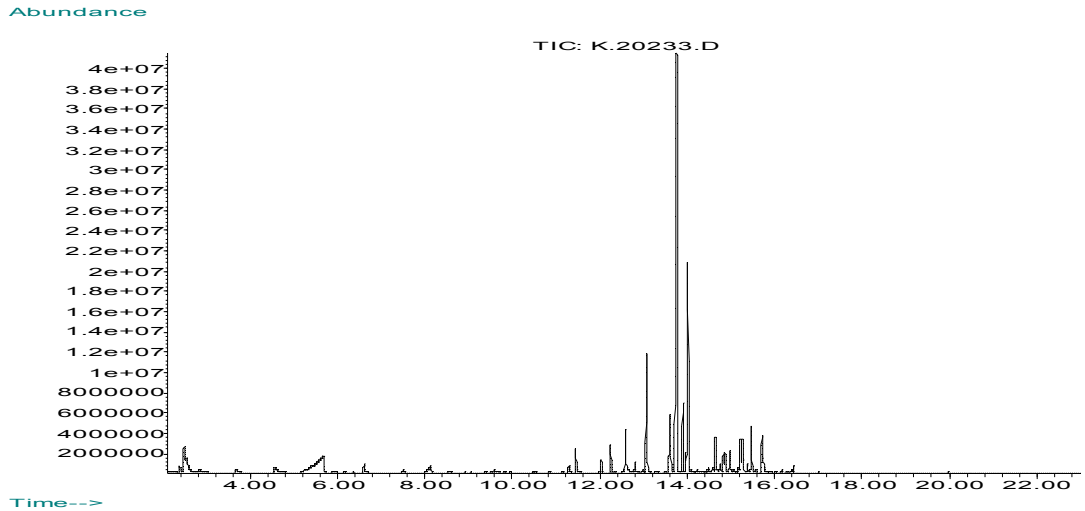

**19773**



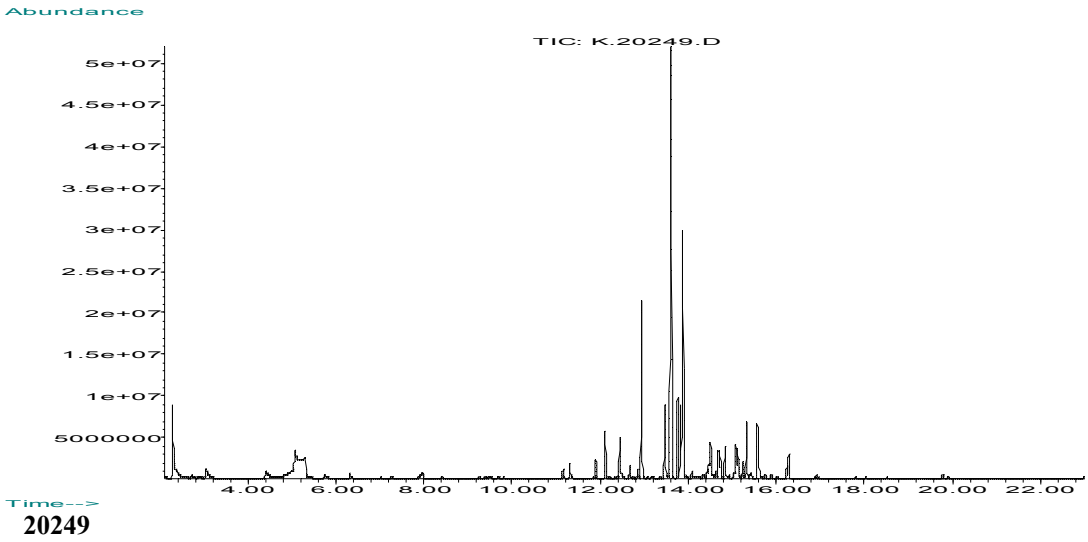

**19774**



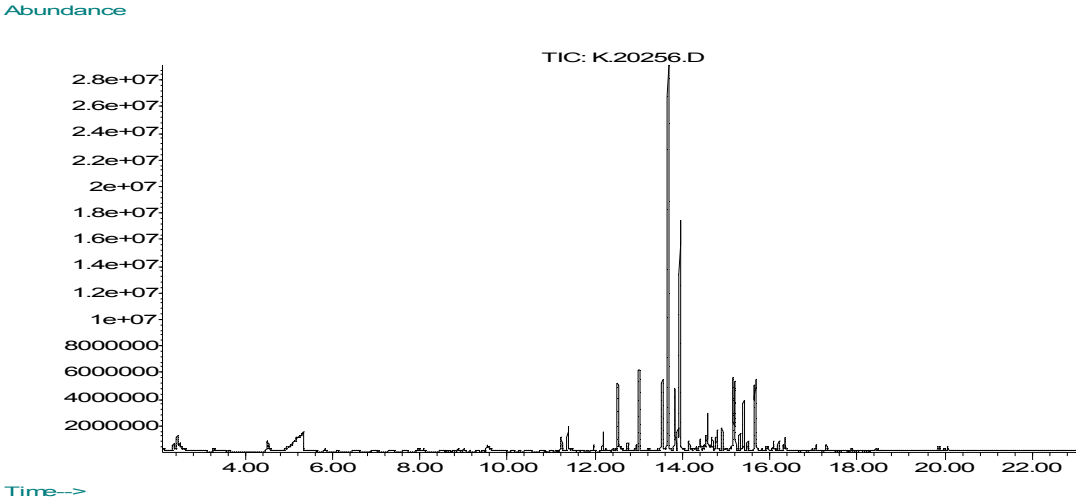

**19775**



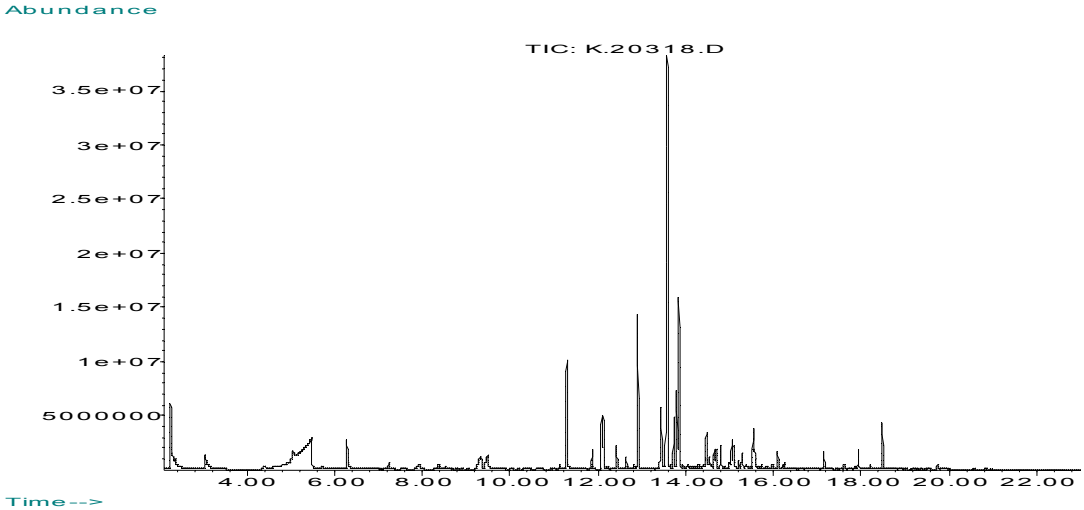

**20233**



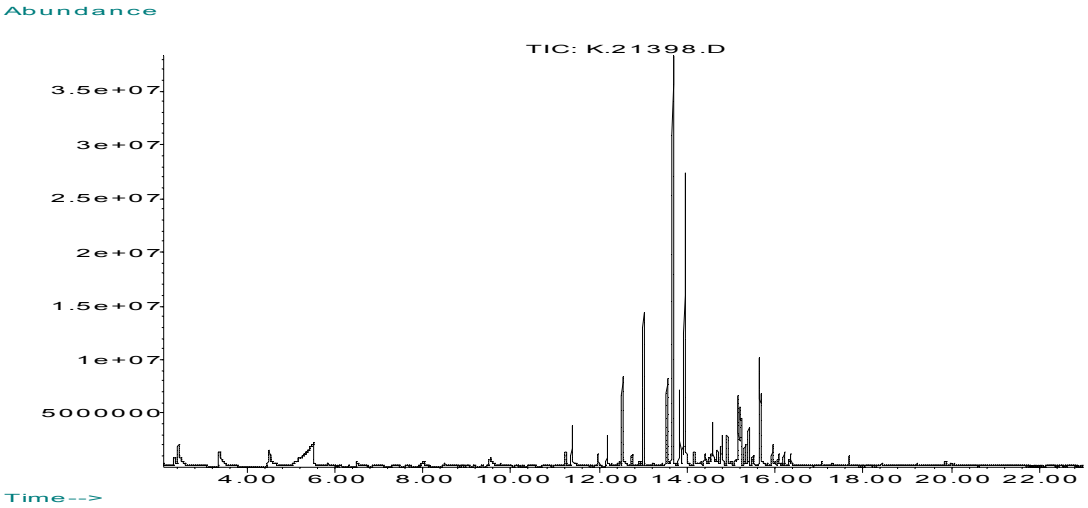

**20249**



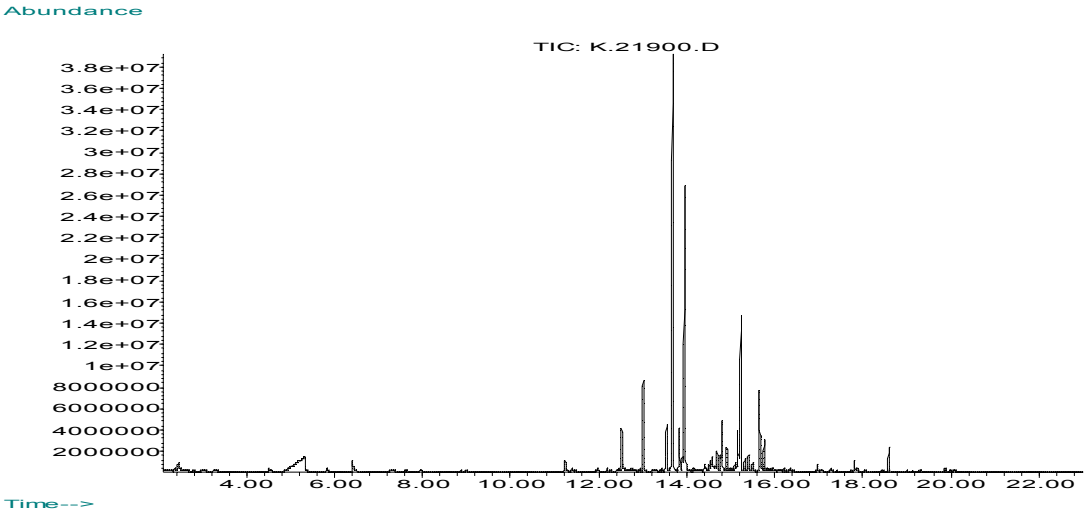

**20256**



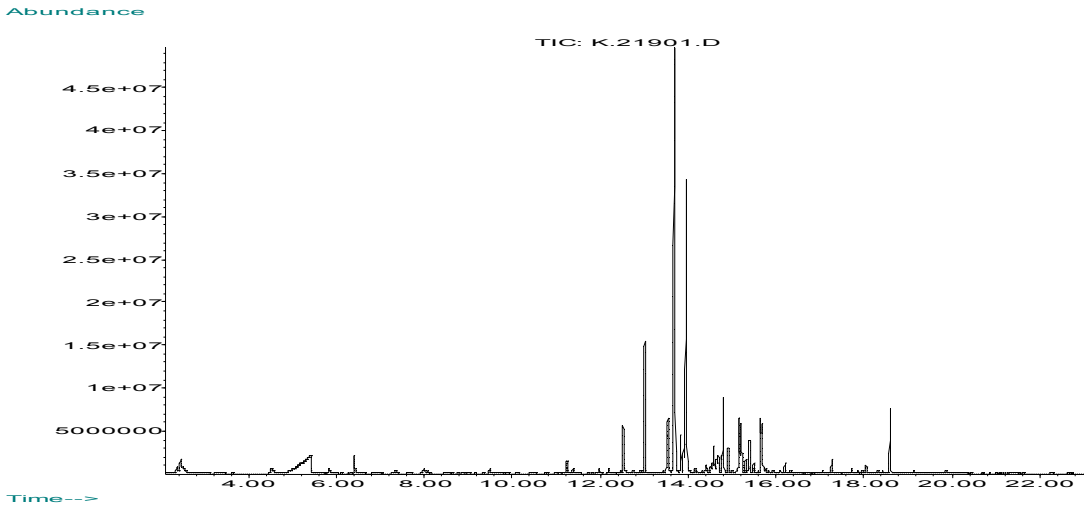

**20318**



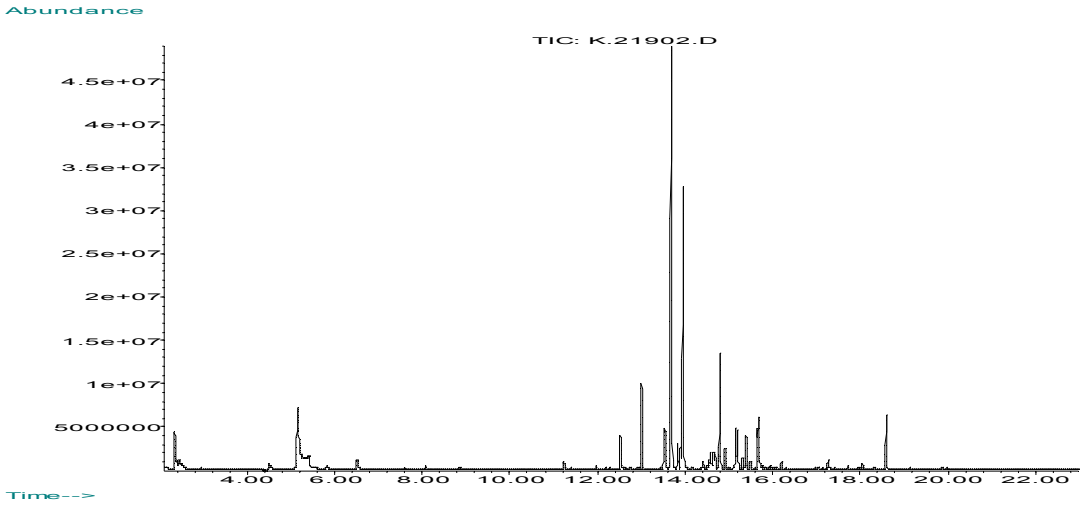

**21398**



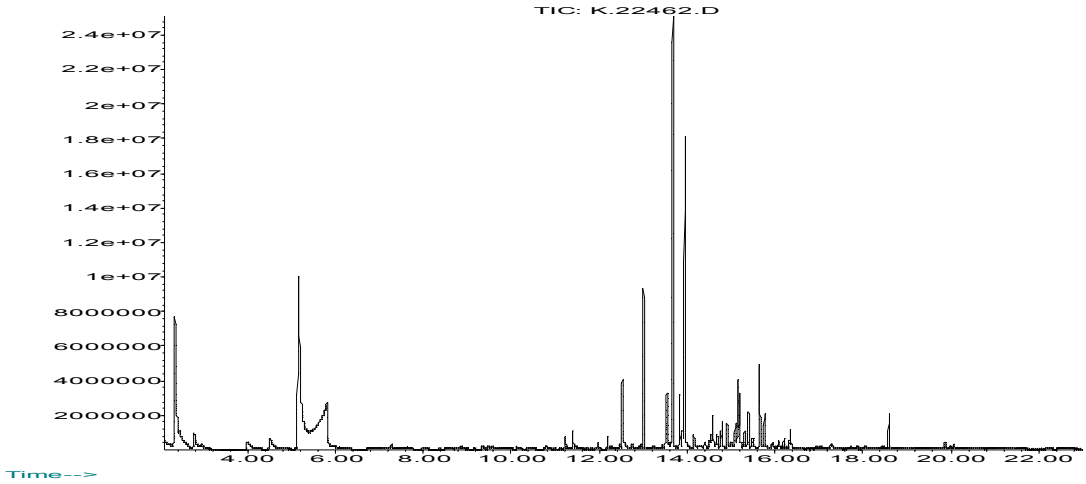

**21900**




**21901**

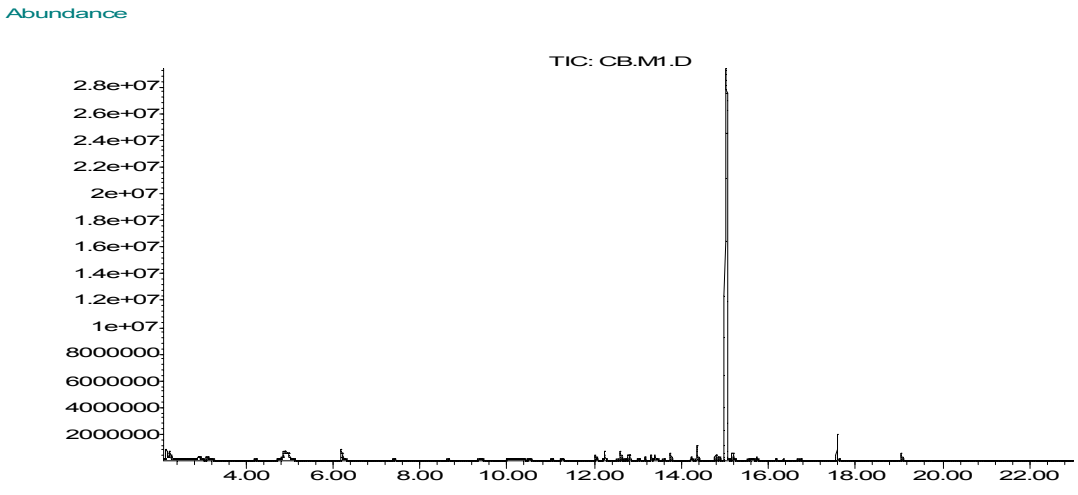

Abundance
TIC: K.21902.D
4.5e+07
4e+07
3.5e+07
3e+07
2.5e+07
2e+07
1.5e+07
1e+07
5000000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**21902**

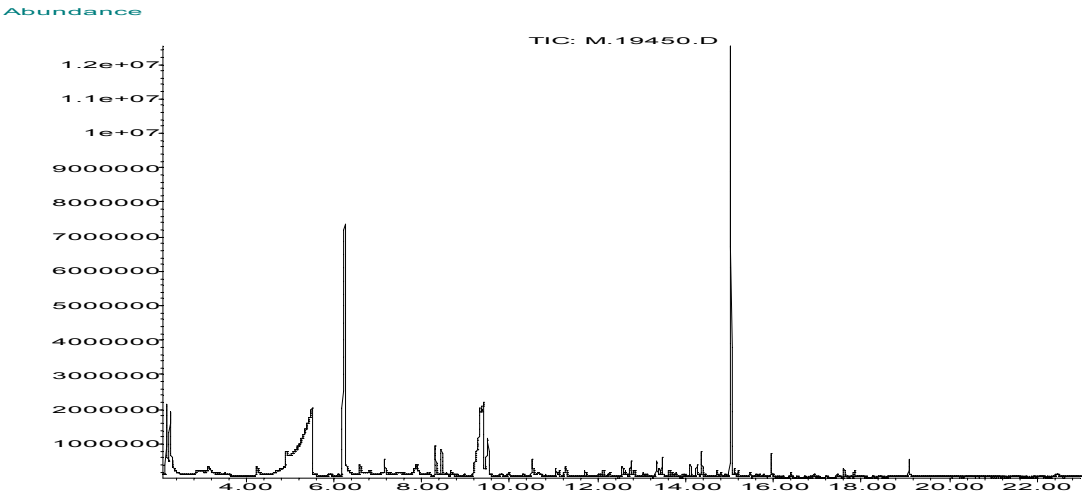

Abundance
TIC: K.22462.D
2.4e+07
2.2e+07
2e+07
1.8e+07
1.6e+07
1.4e+07
1.2e+07
1e+07
8000000
6000000
4000000
2000000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**22462**

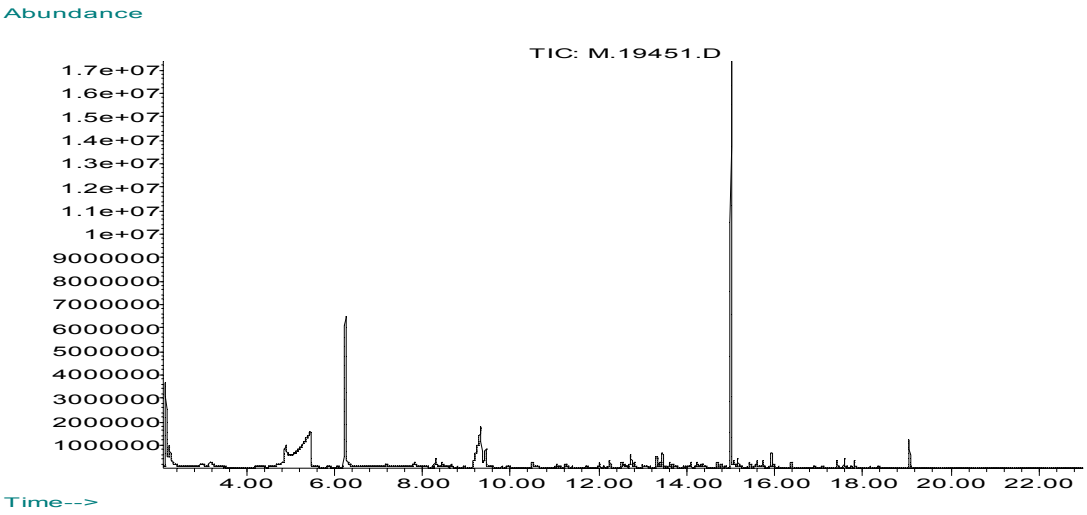
# Appendix 5: Chromatograms of Manuka Honey Samples



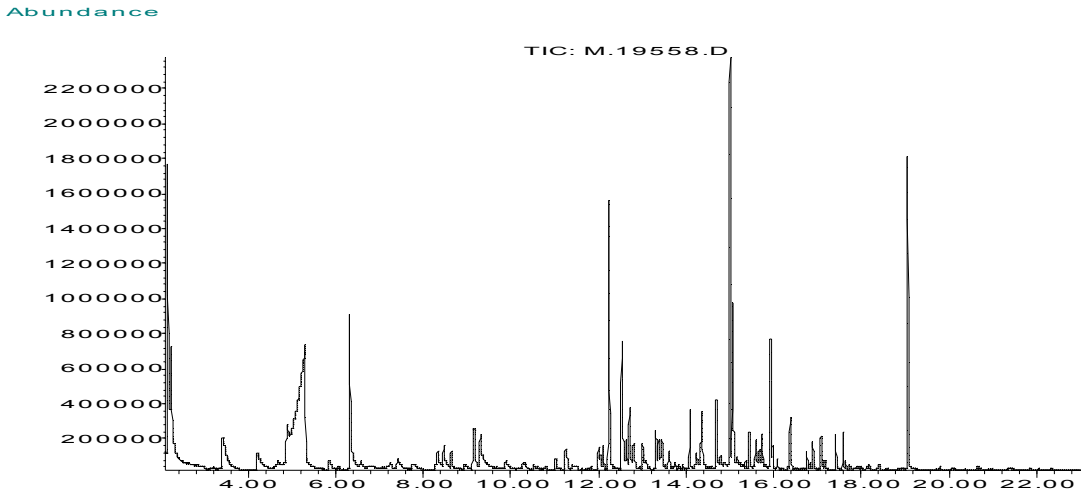

**CB.M1**



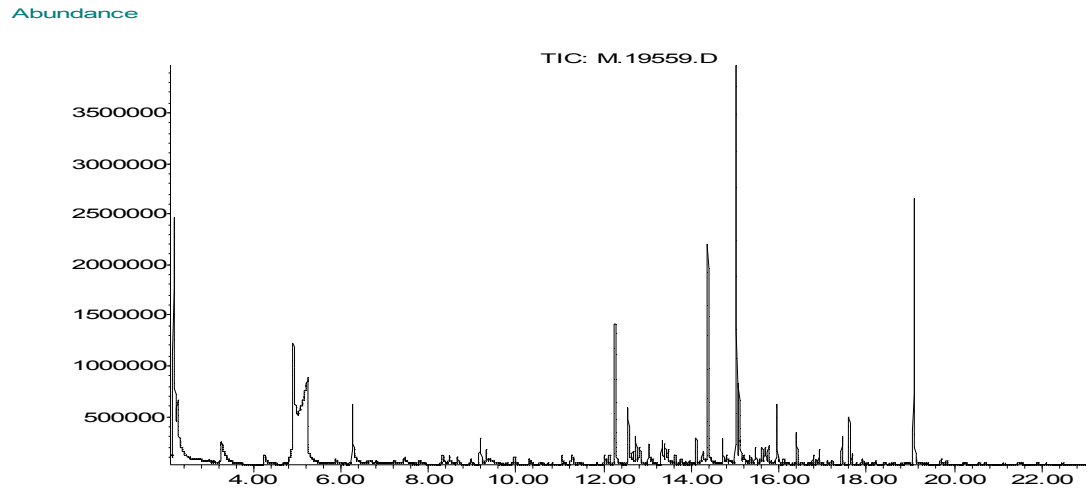

**19450**



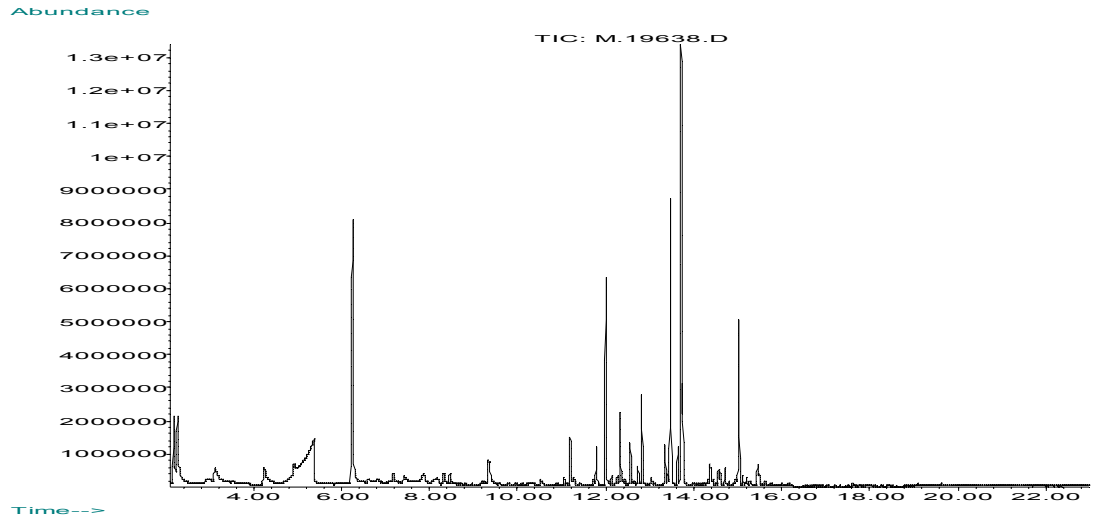

**19451**



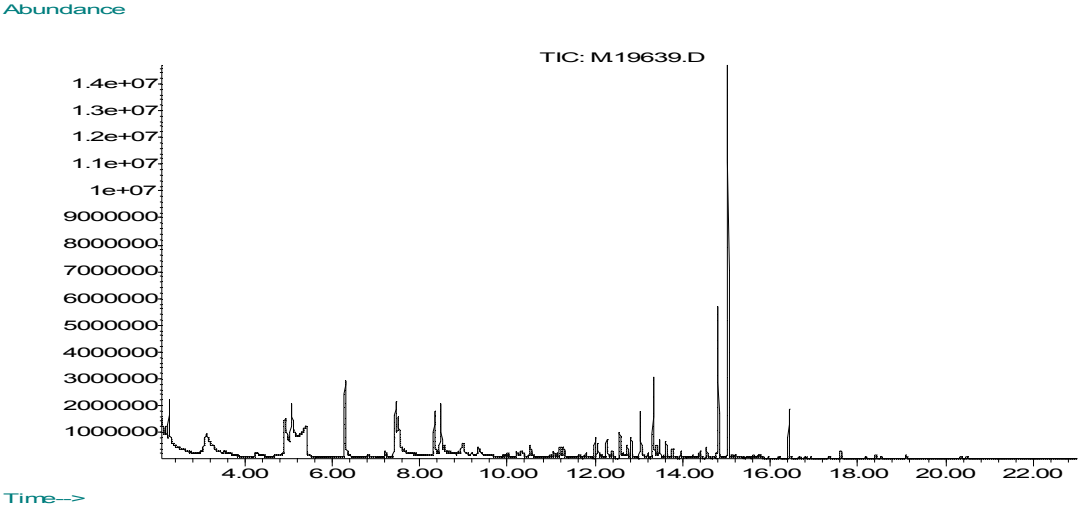

**19558**



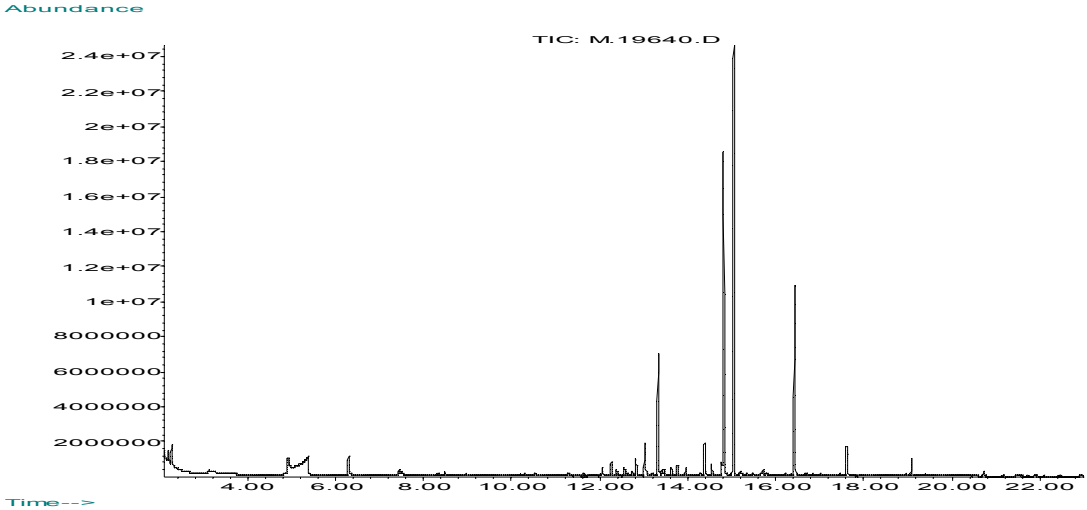

**19559**



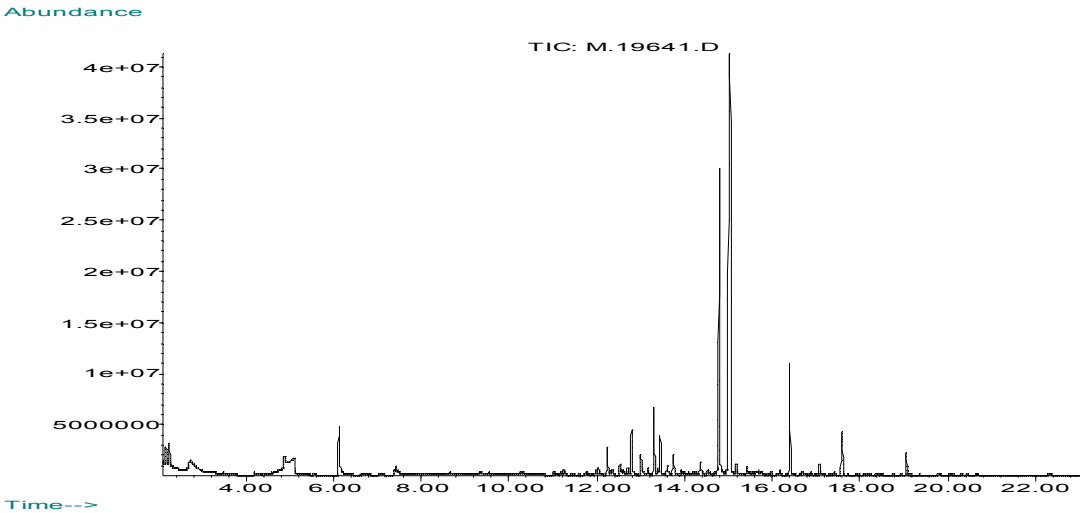

**19638**



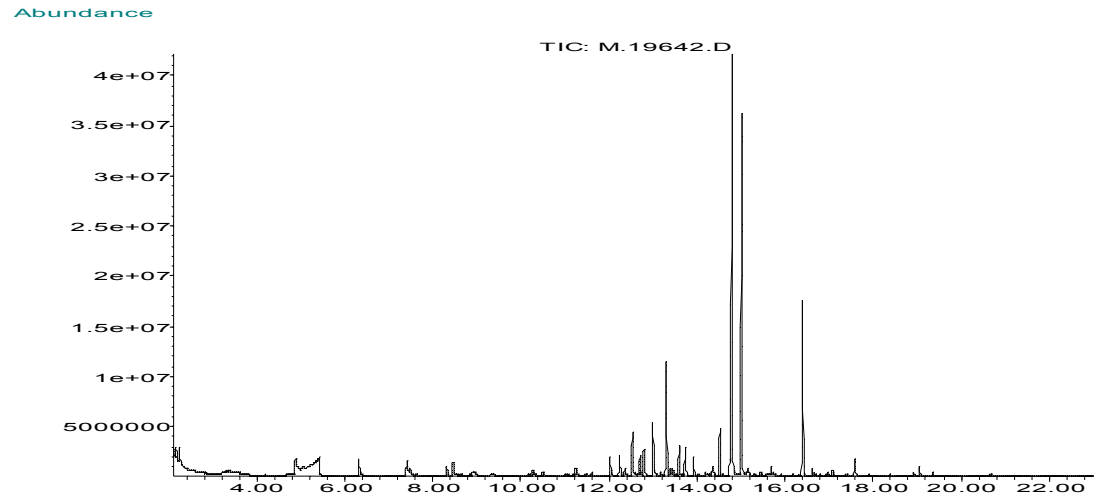

**19639**



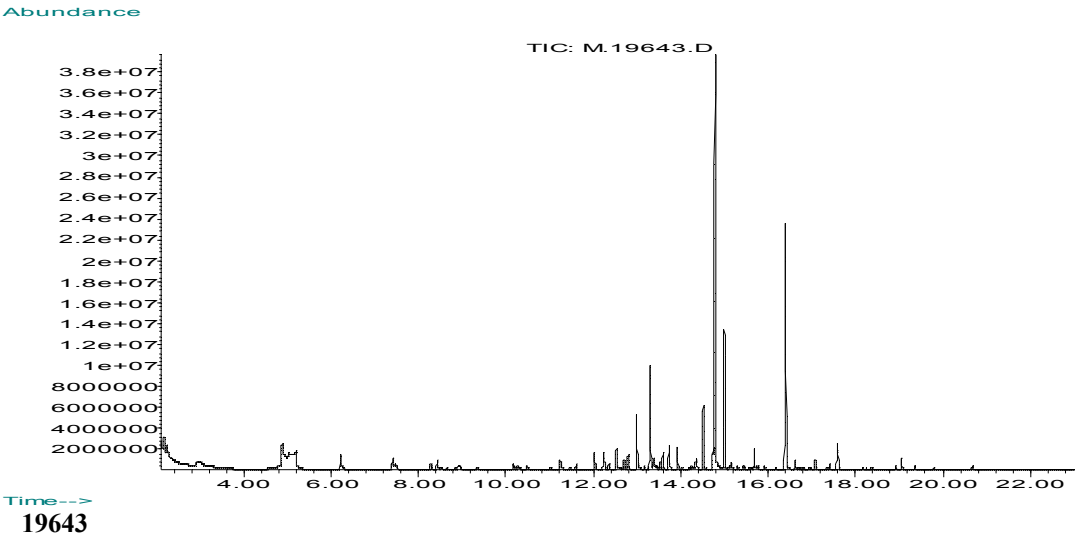

**19640**



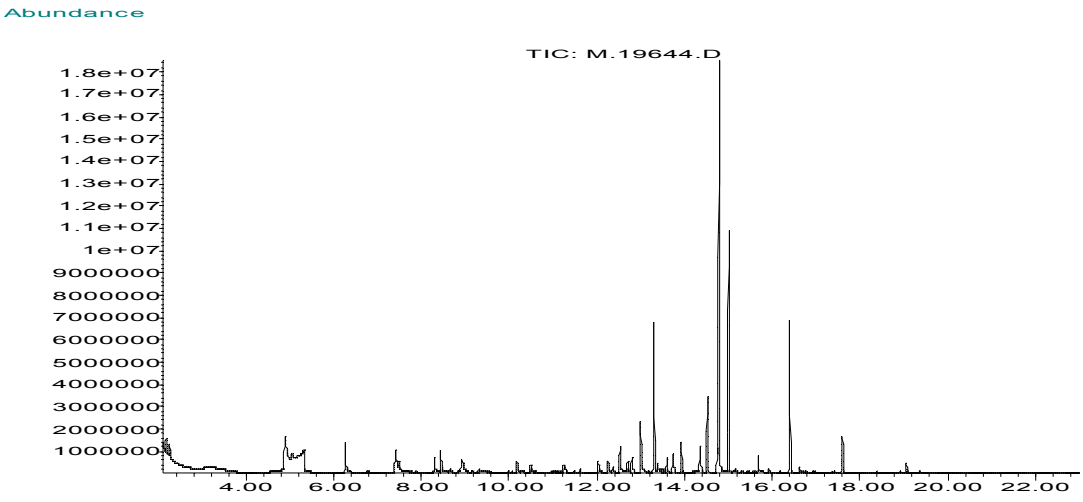

**19641**



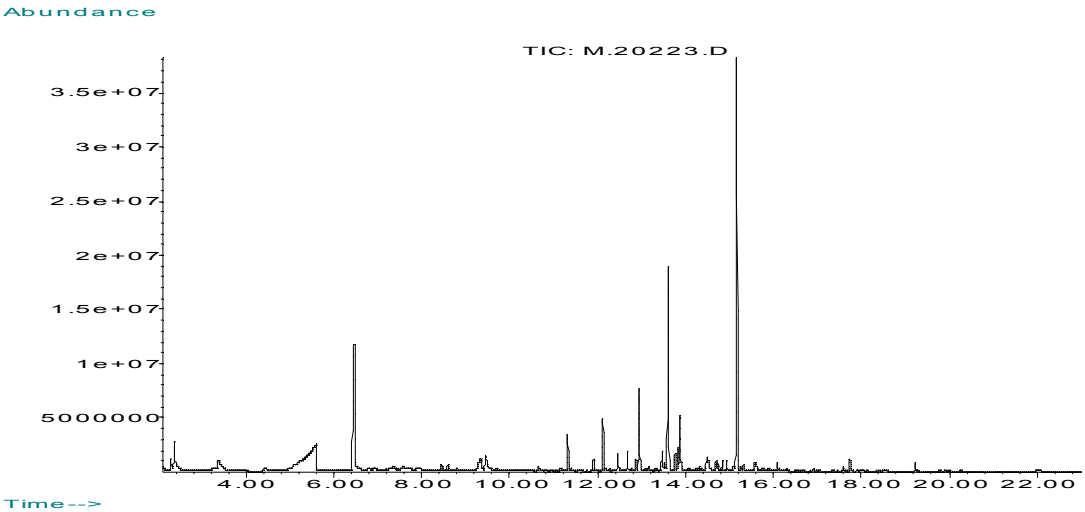

**19642**



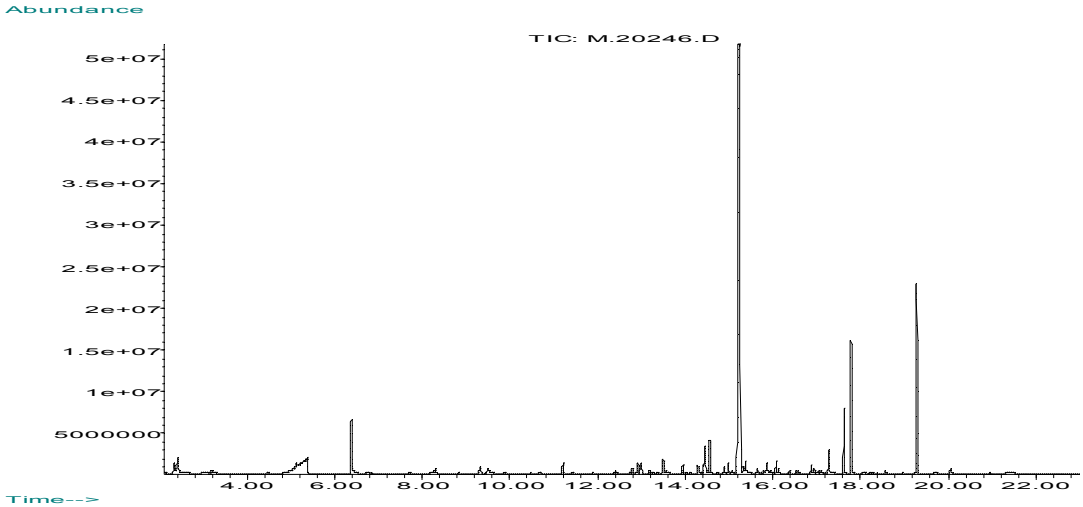

**19643**



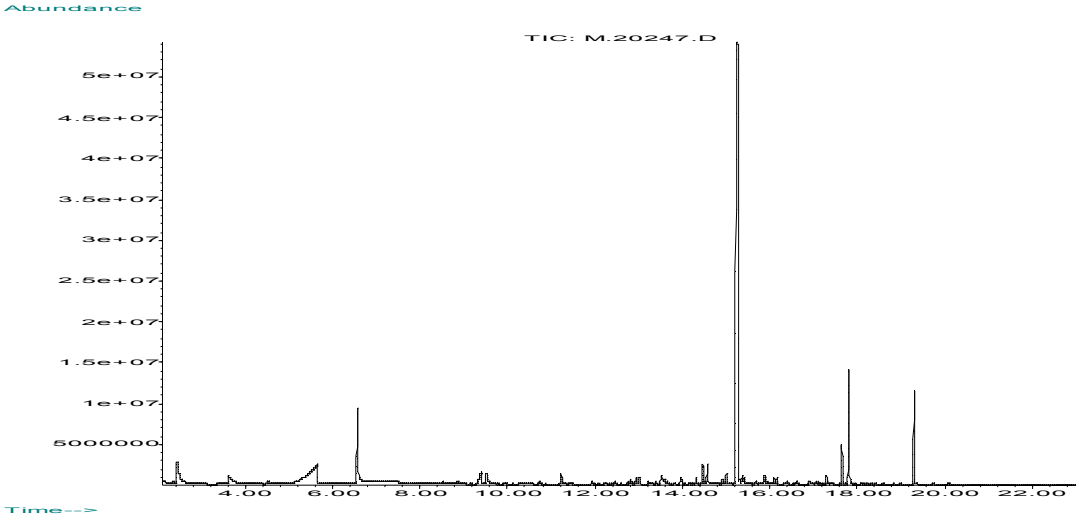

**19644**



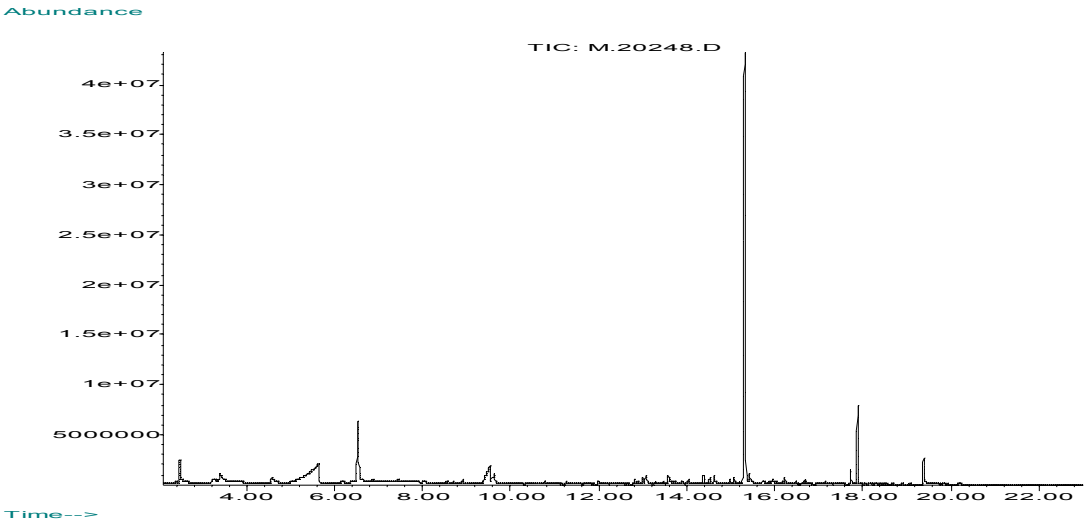

**20223**



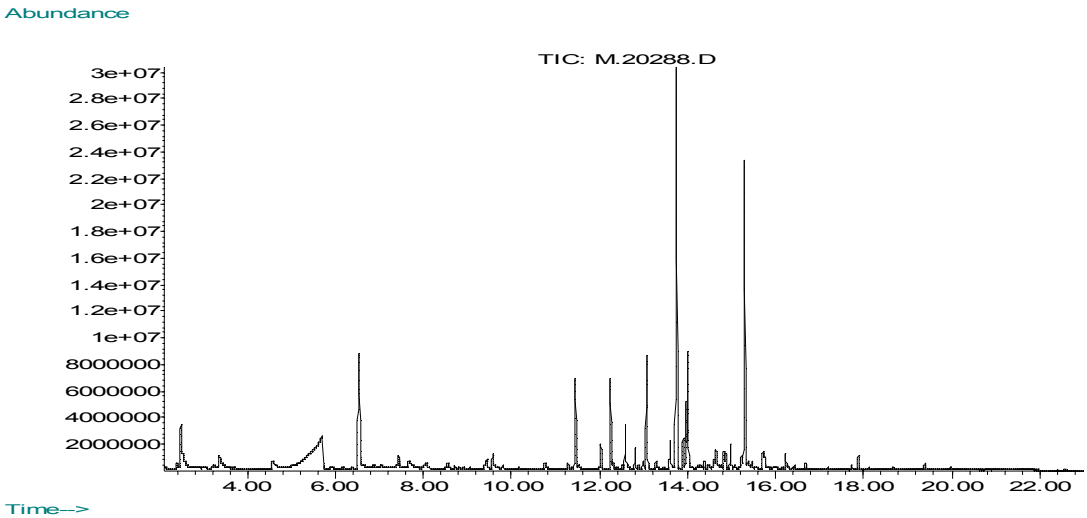

**20246**



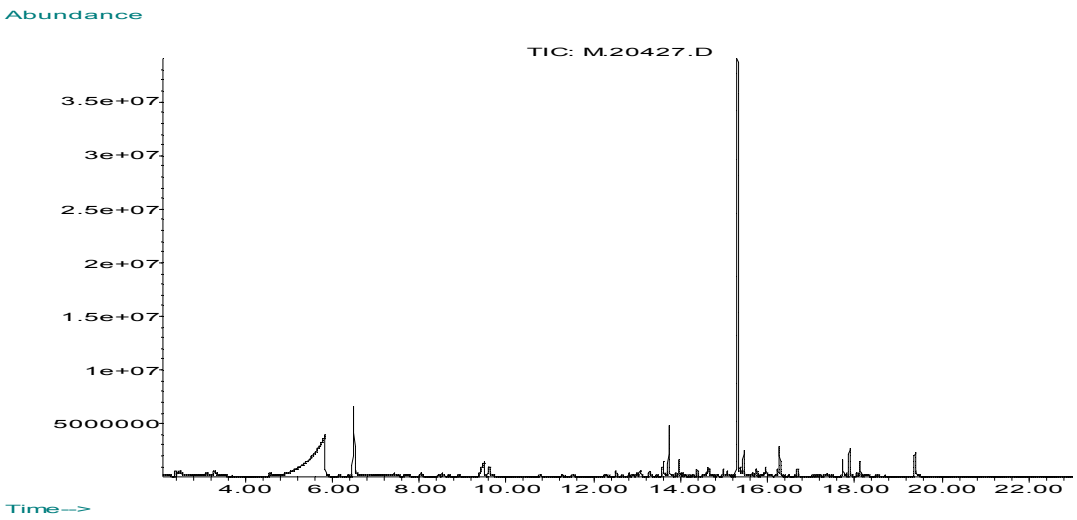

**20247**



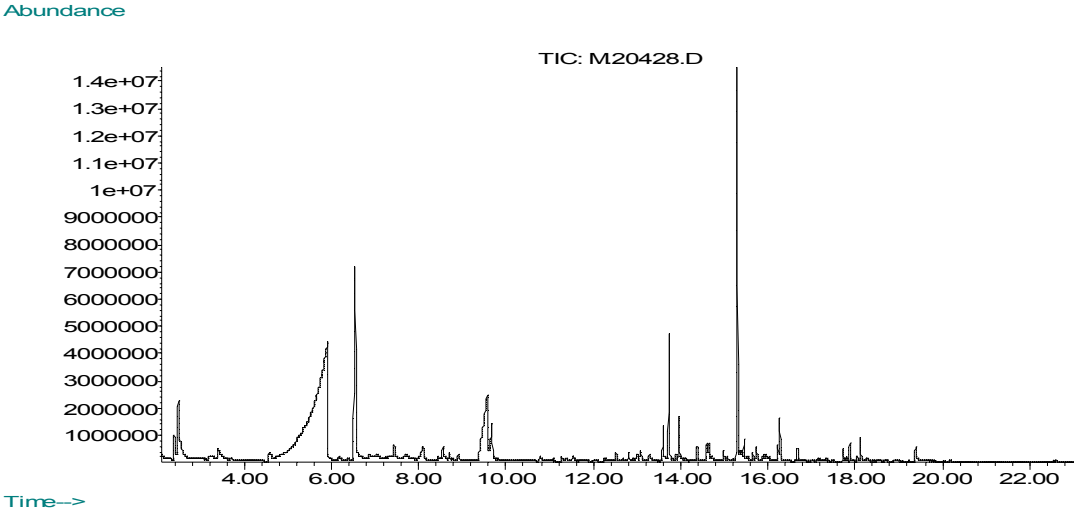

**20248**



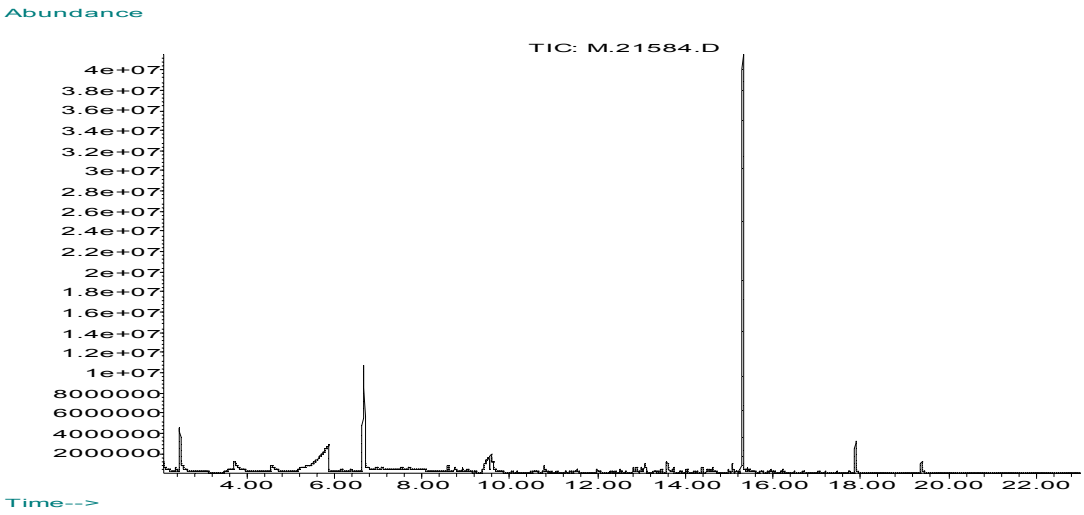

**20288**



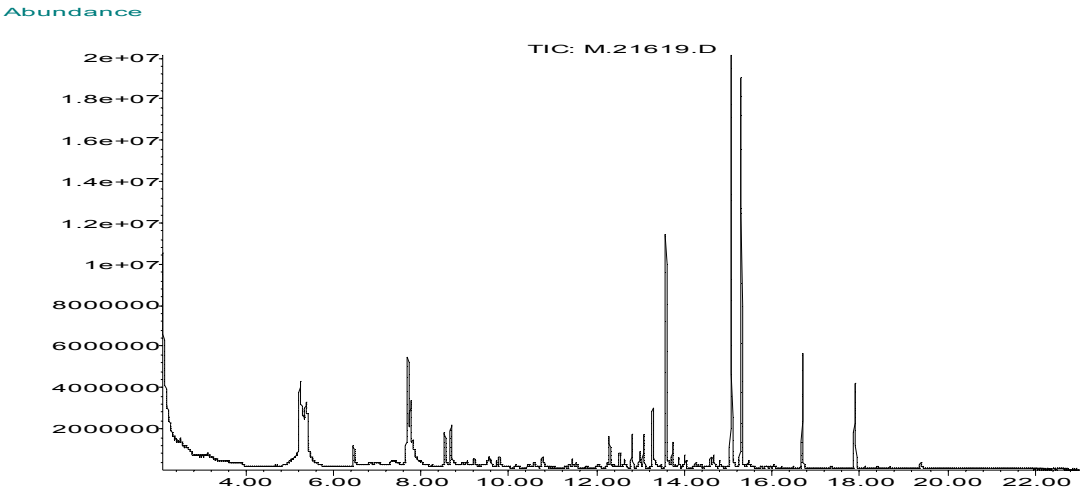

**20427**



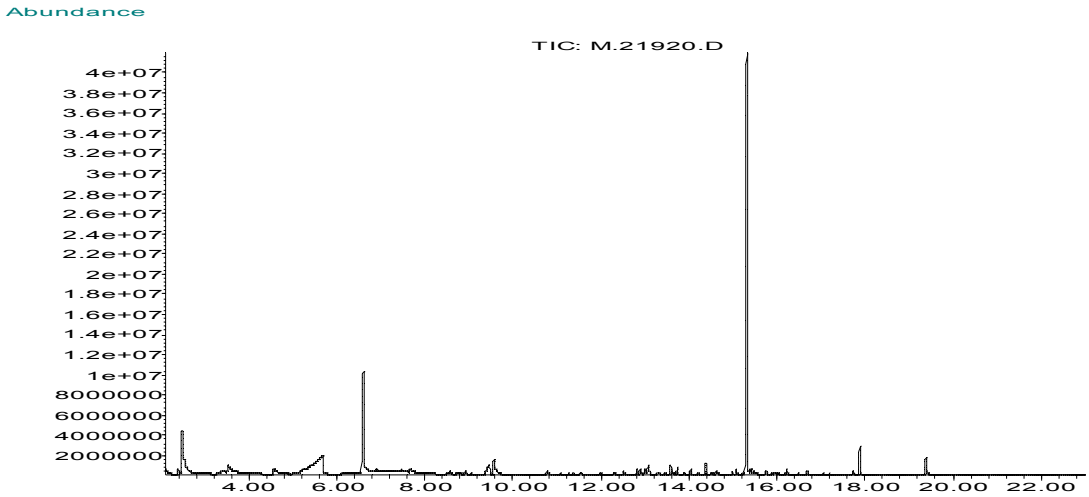

**20428**



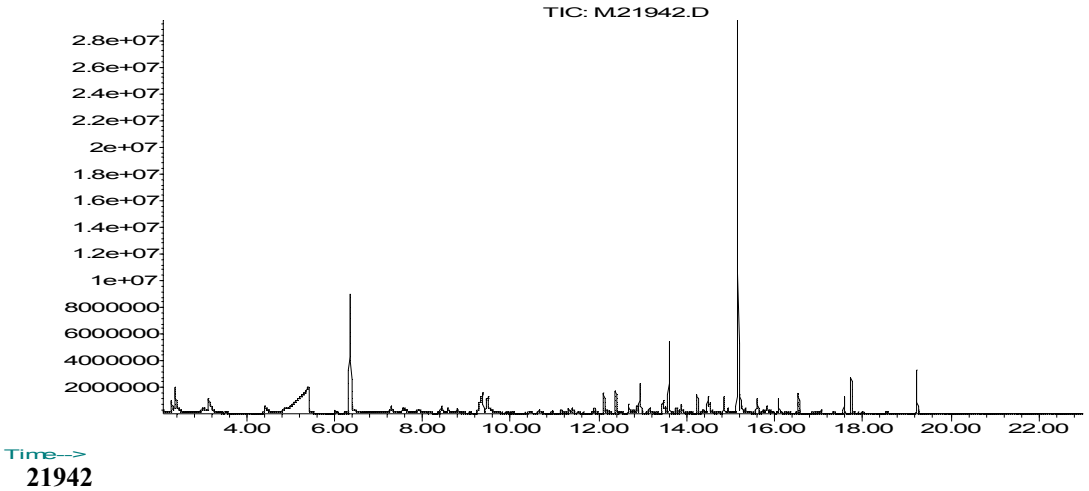

**21584**



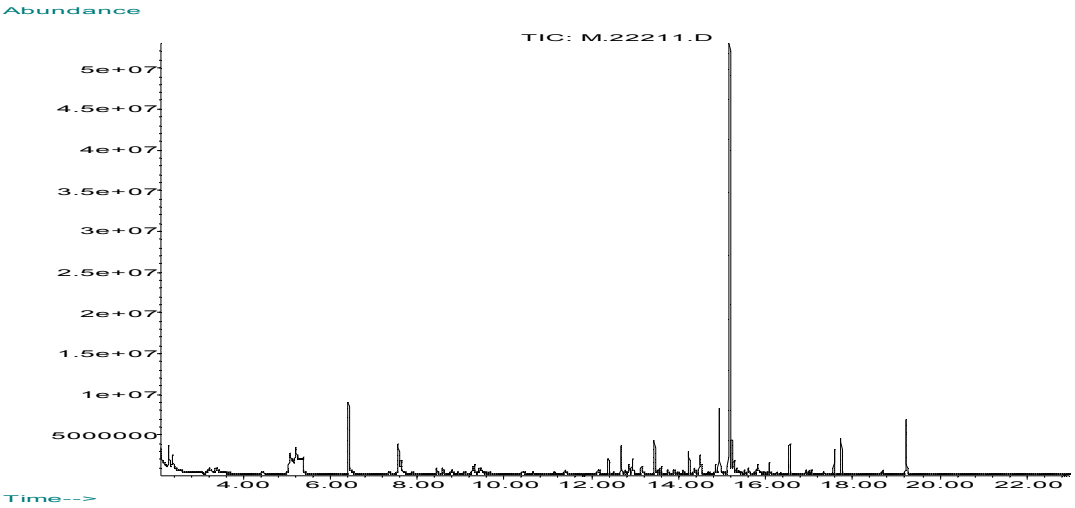

**21619**



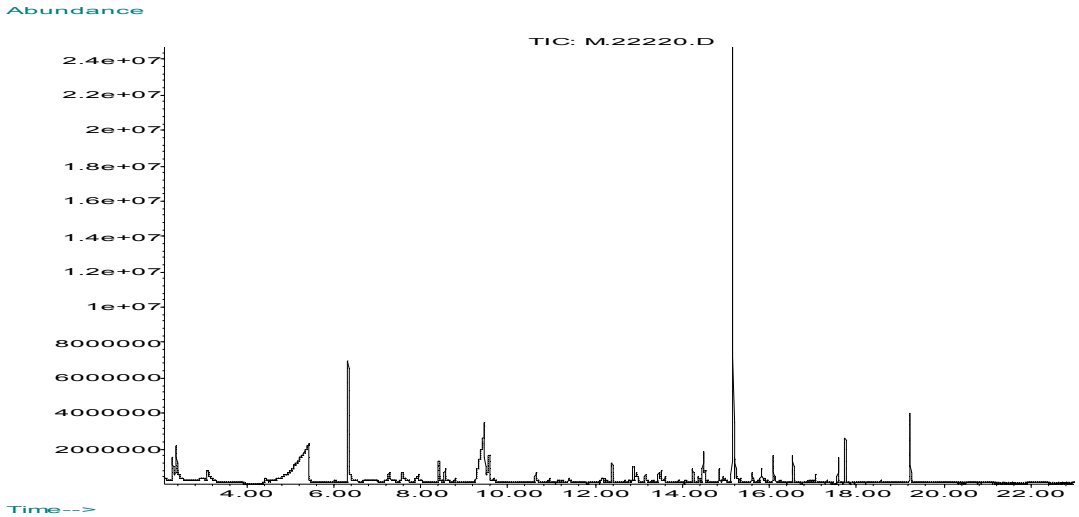

**21920**



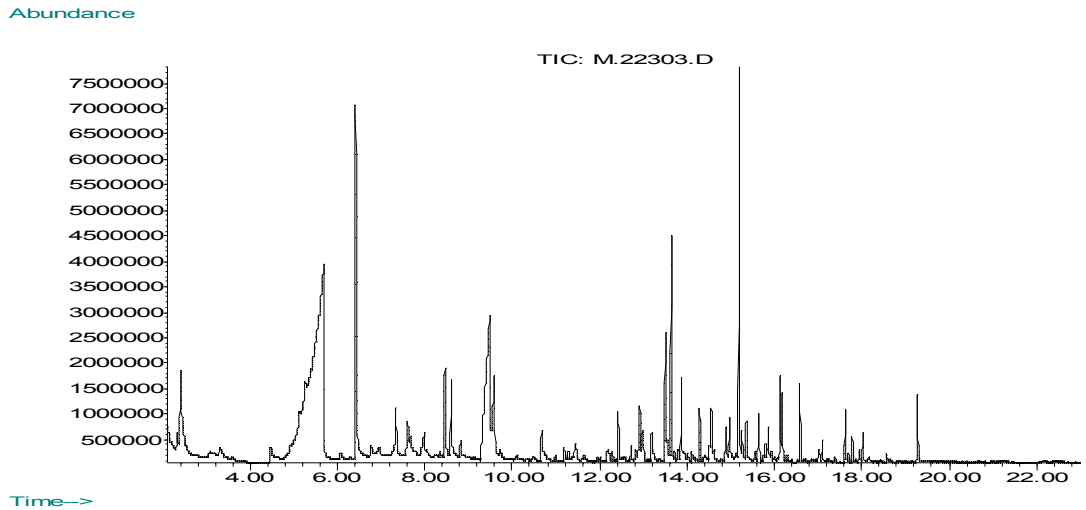

**21942**



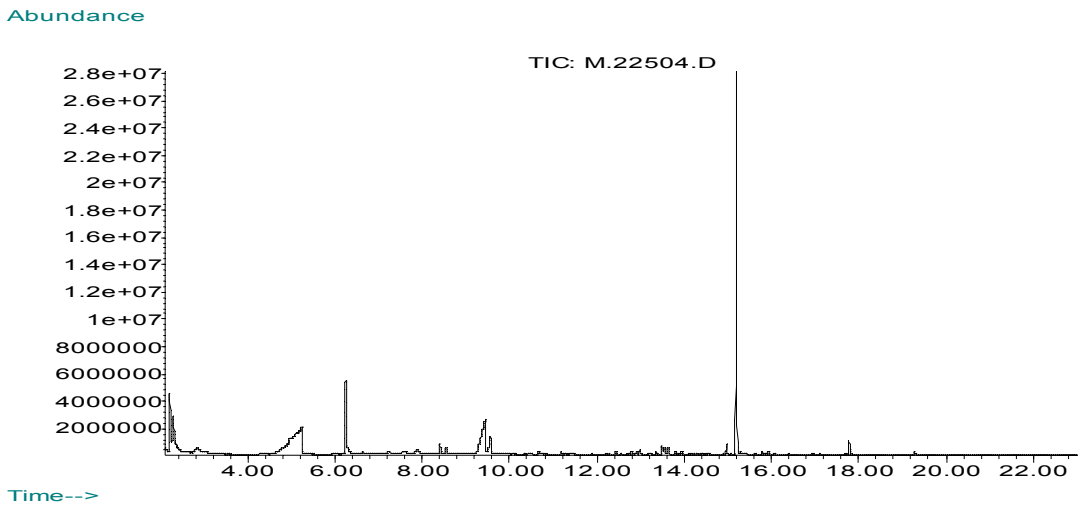

**22211**



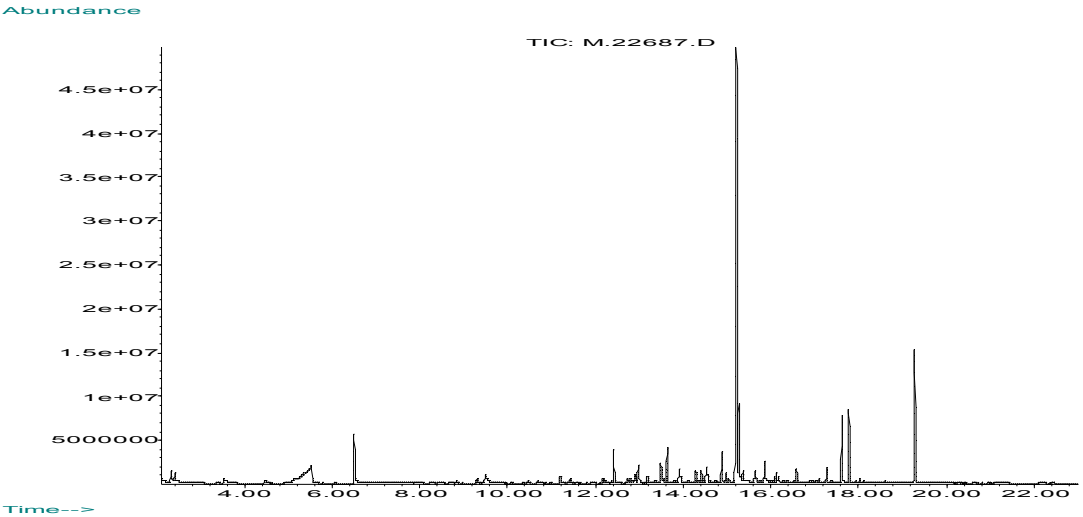

**22220**



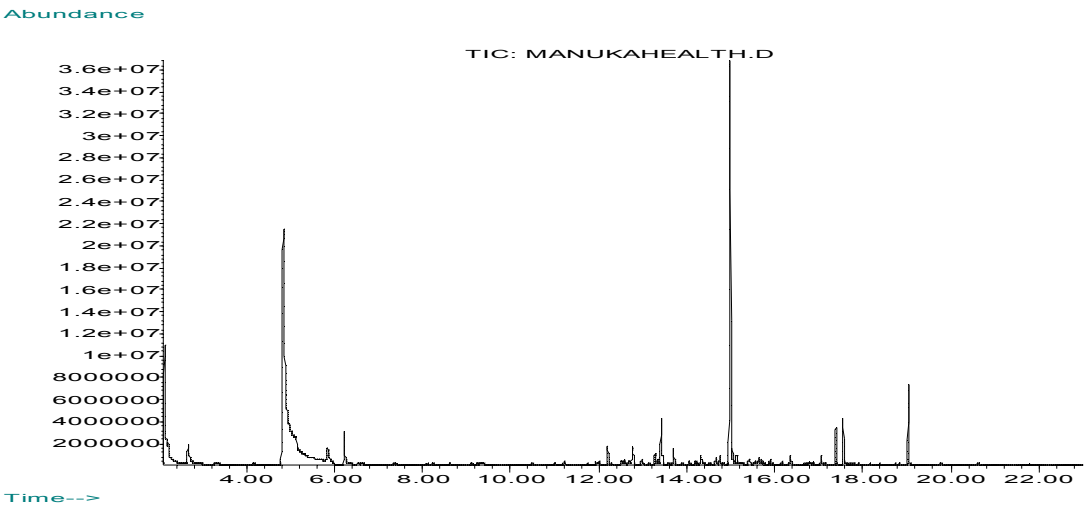

**22303**




**22504**

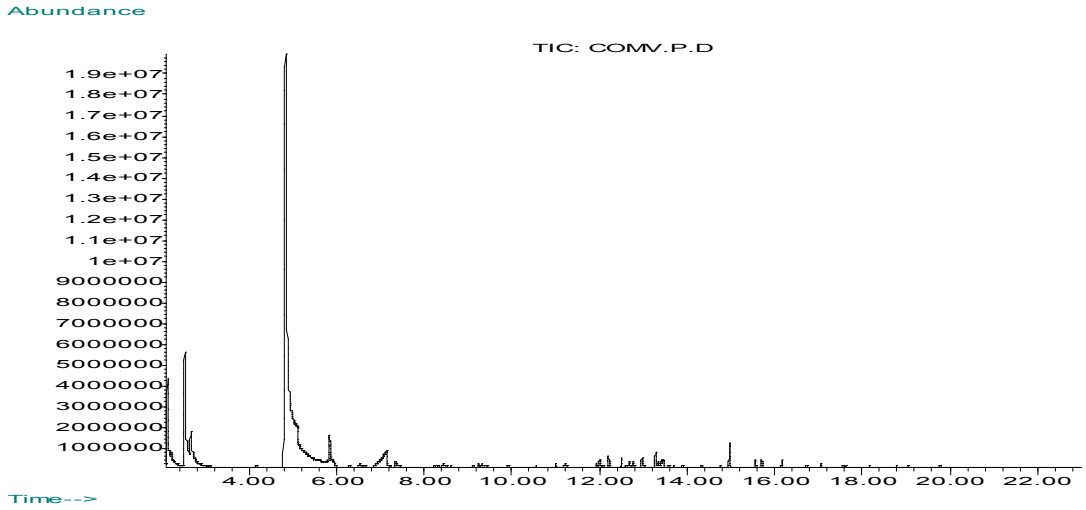



**22687**

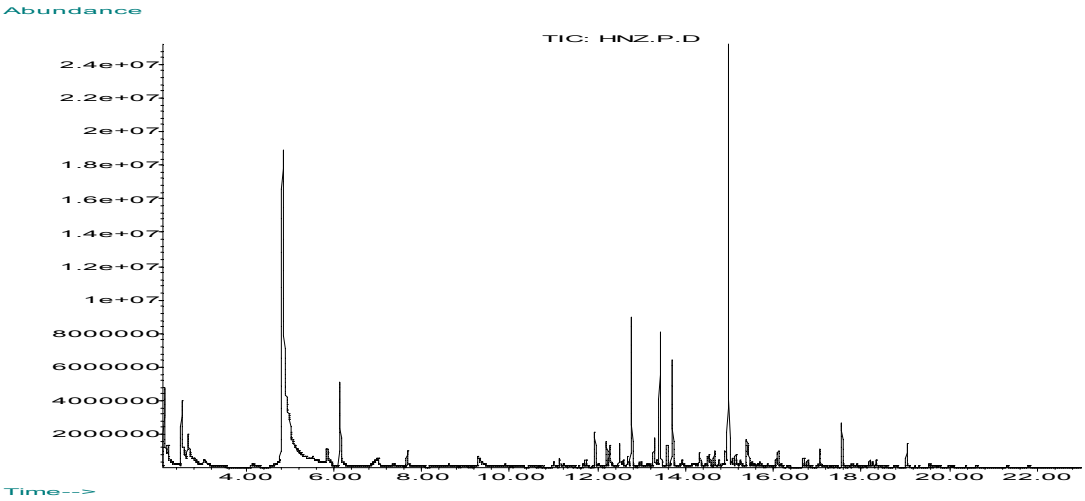



**Manuka Health**

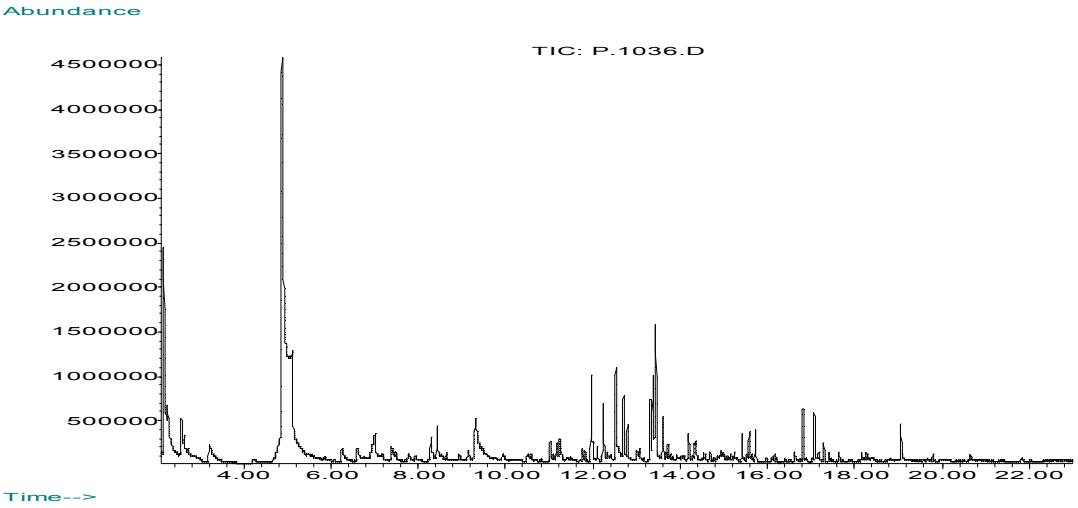
# Appendix 6: Chromatograms of Pohutukawa Honey Samples



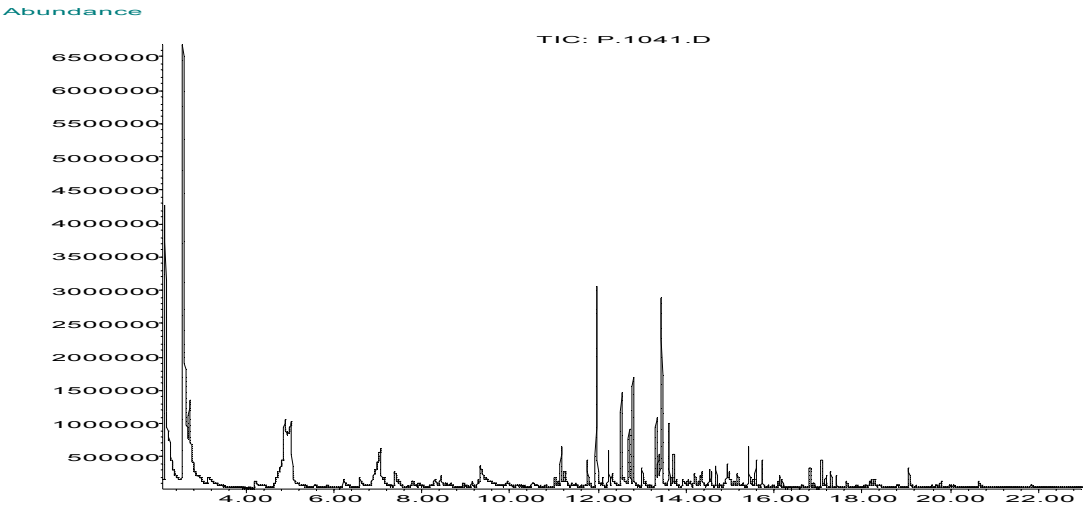

**Comv.P**



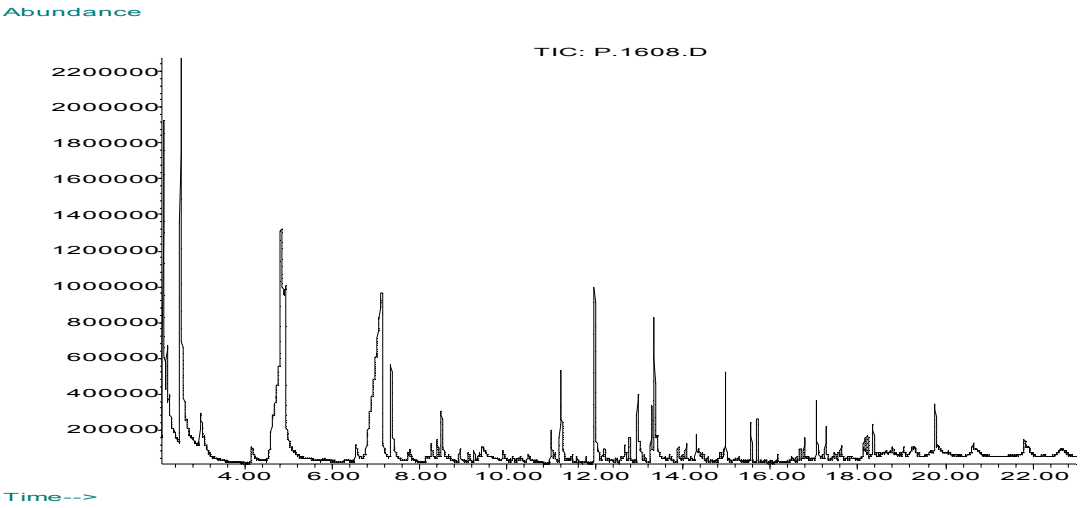

**HNZ.P**



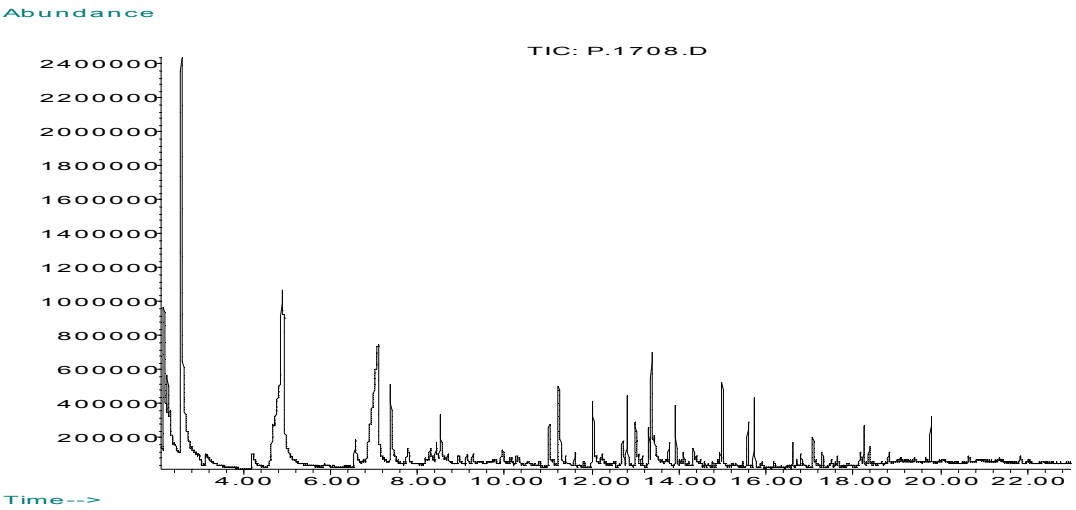

**1036**



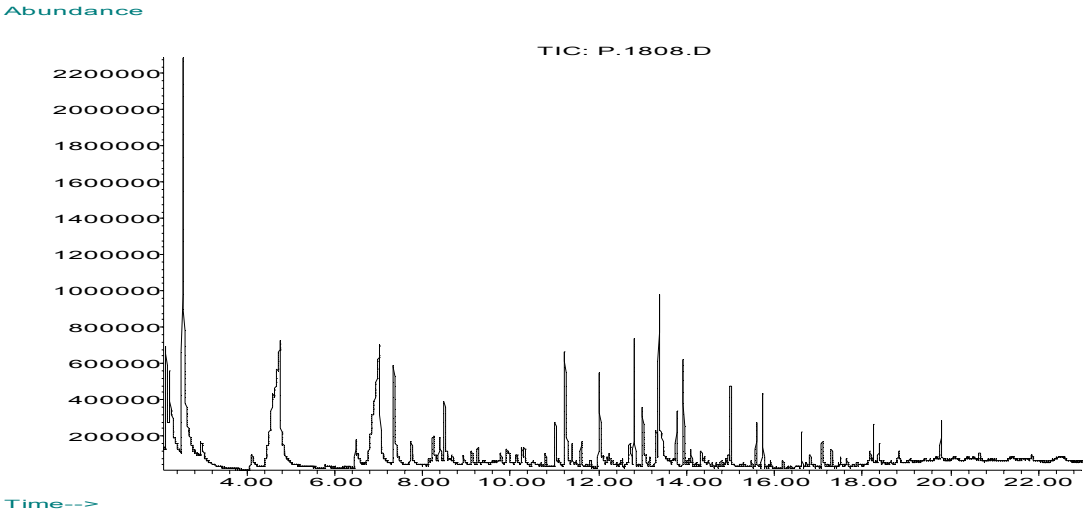

**1041**



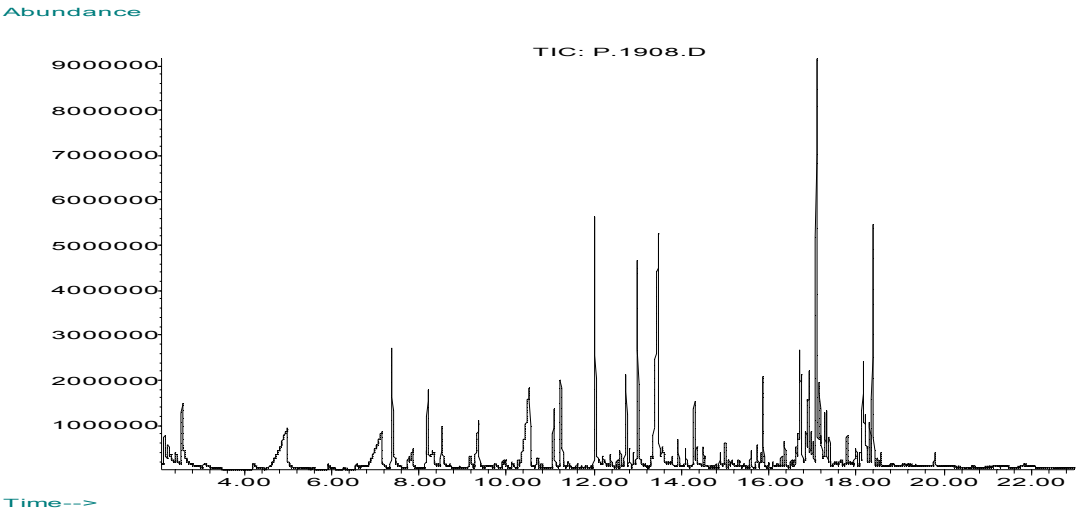

**1608**



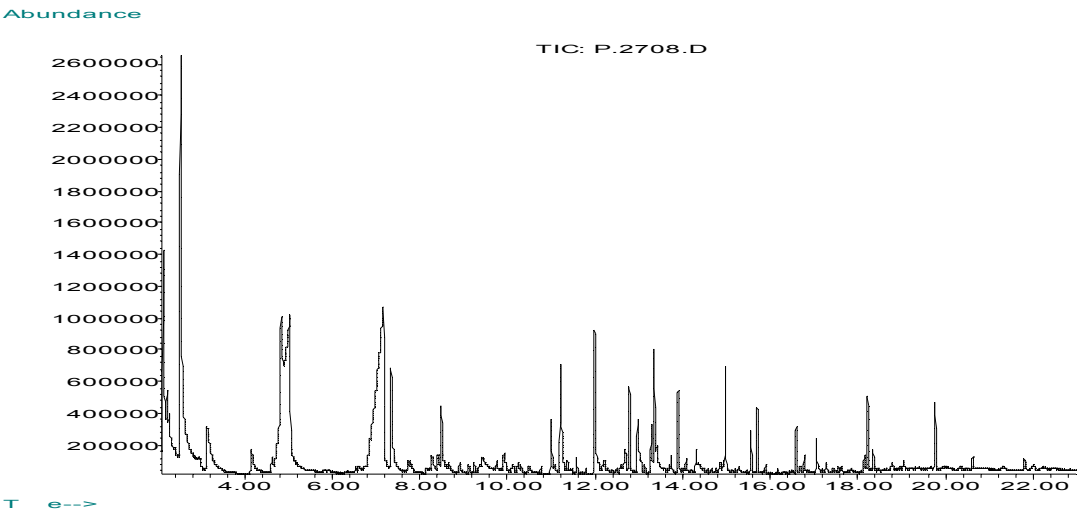

**1708**



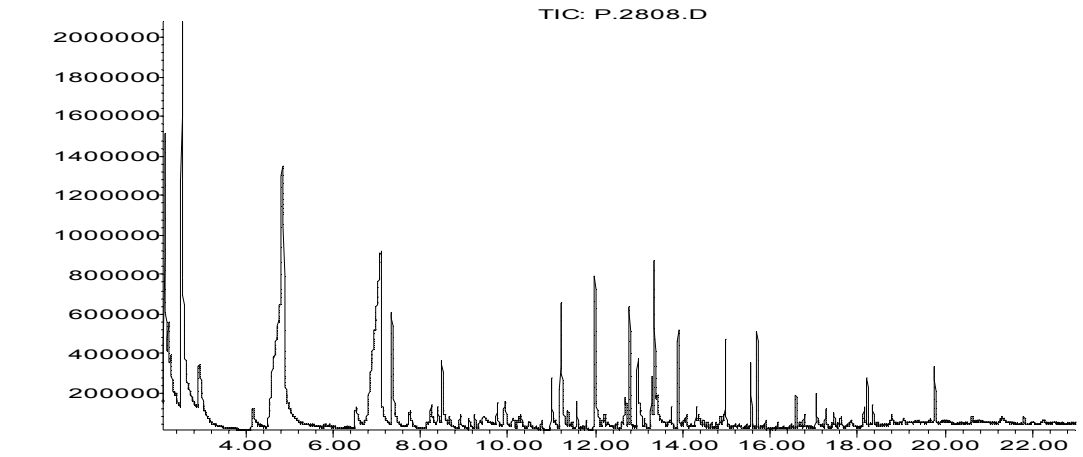

**1808**



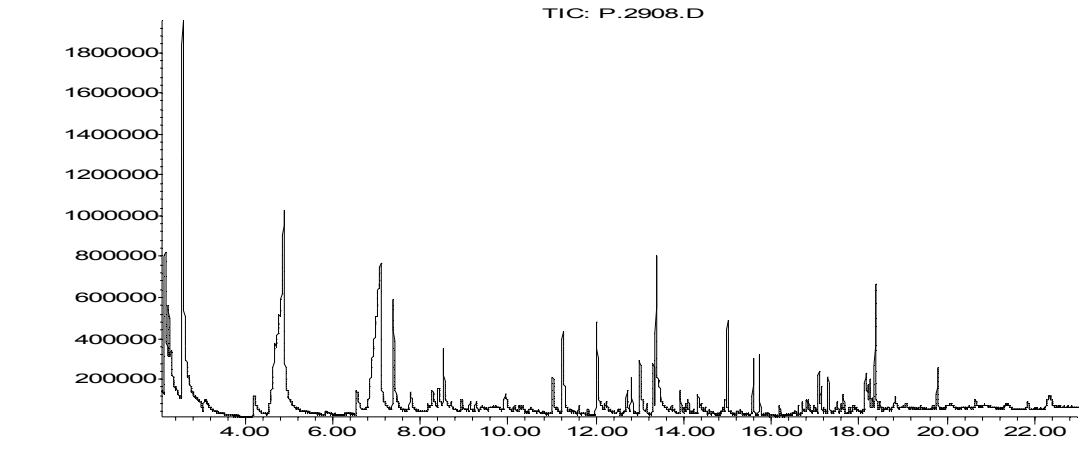

**1908**



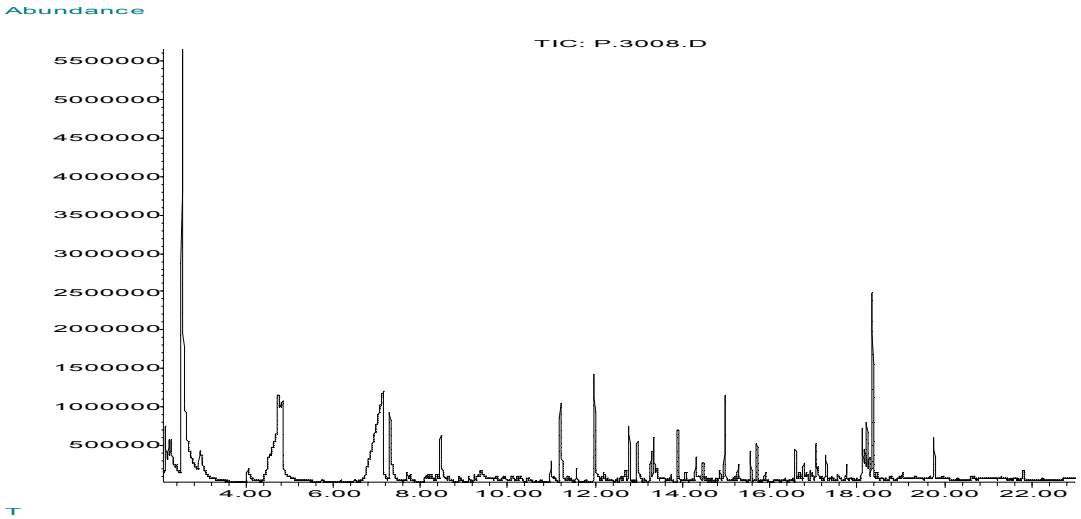

**2708**



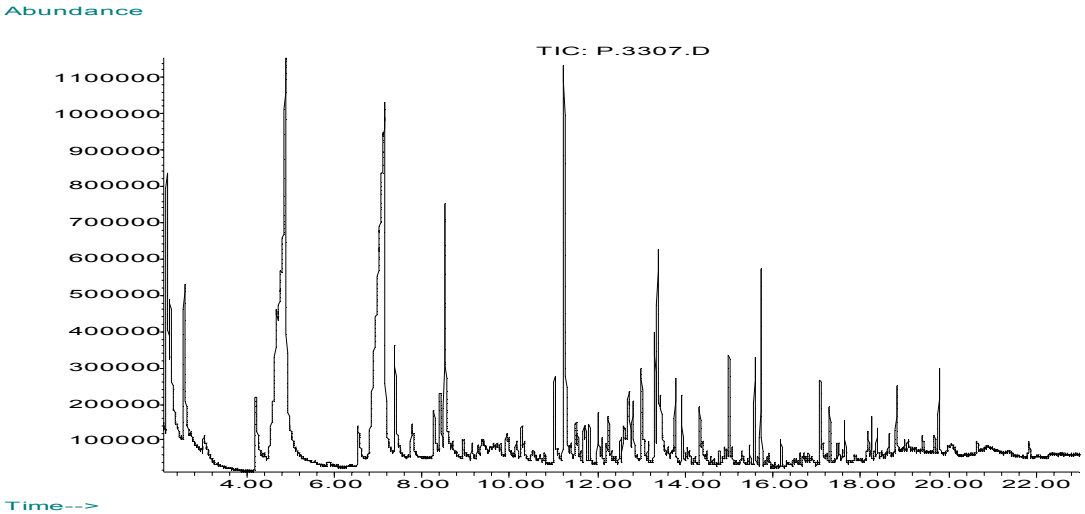

**2808**



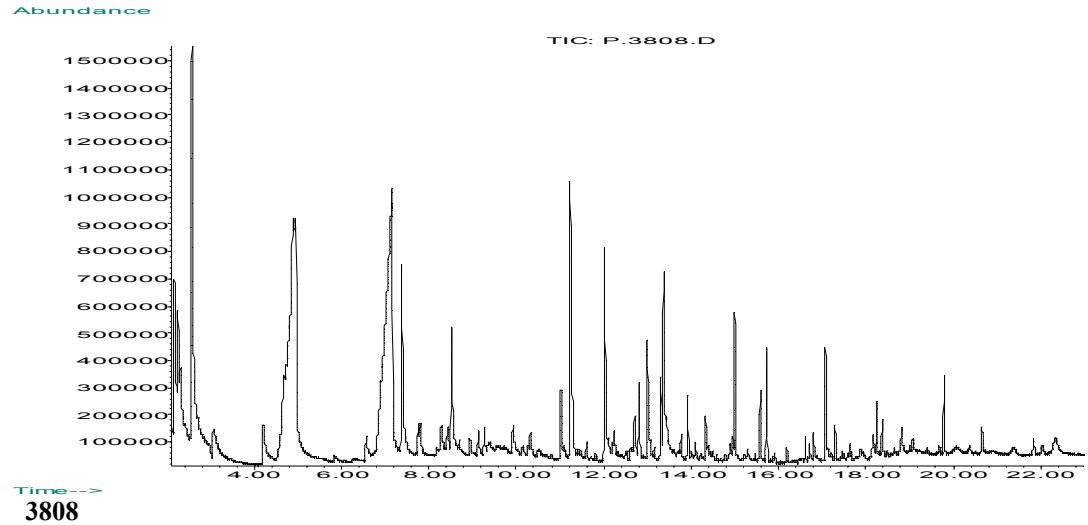

**2908**



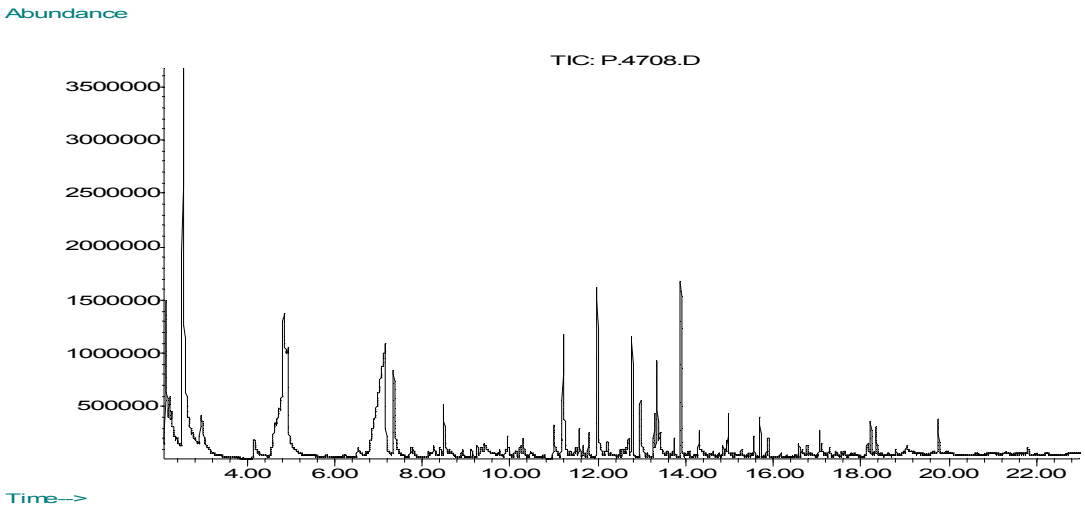

**3008**



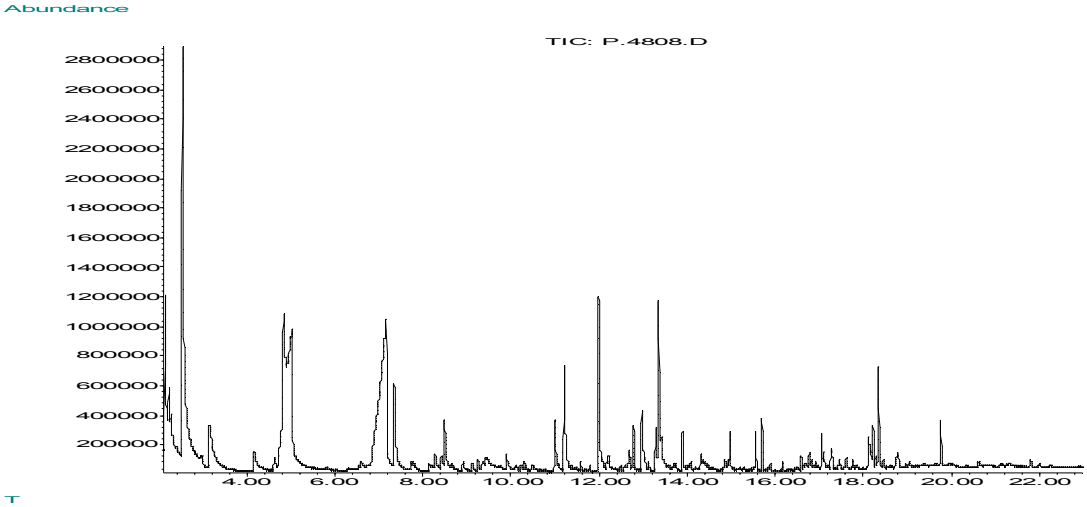

**3307**



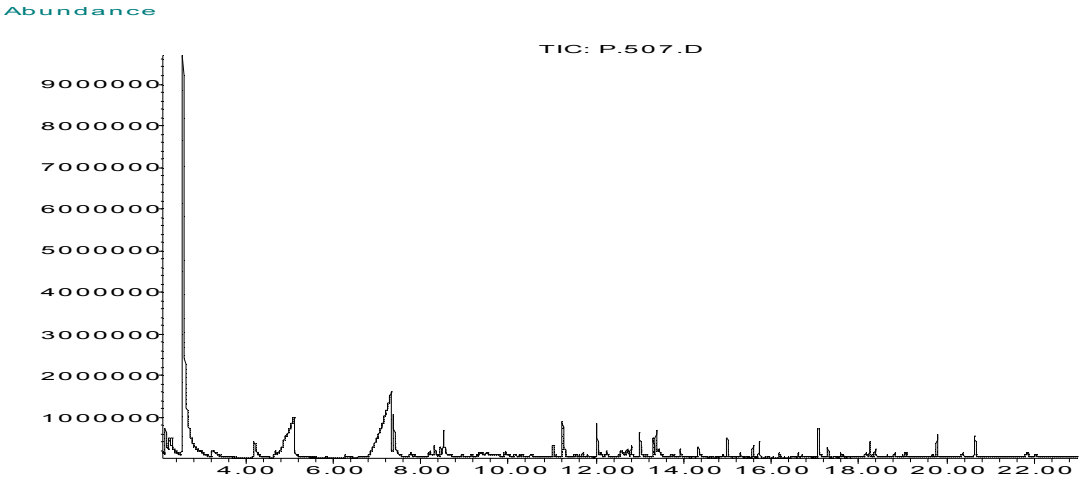

**3808**


Abundance
TIC: P.4708.D
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

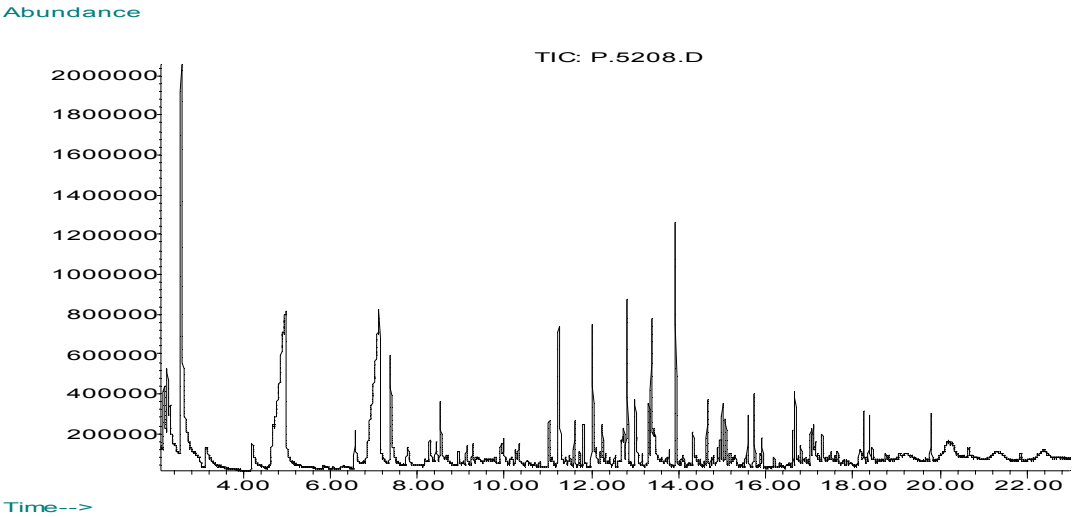
**4708**


Abundance
TIC: P.4808.D
2800000
2600000
2400000
2200000
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
T

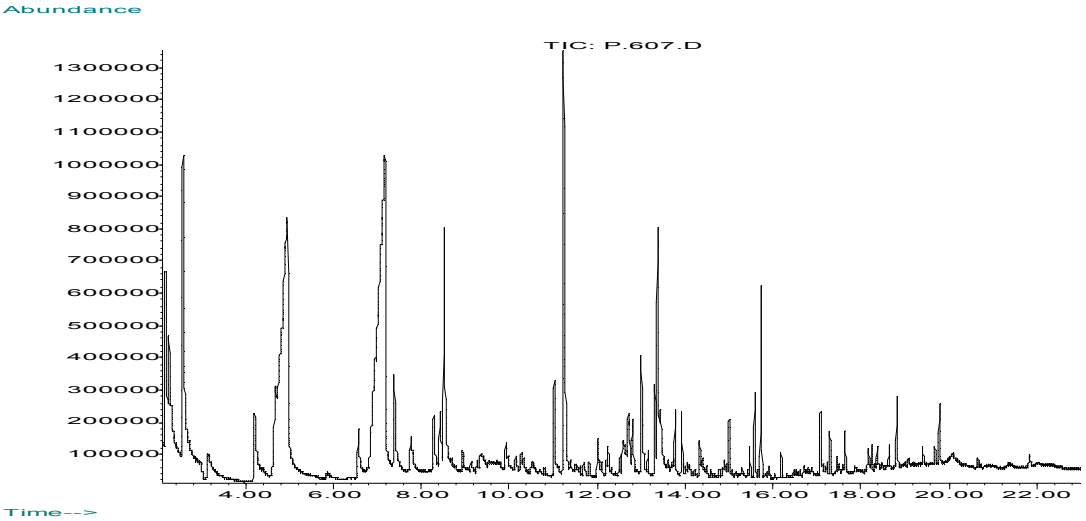
**4808**


Abundance
TIC: P.507.D
9000000
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**507**

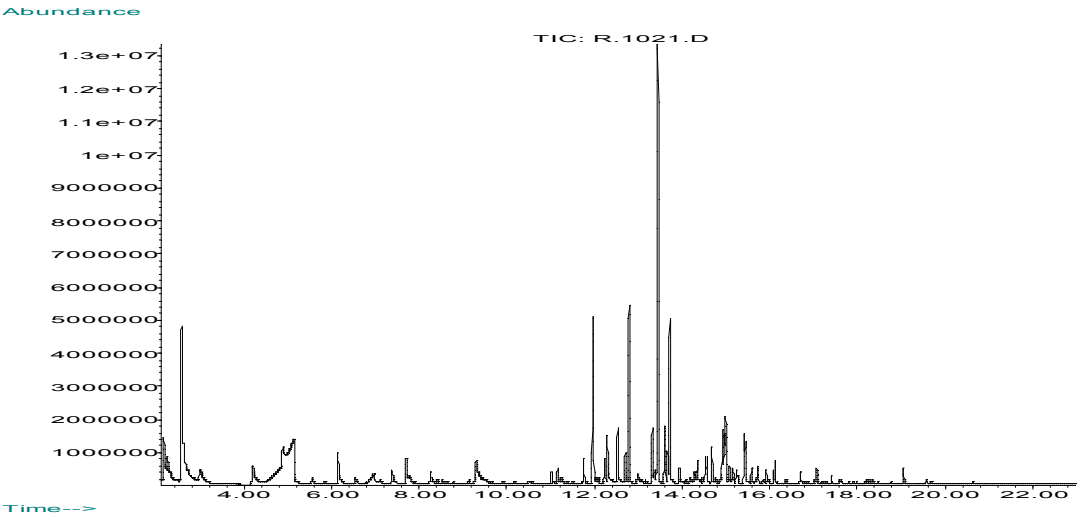

Abundance
TIC: P.5208.D
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**5208**

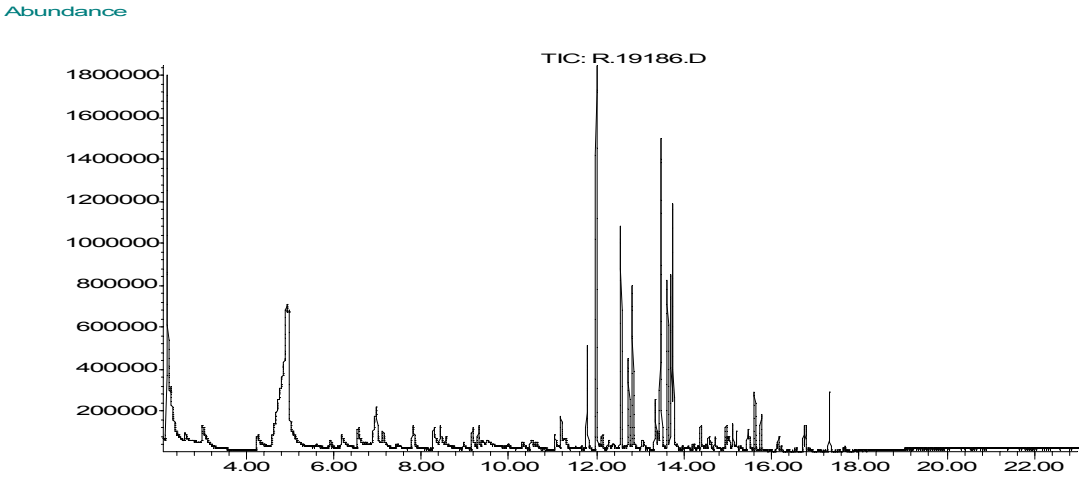

Abundance
TIC: P.607.D
1300000
1200000
1100000
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**607**

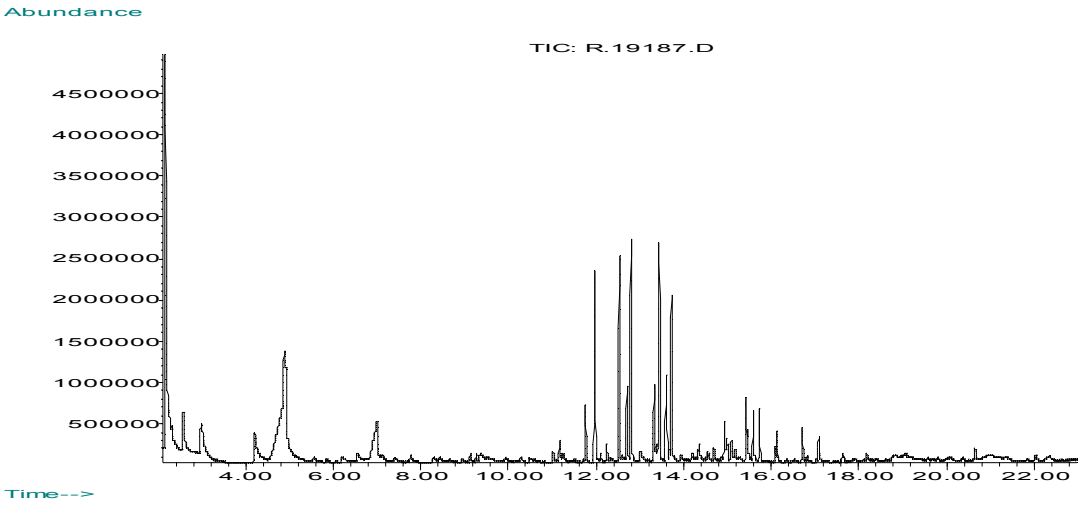
# Appendix 7: Chromatograms of Rata Honey Samples



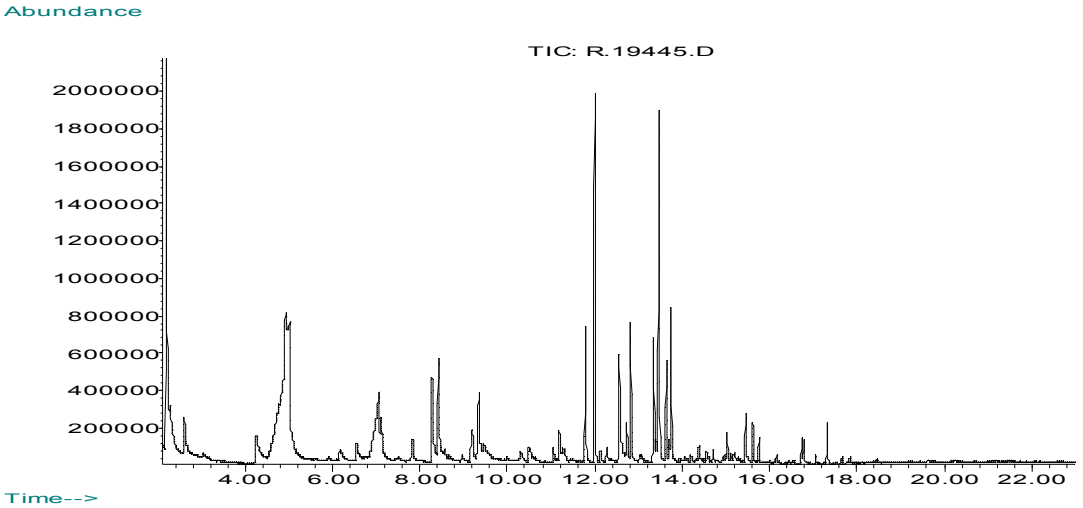

**1021**



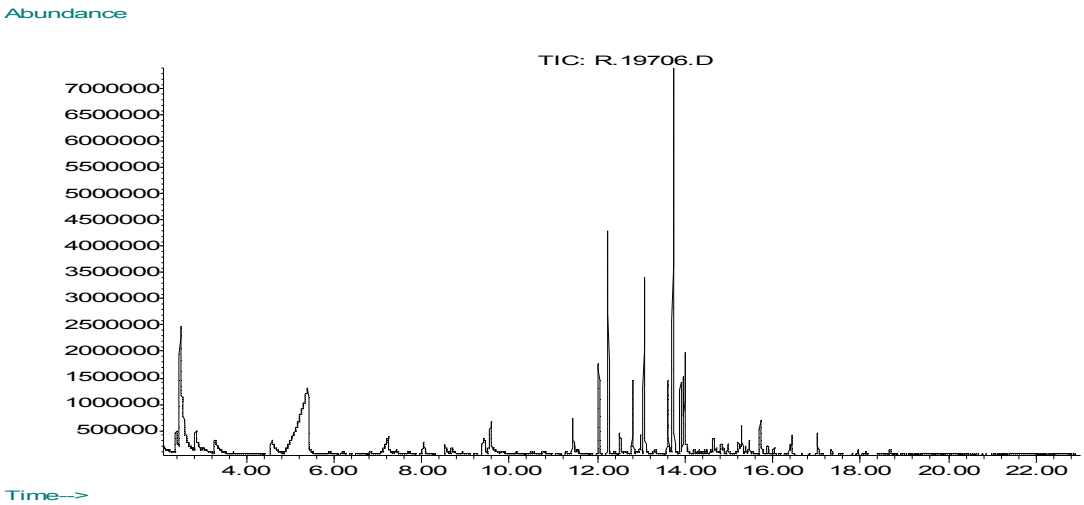

**19186**



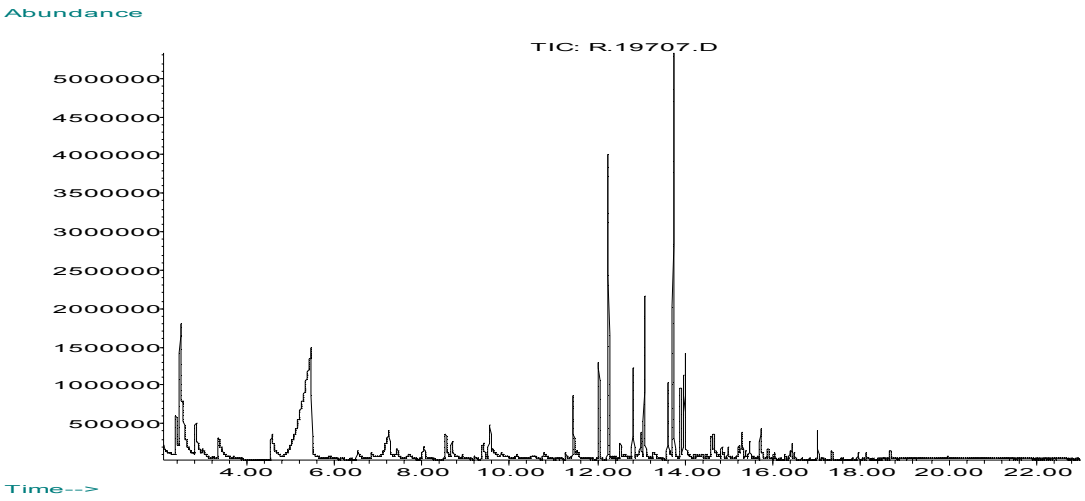

**19187**



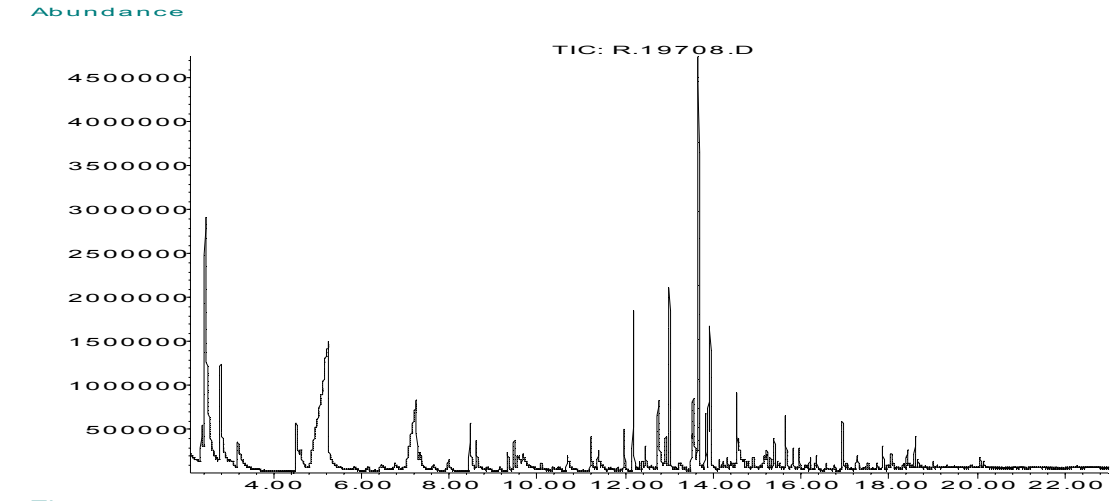

**19445**



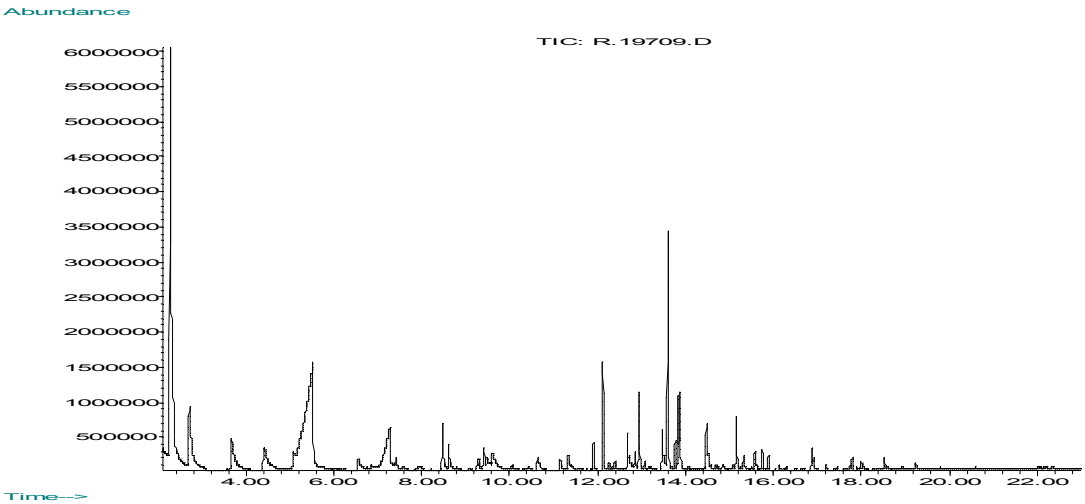

**19706**



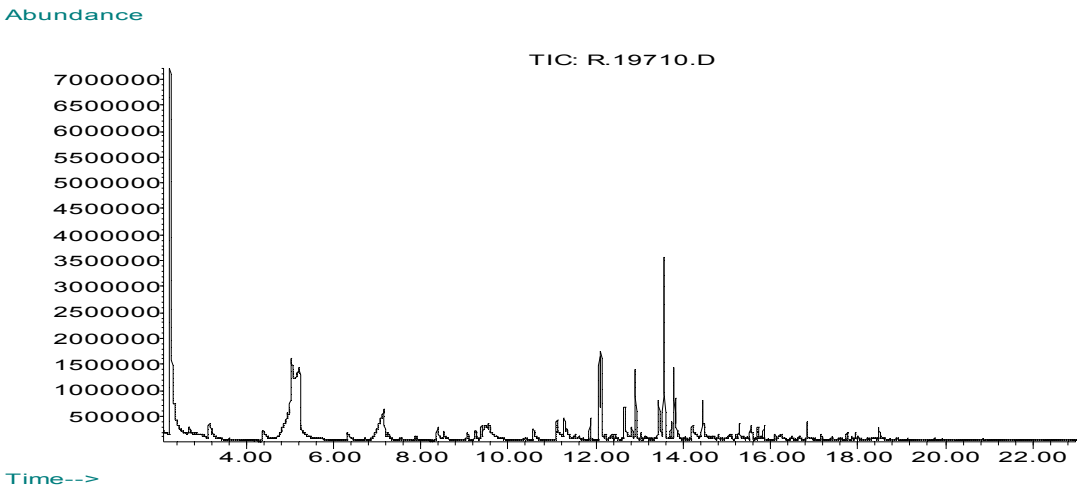

**19707**



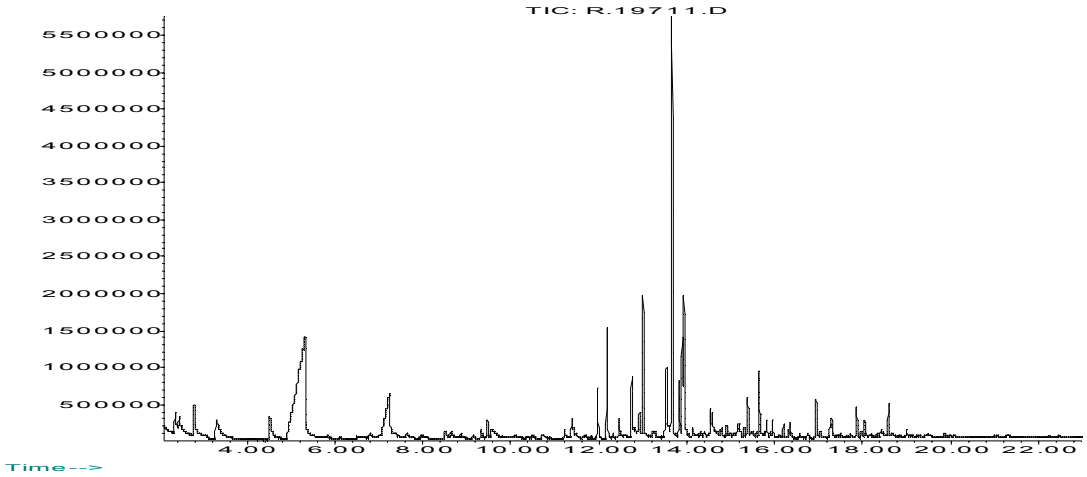

**19708**



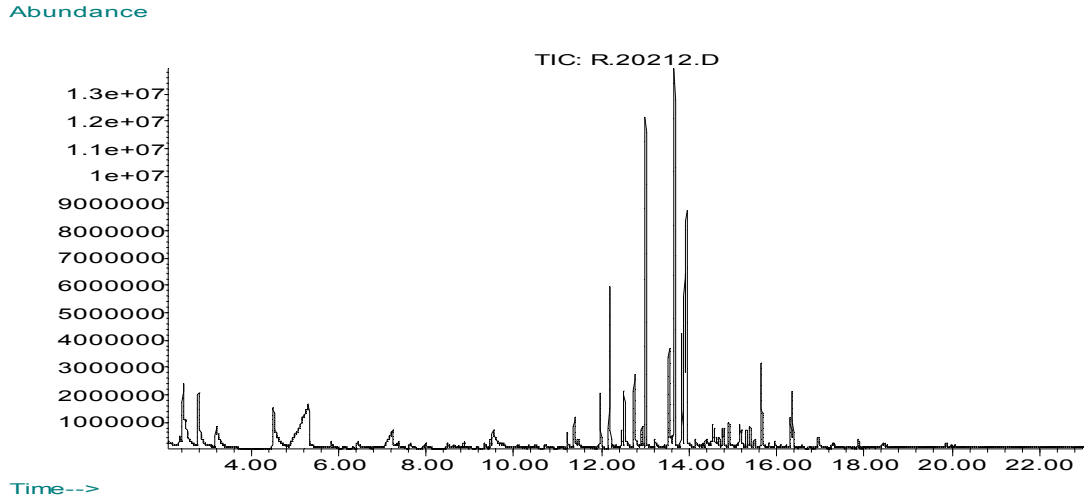

**19709**



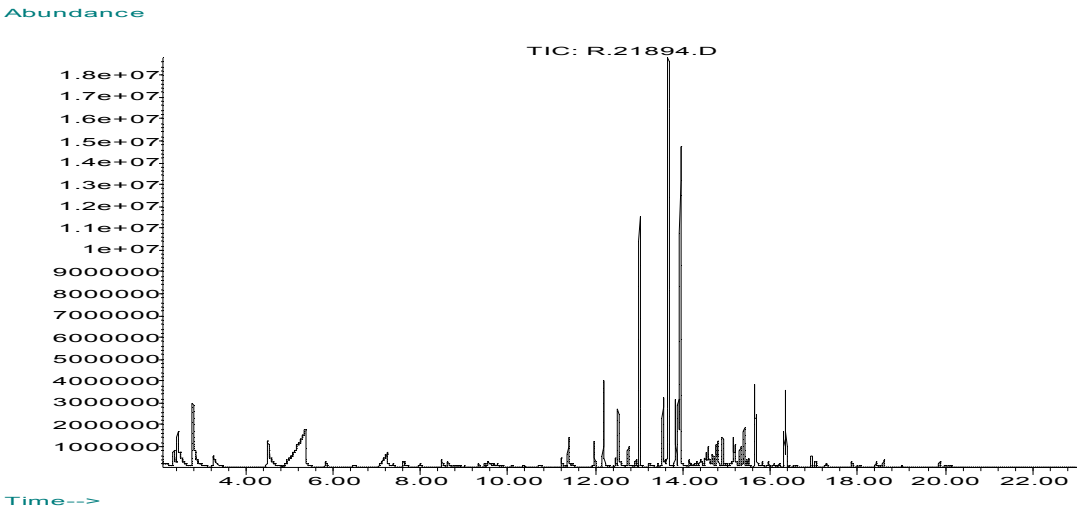

**19710**



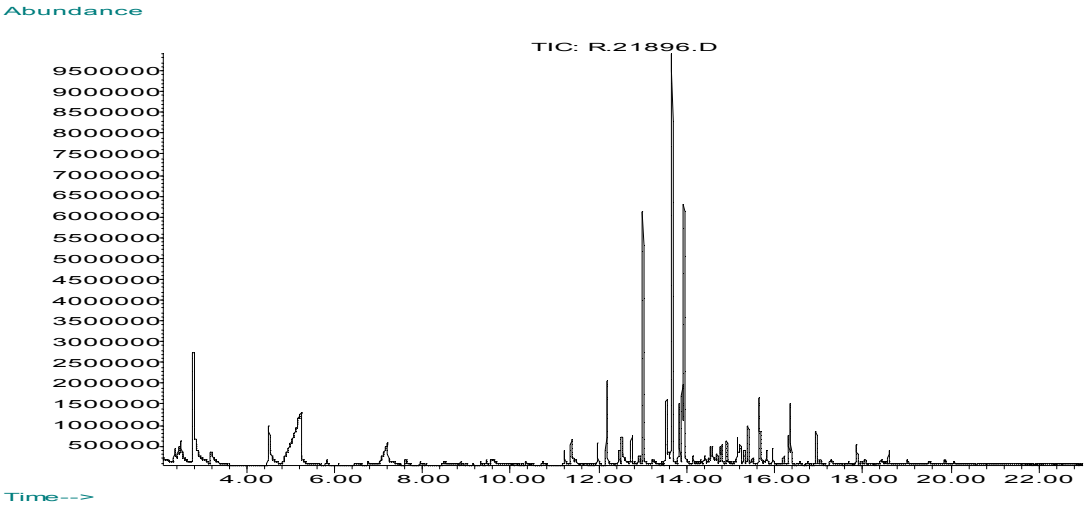

**19711**



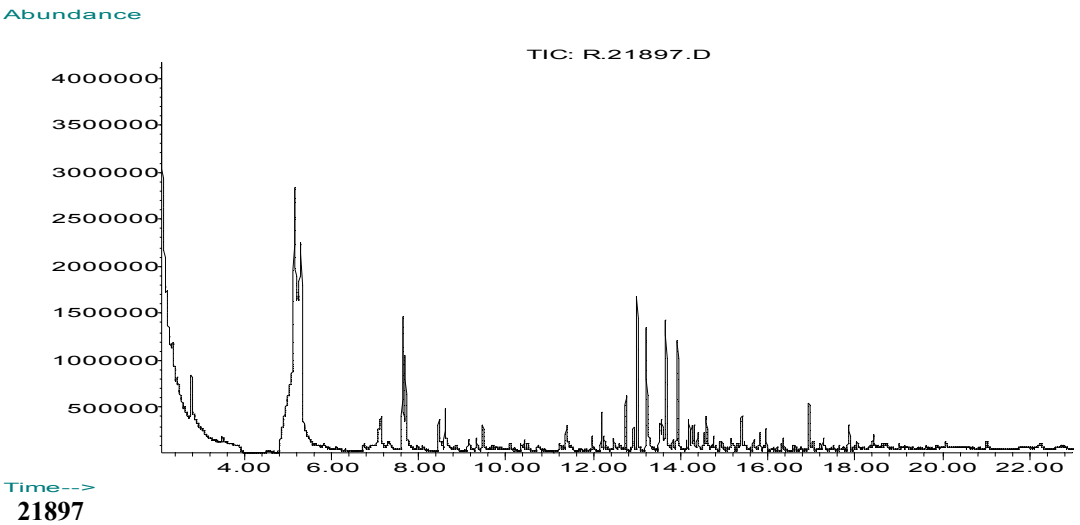

**20212**



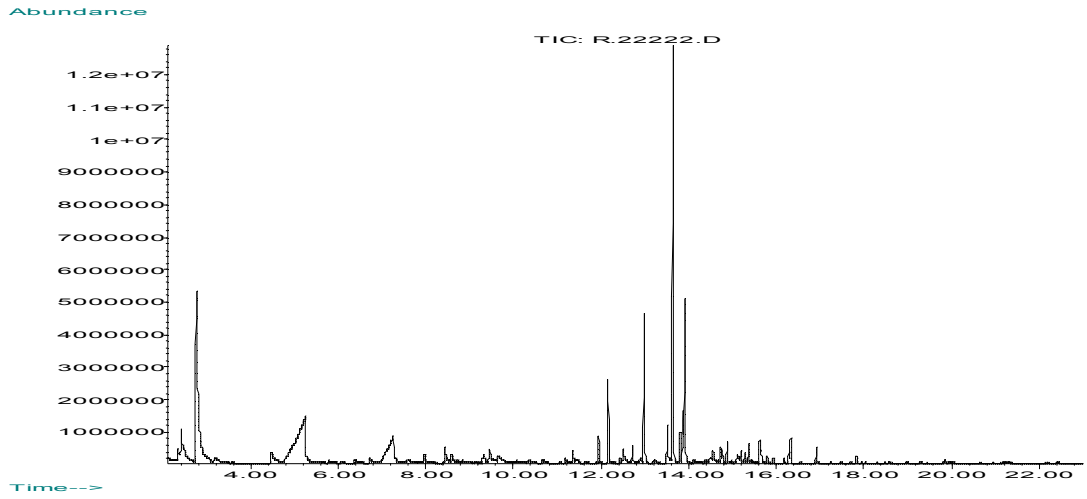

**21894**



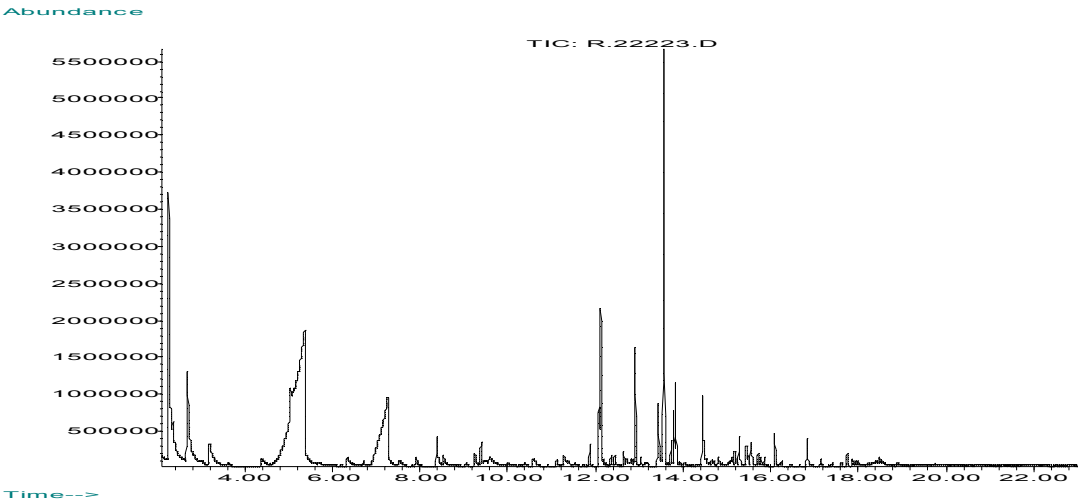

**21896**



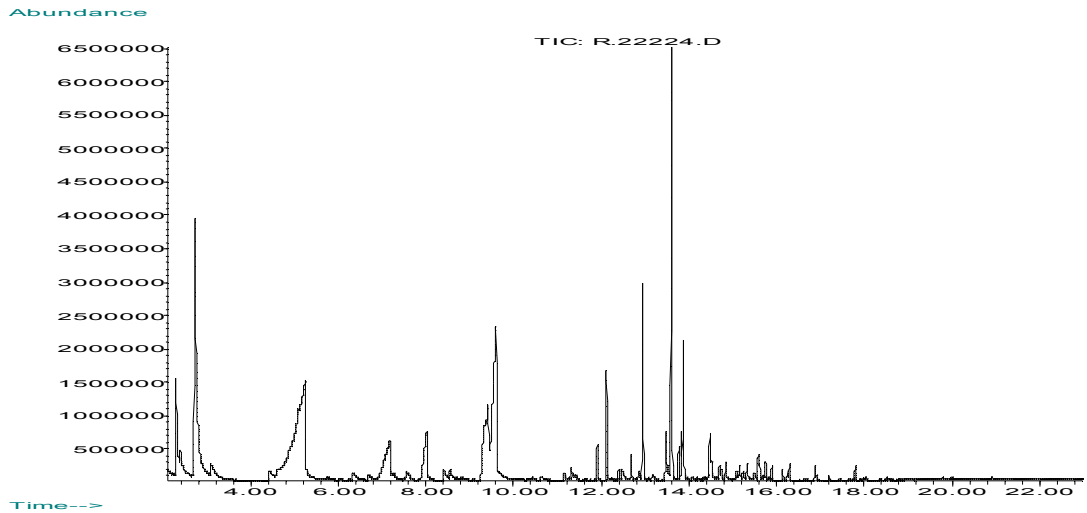

**21897**



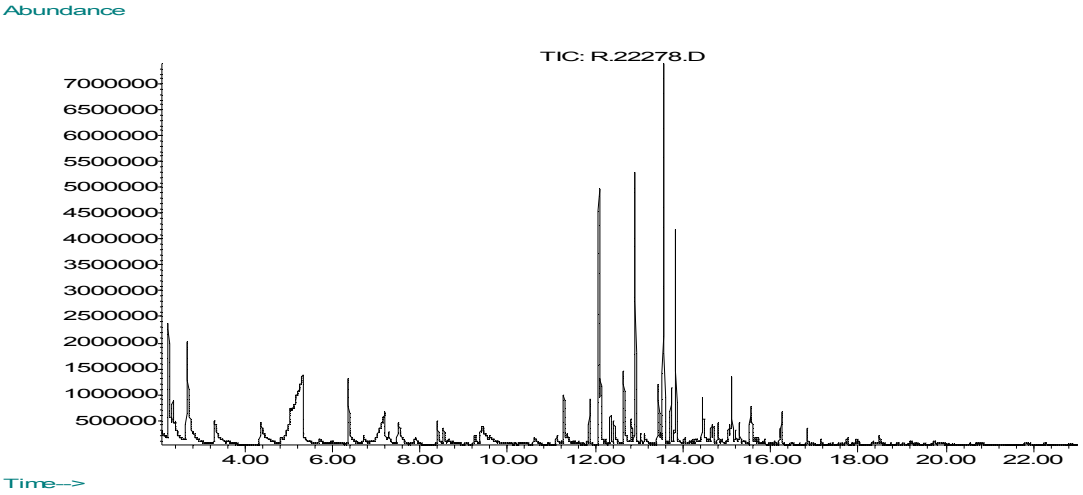

**22222**



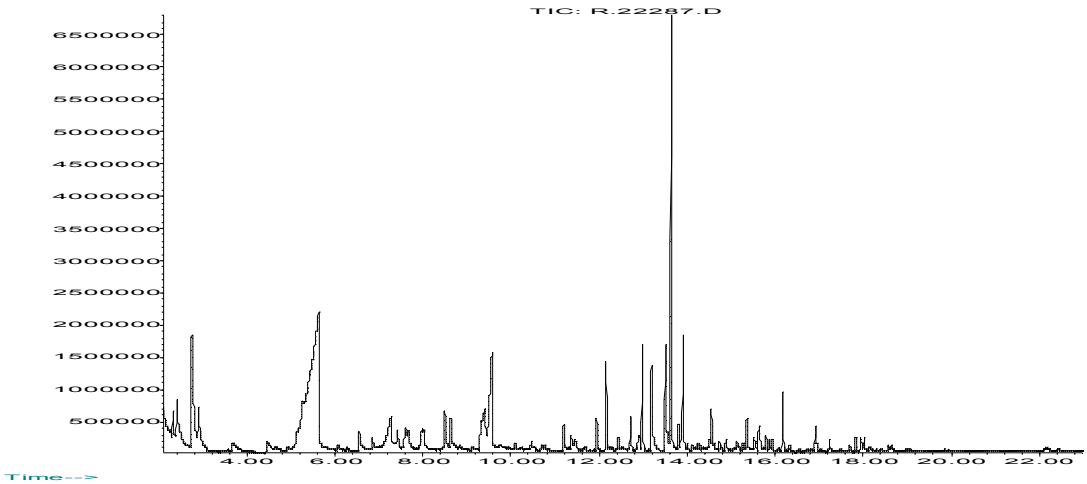

**22223**



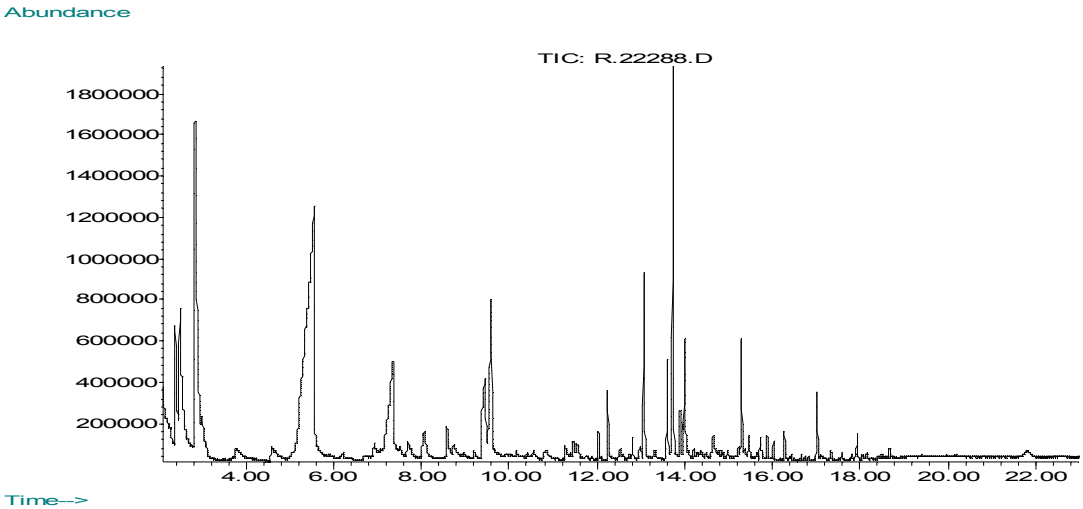

**22224**



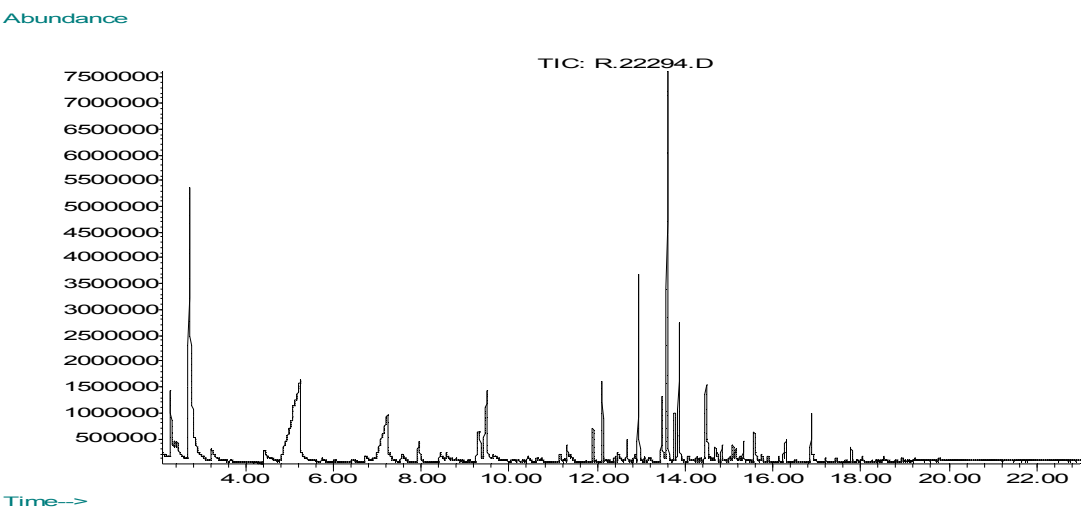

**22278**



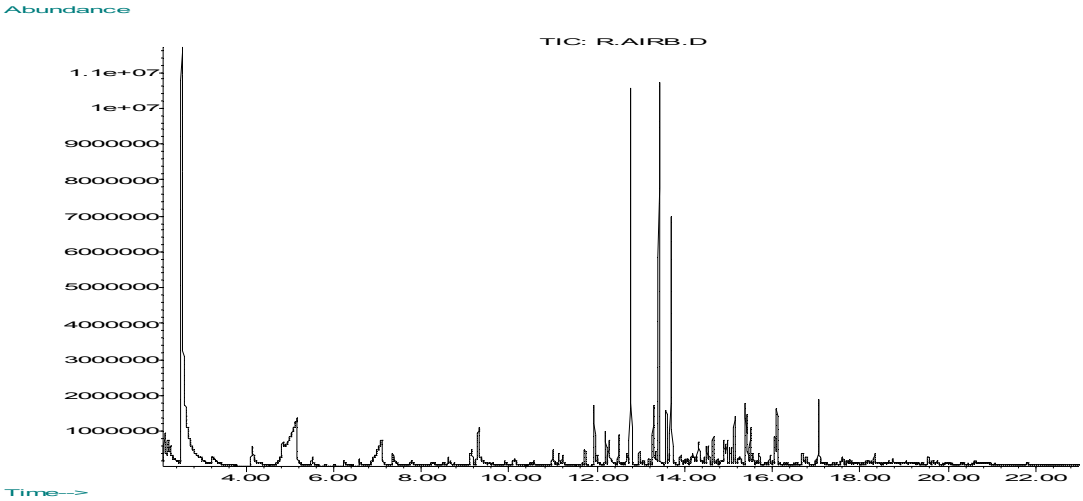

**22287**




**22288**

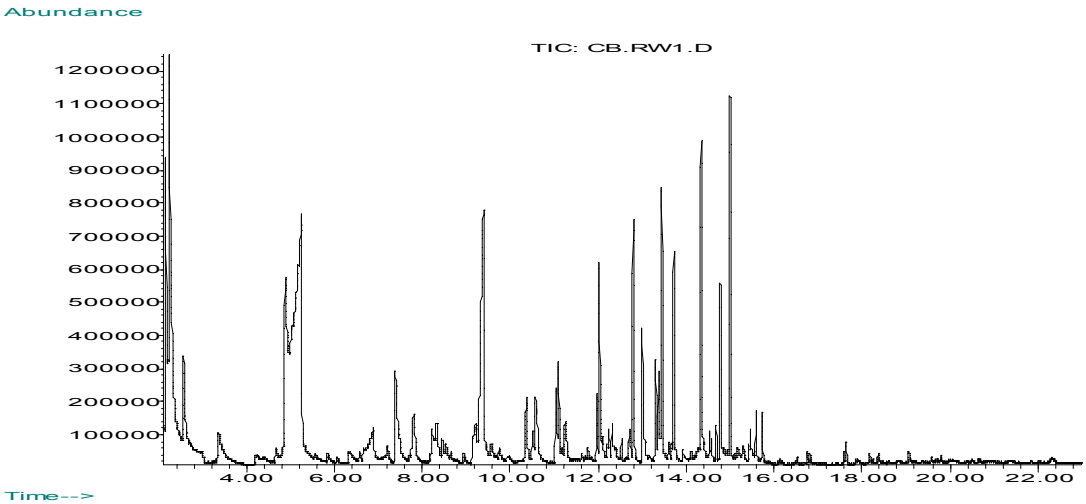

Abundance
TIC: R.22294.D
7500000
7000000
6500000
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**22294**

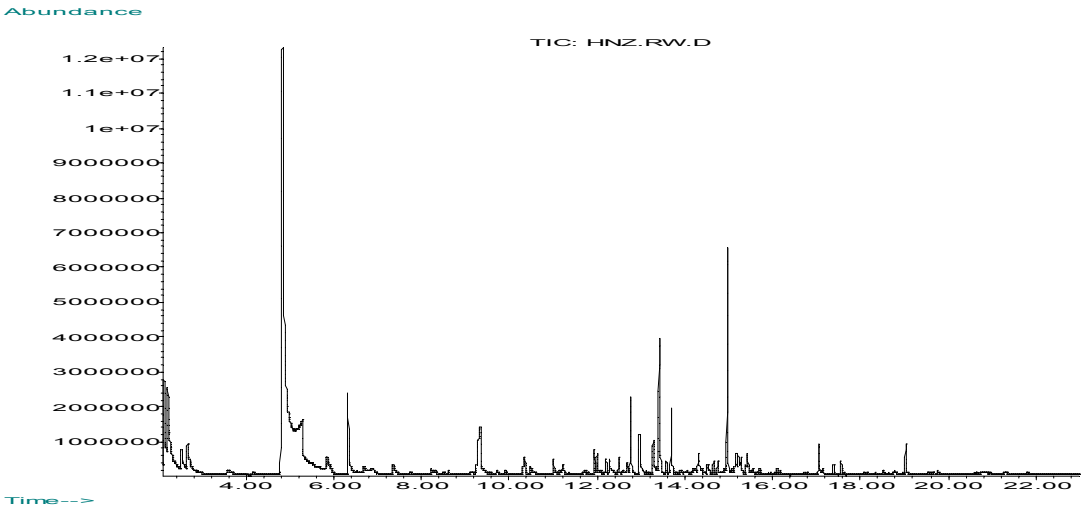

Abundance
TIC: R.AIRB.D
1.1e+07
1e+07
9000000
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**R.Airb.**

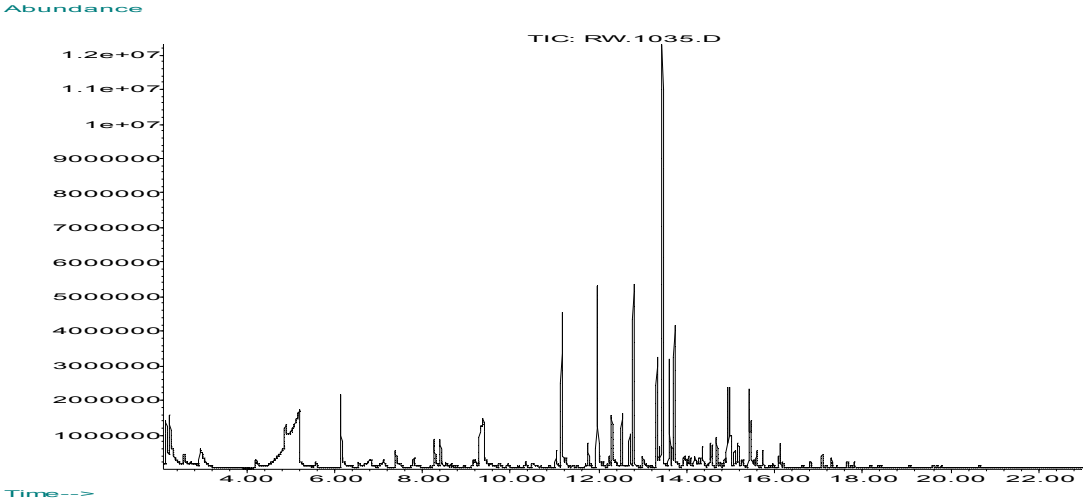
# Appendix 8: Chromatograms of Rewarewa Honey Samples



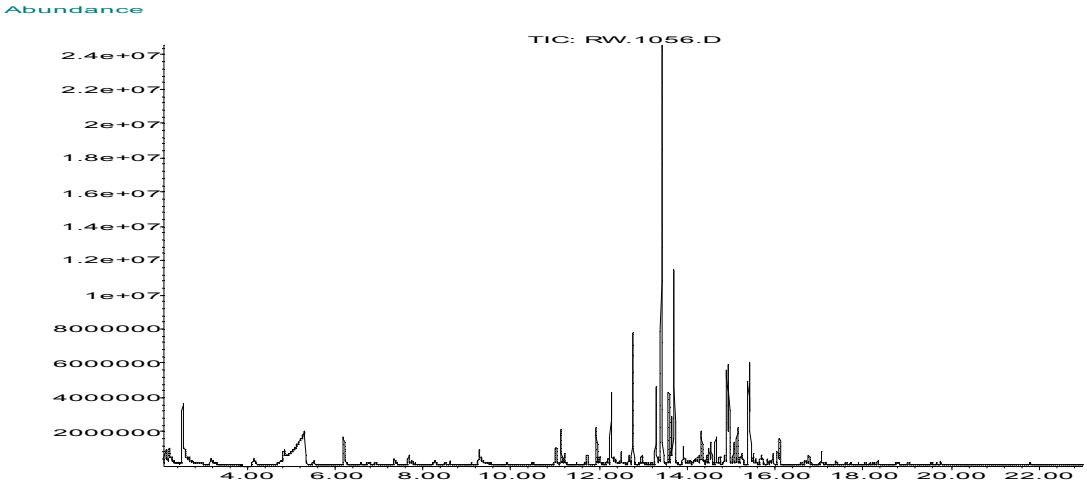

**CB.RW.1**



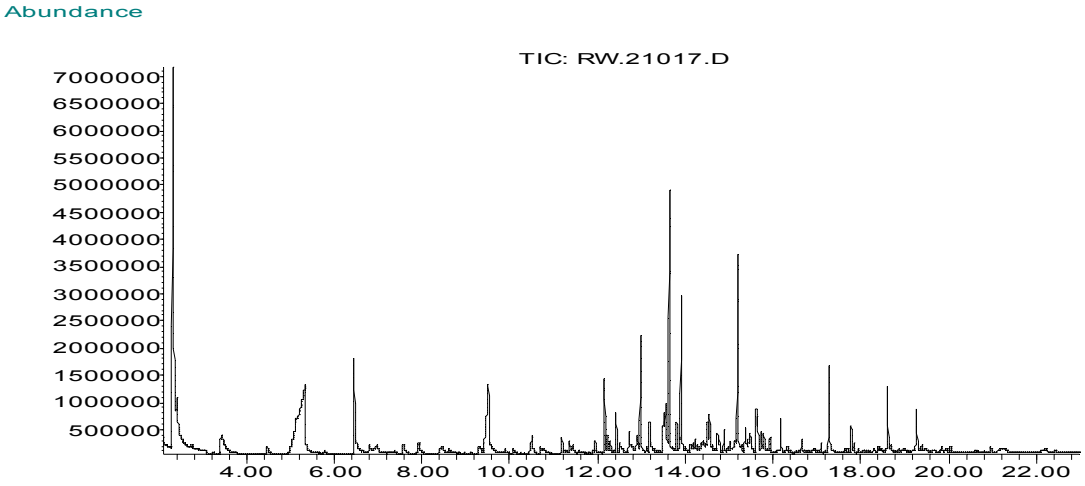

**HNZ.RW**



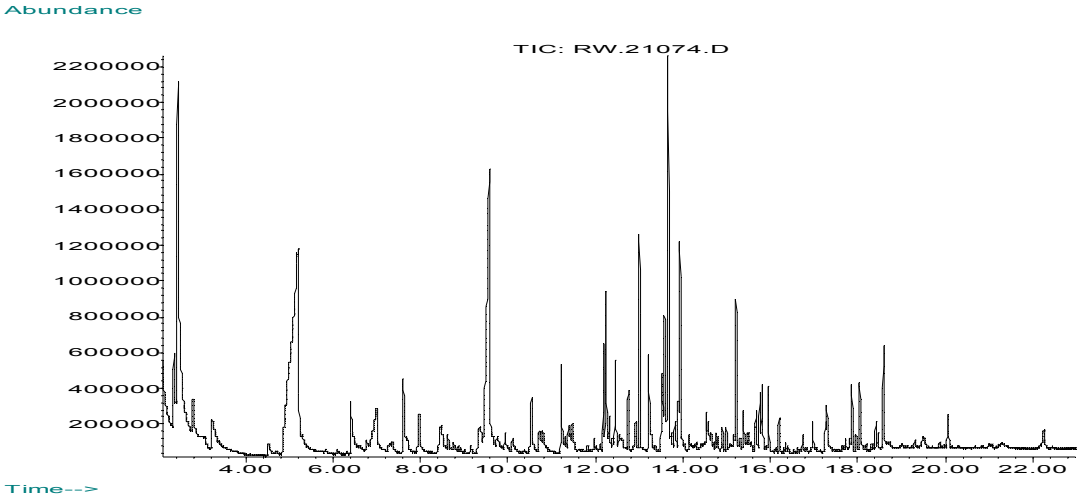

**1035**



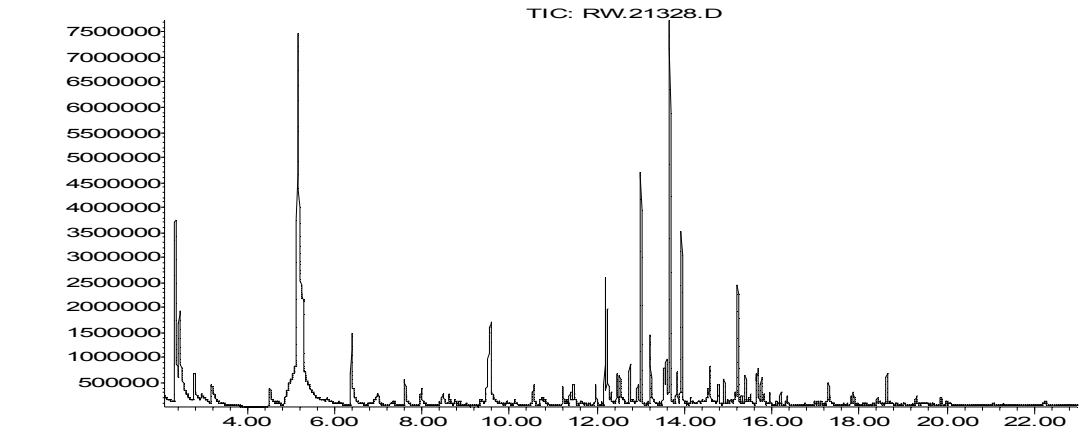

**1056**



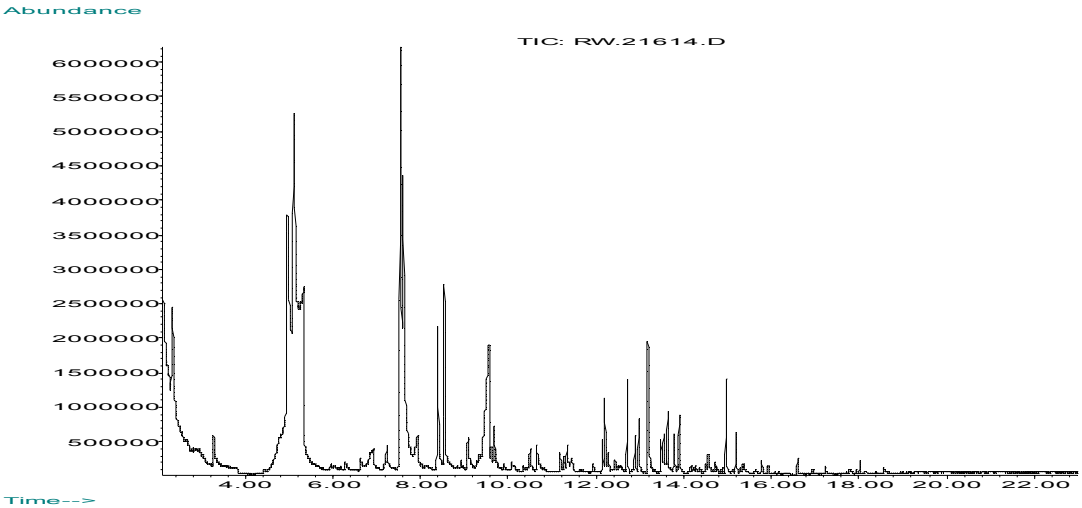

**21017**



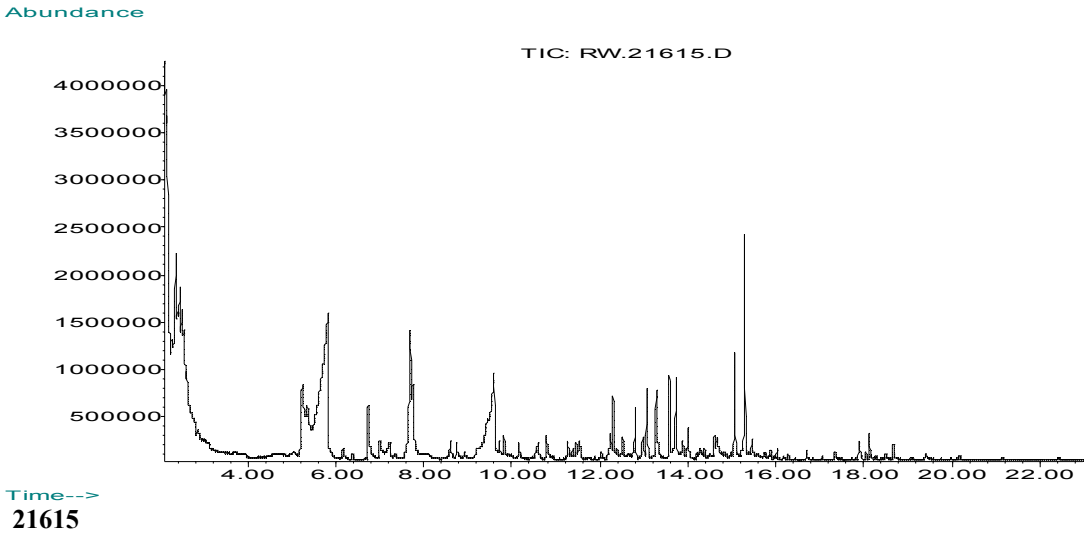

**21074**



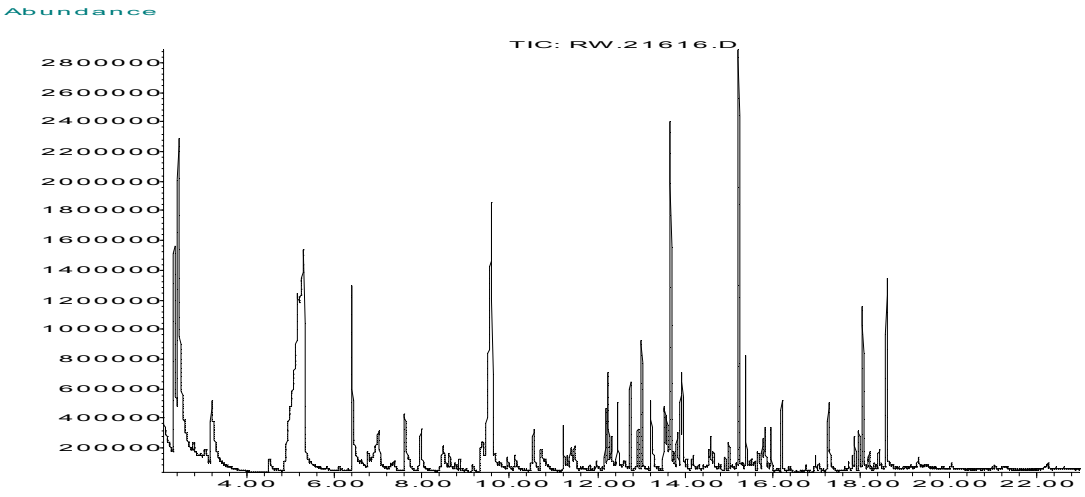

**21328**



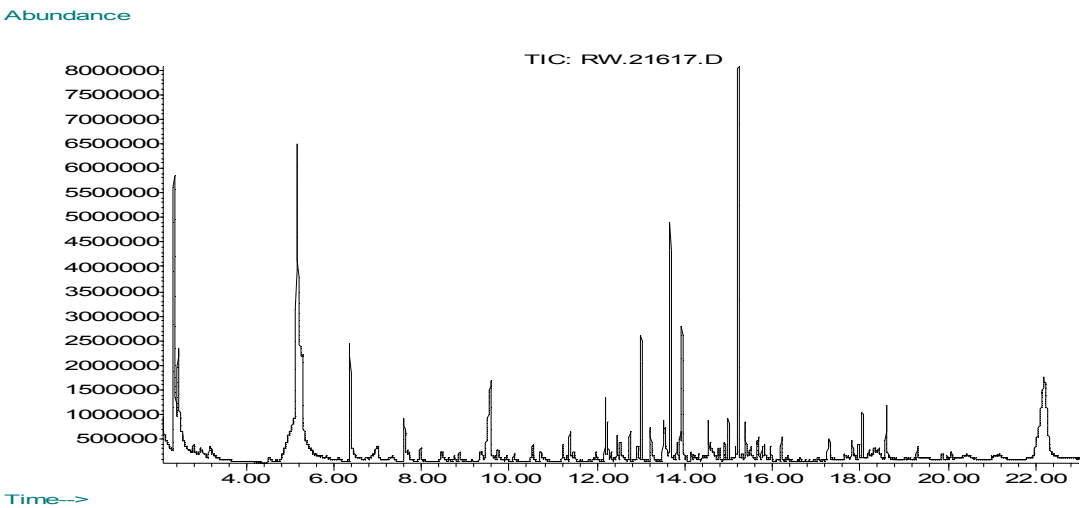

**21614**



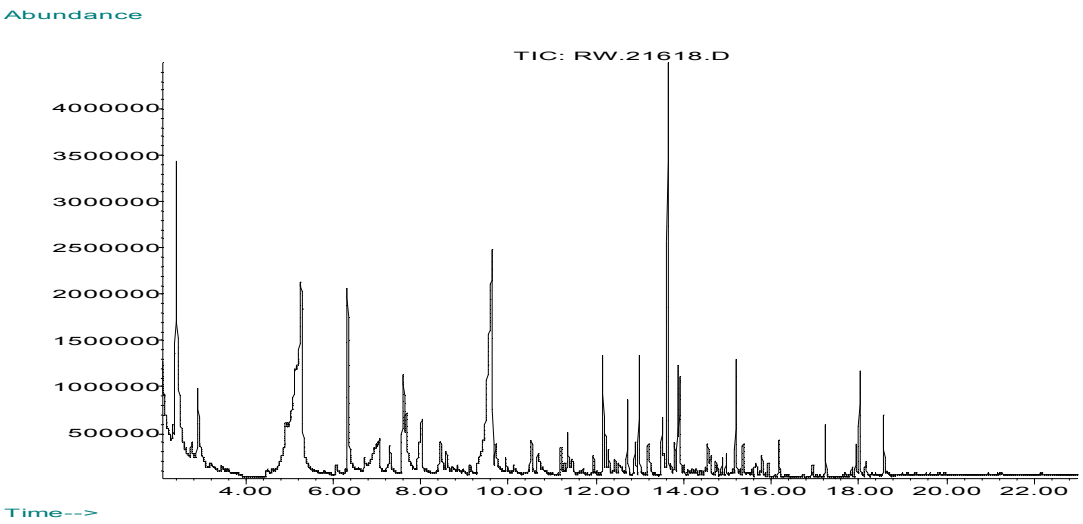

**21615**



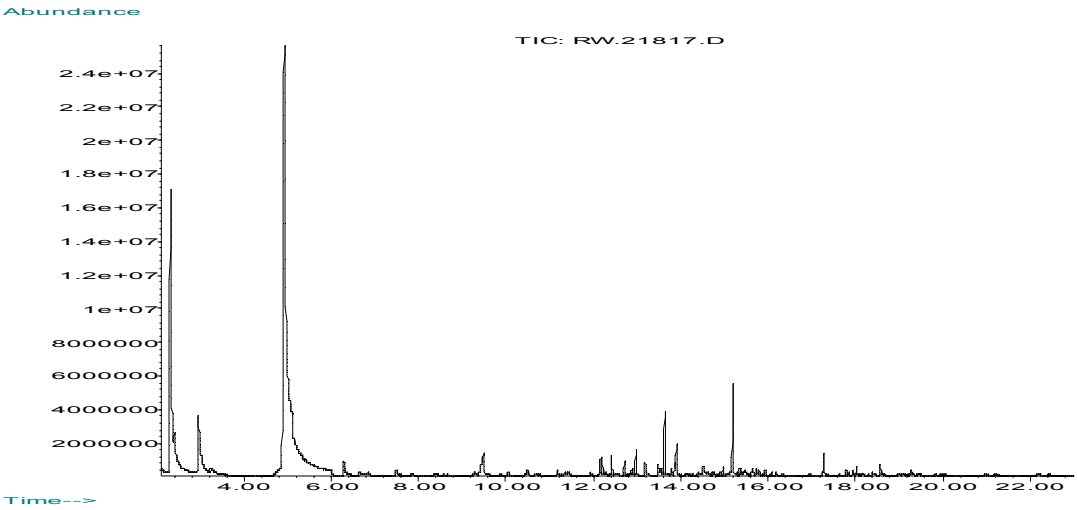

**21616**



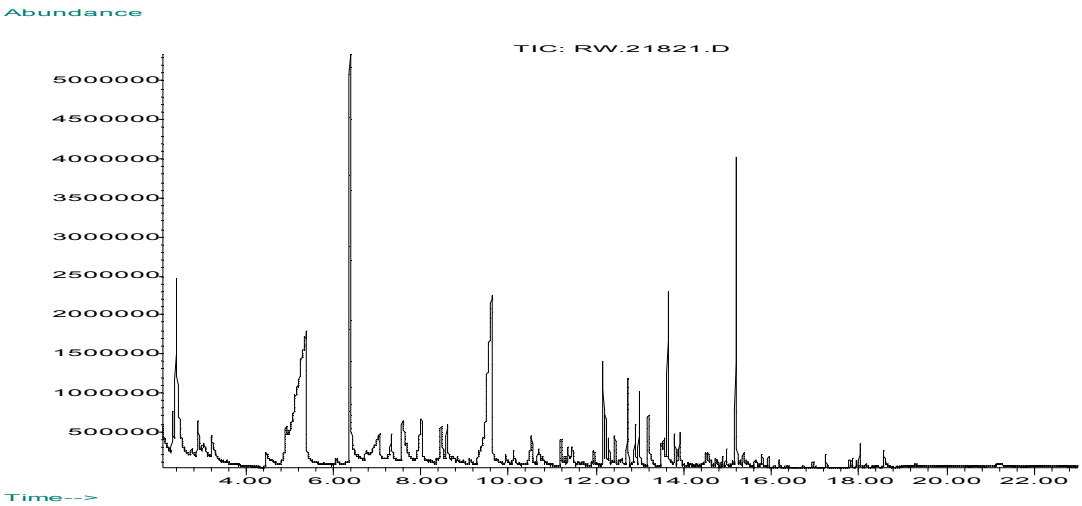

**21617**



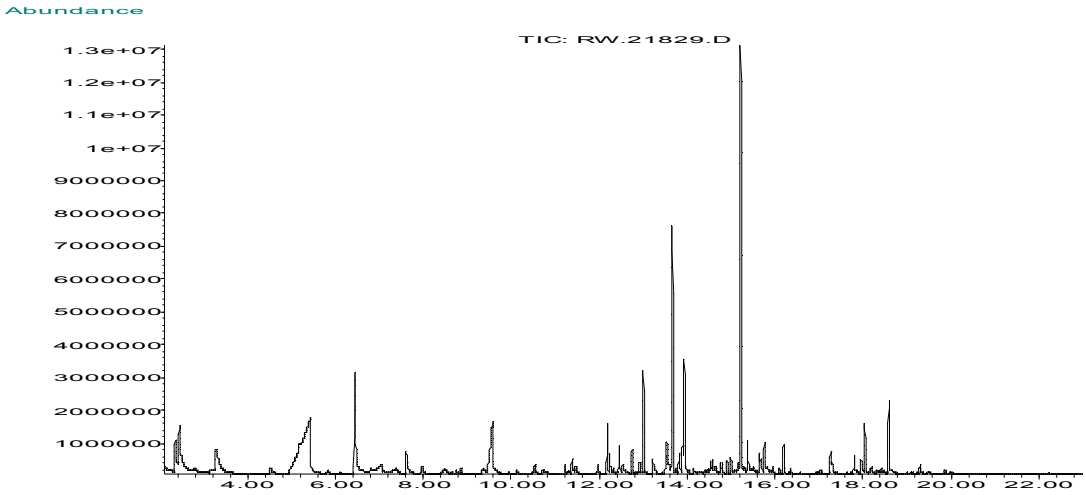

**21618**



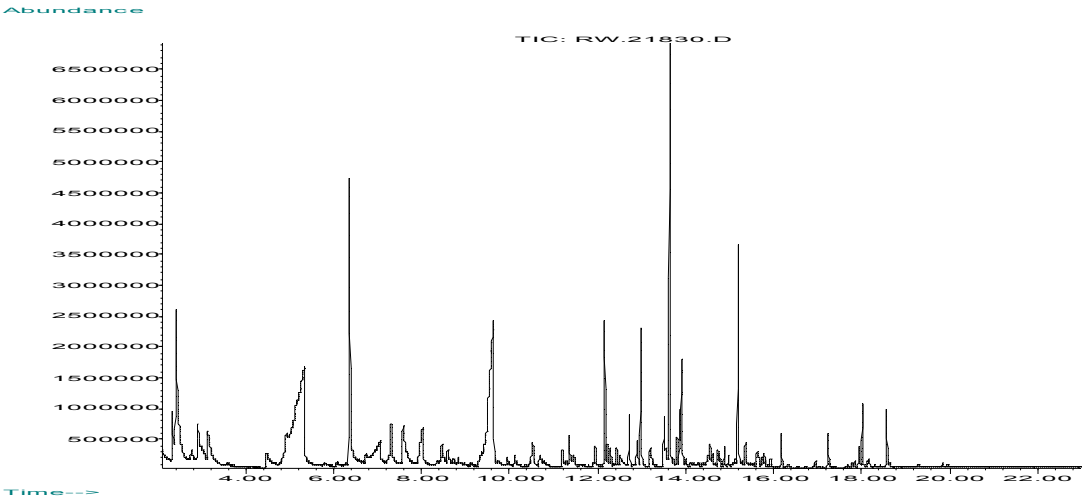

**21817**



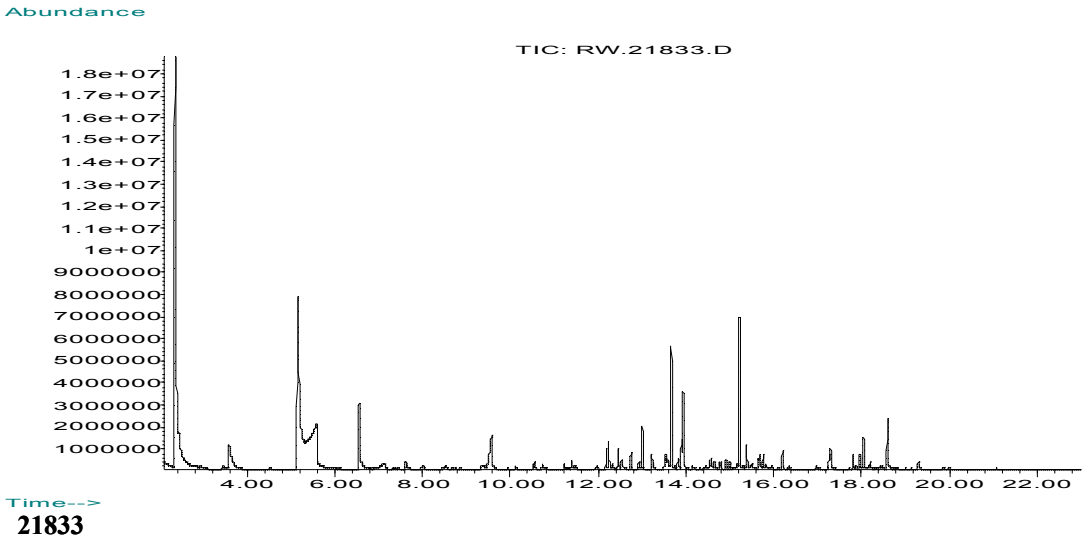

**21821**



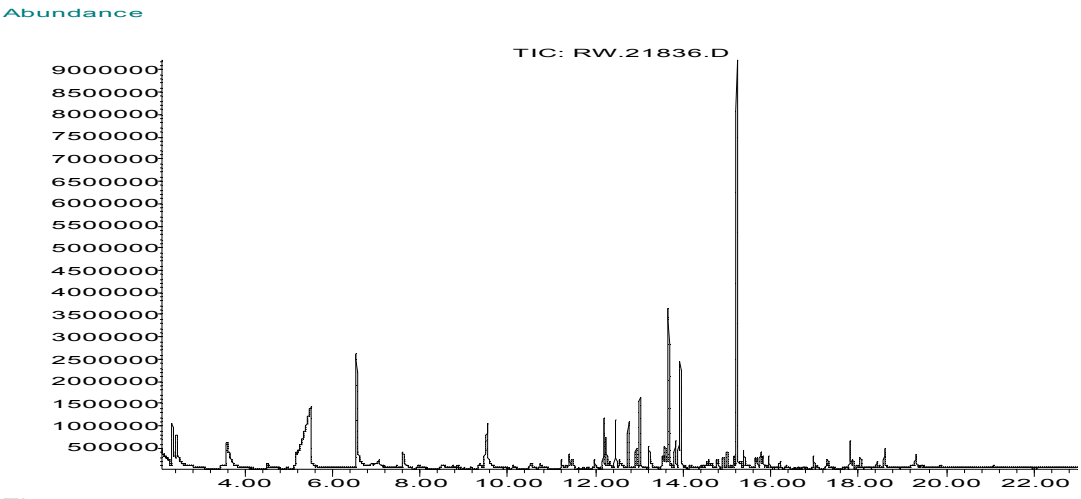

**21829**



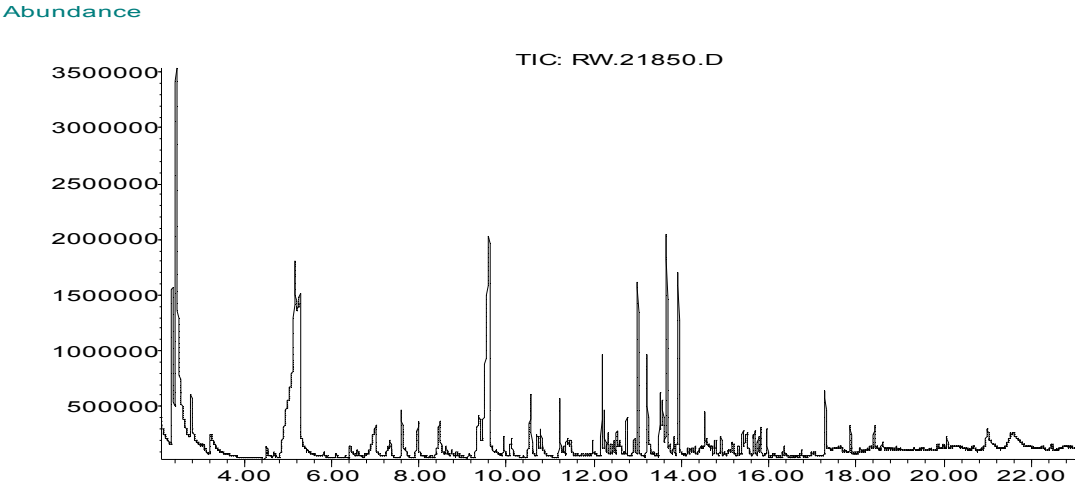

**21830**




**21833**

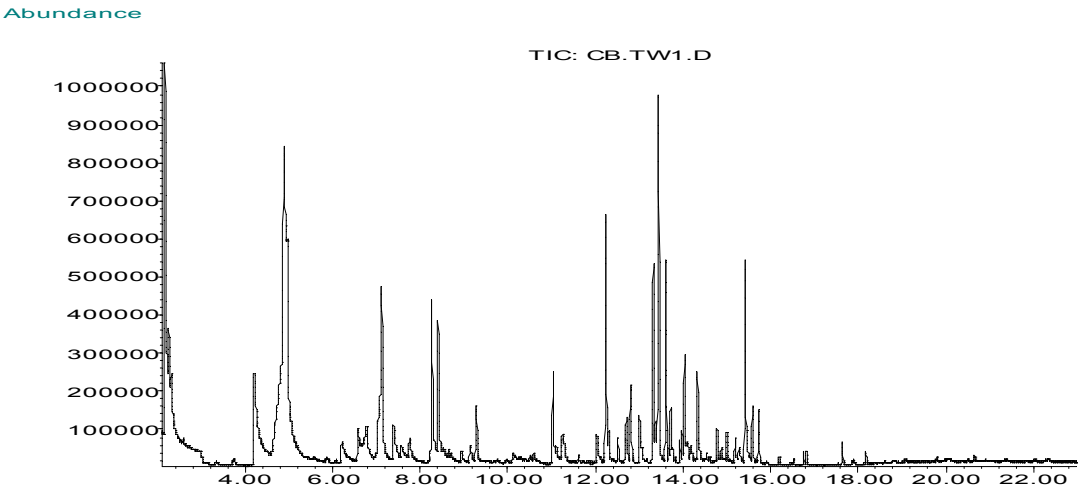

Abundance
TIC: RW.21836.D
9000000
8500000
8000000
7500000
7000000
6500000
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**21836**

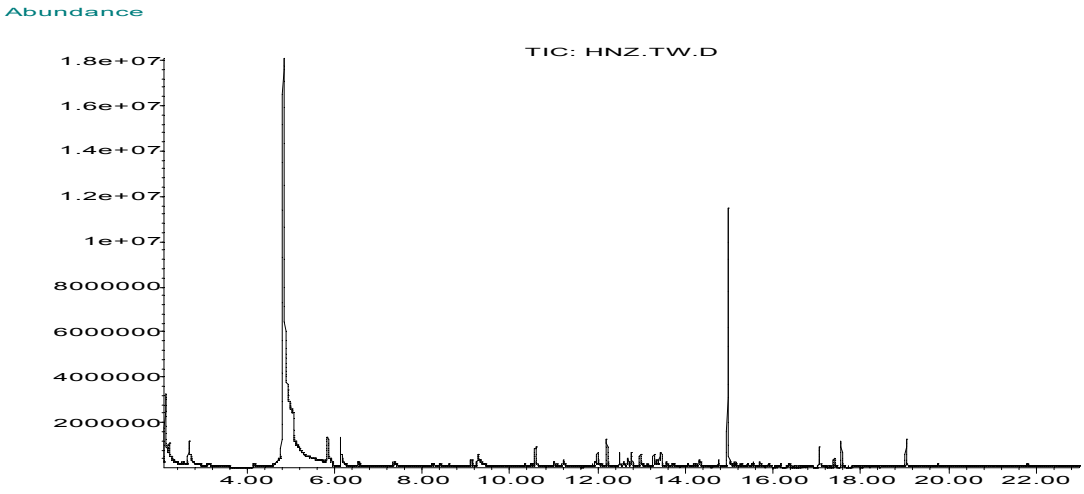

Abundance
TIC: RW.21850.D
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**21850**

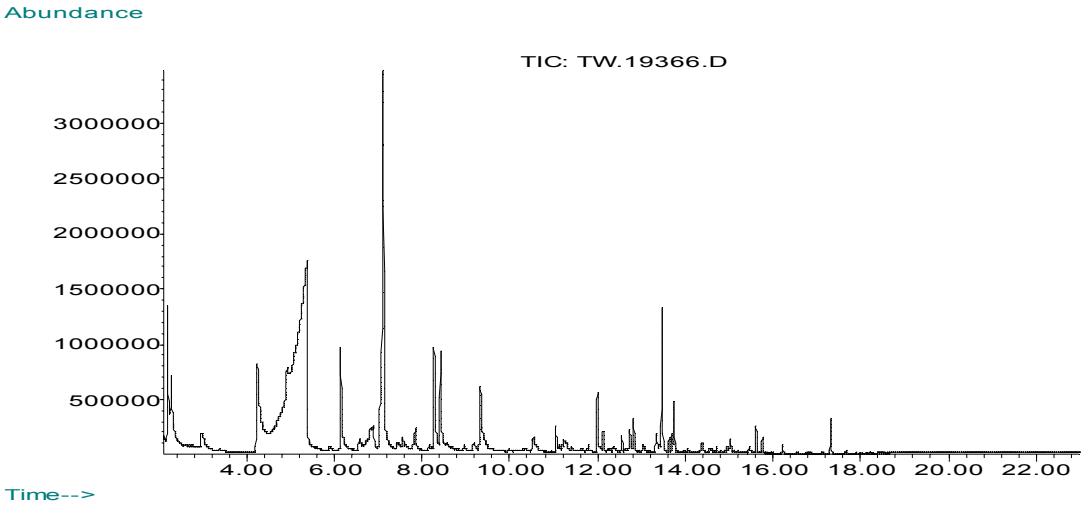
# Appendix 9: Chromatograms of Tawari Honey Samples



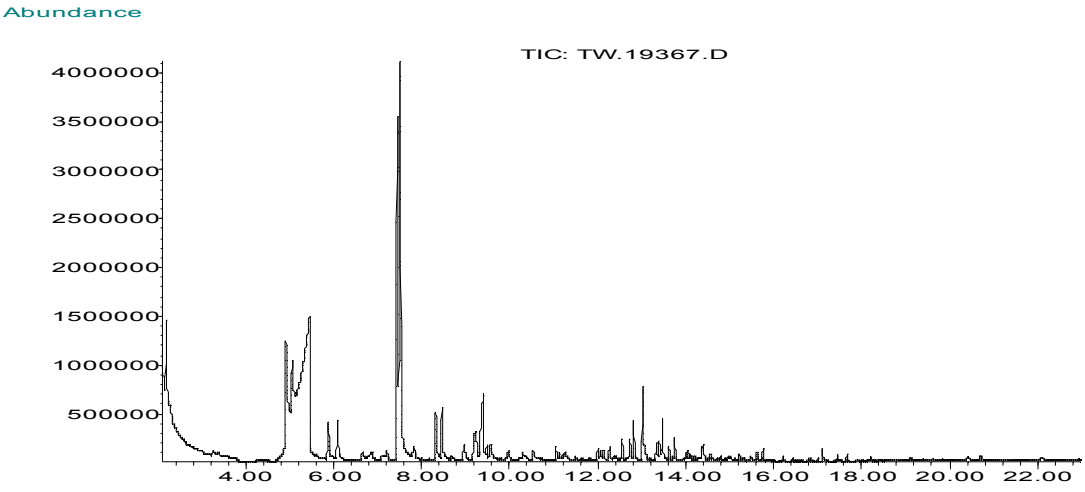

**CB.TW.1**



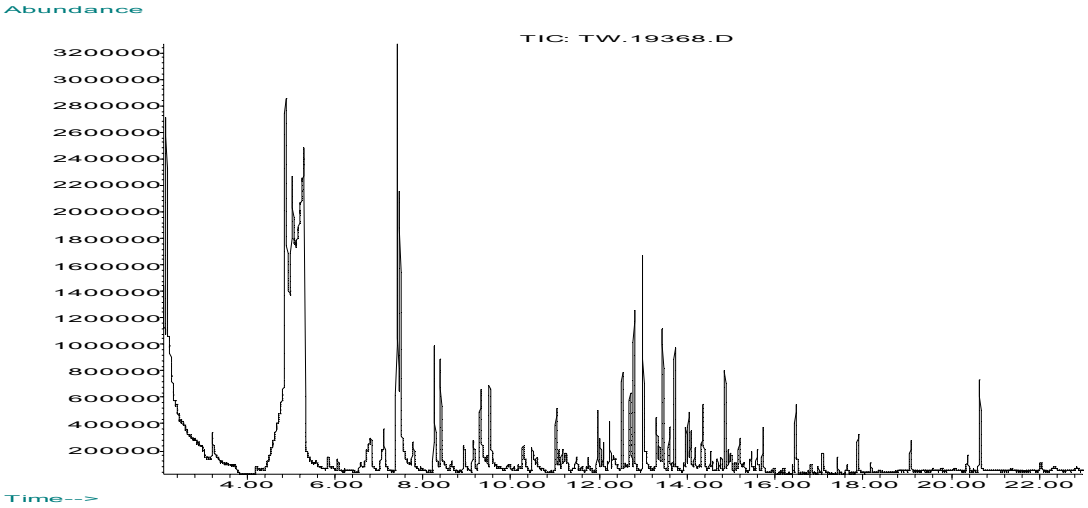

**HNZ.TW**



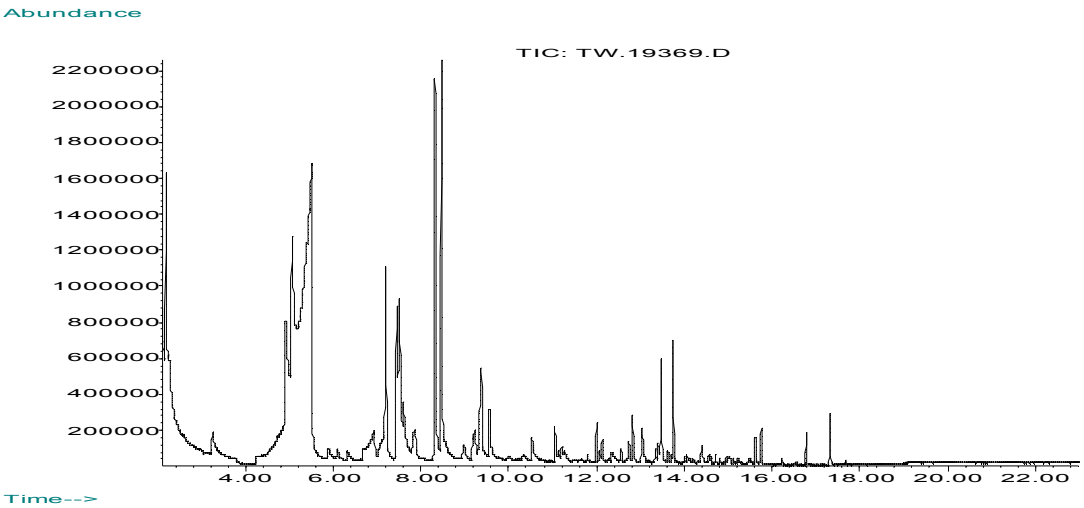

**19366**



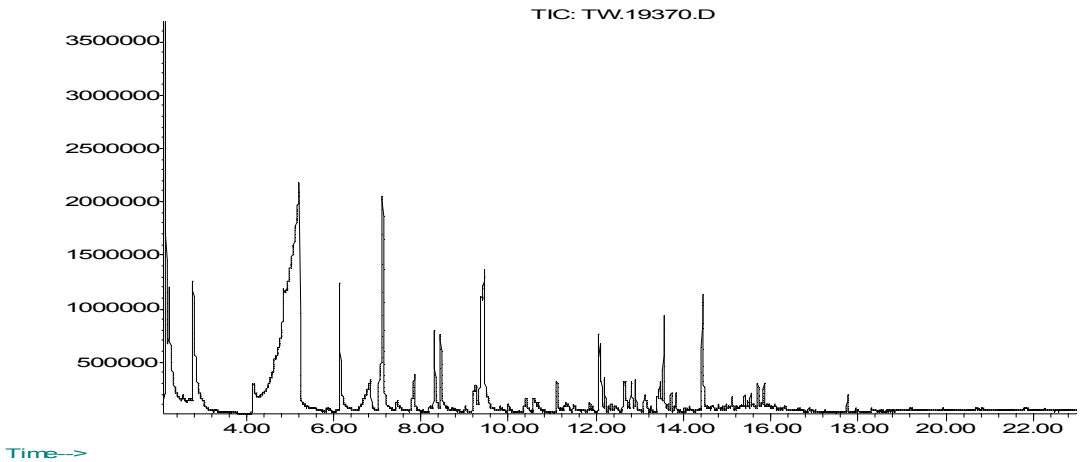

**19367**



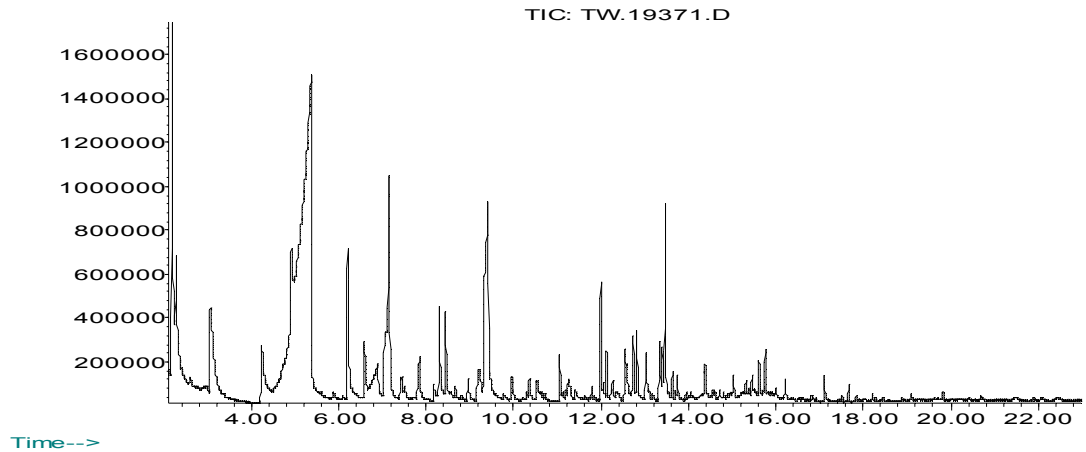

**19368**



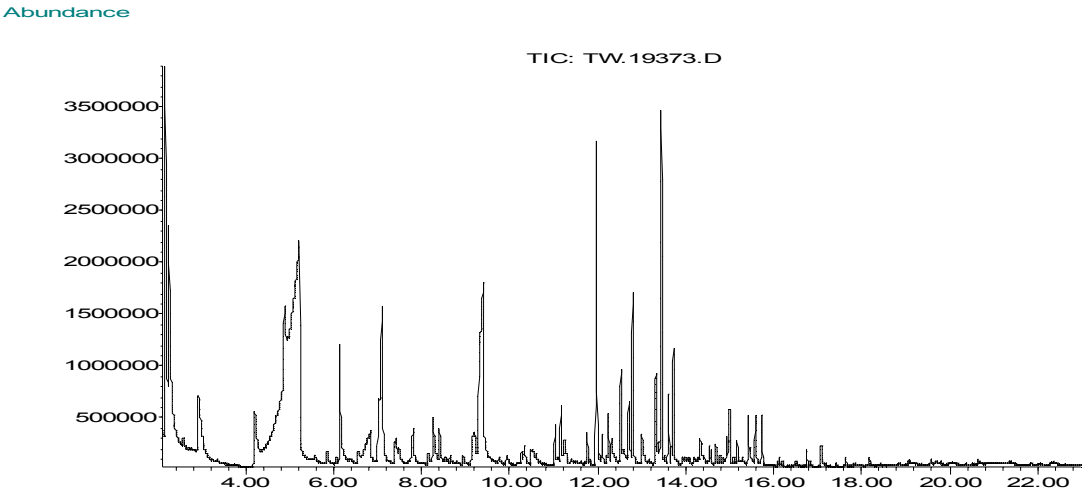

**19369**



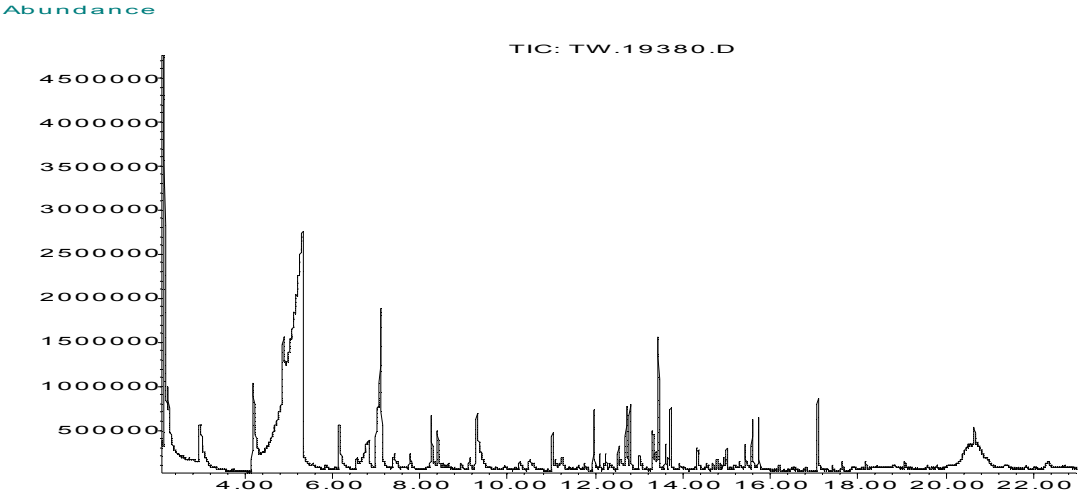

**19370**



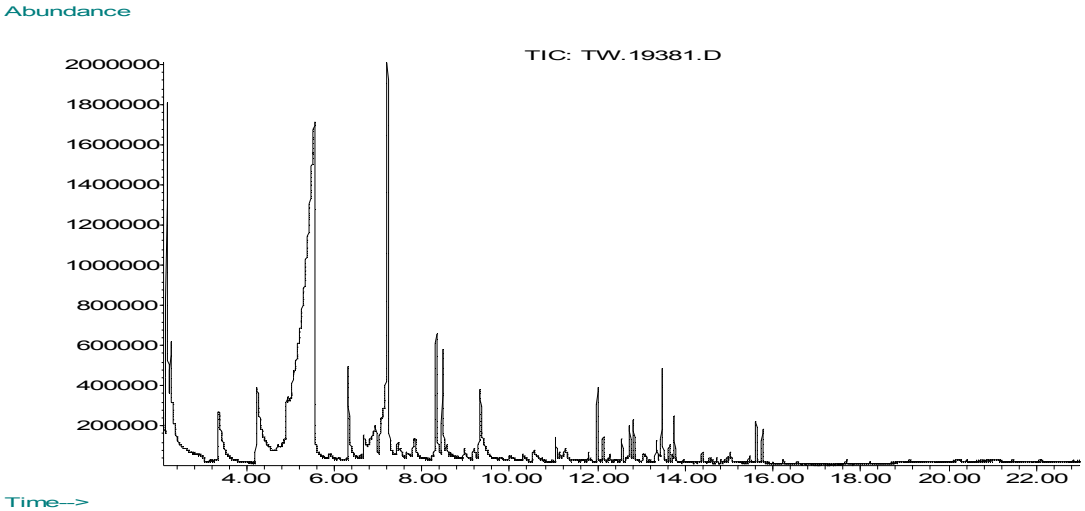

**19371**



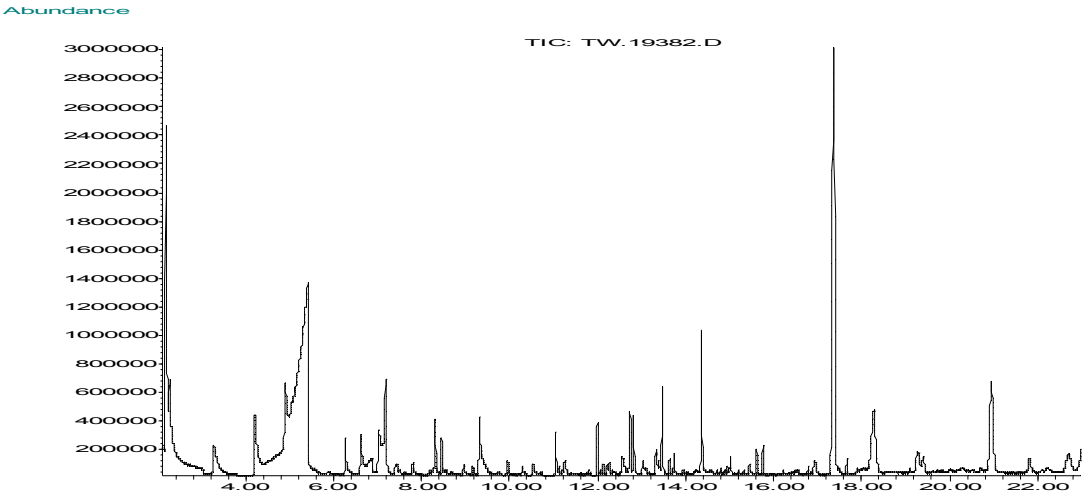

**19373**



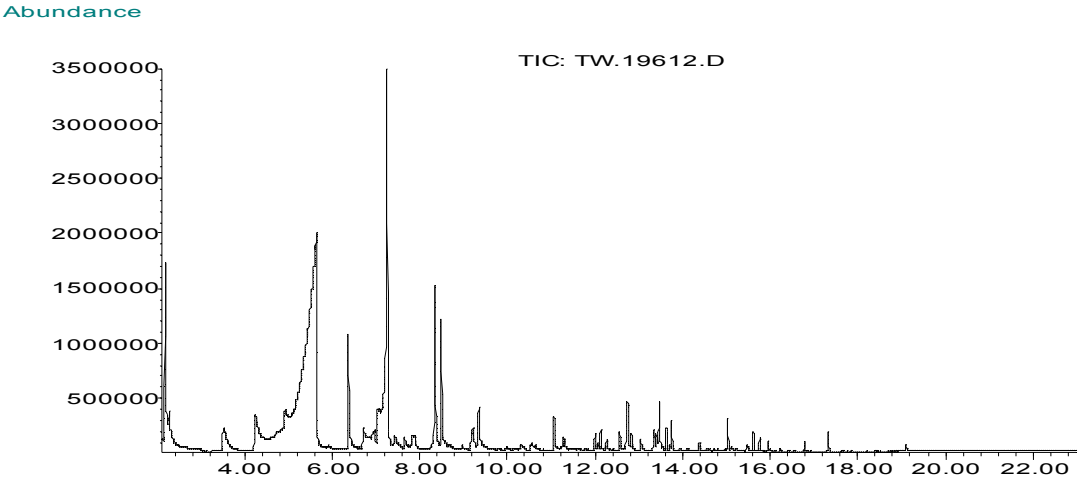

**19380**



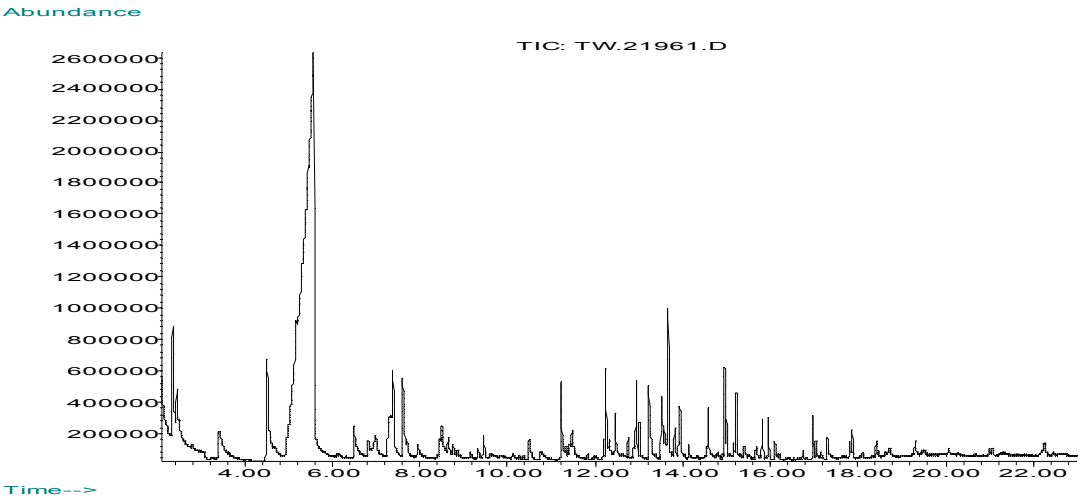

**19381**



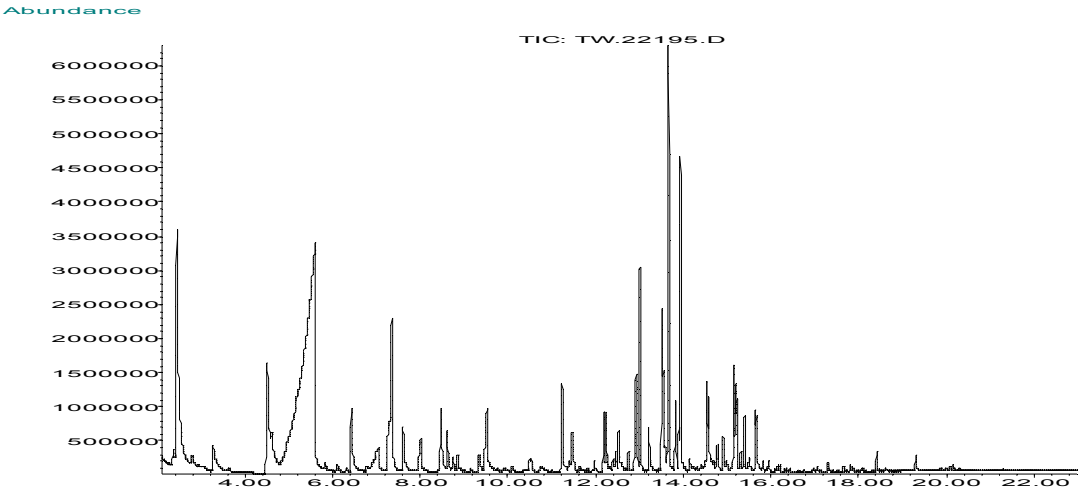

**19382**



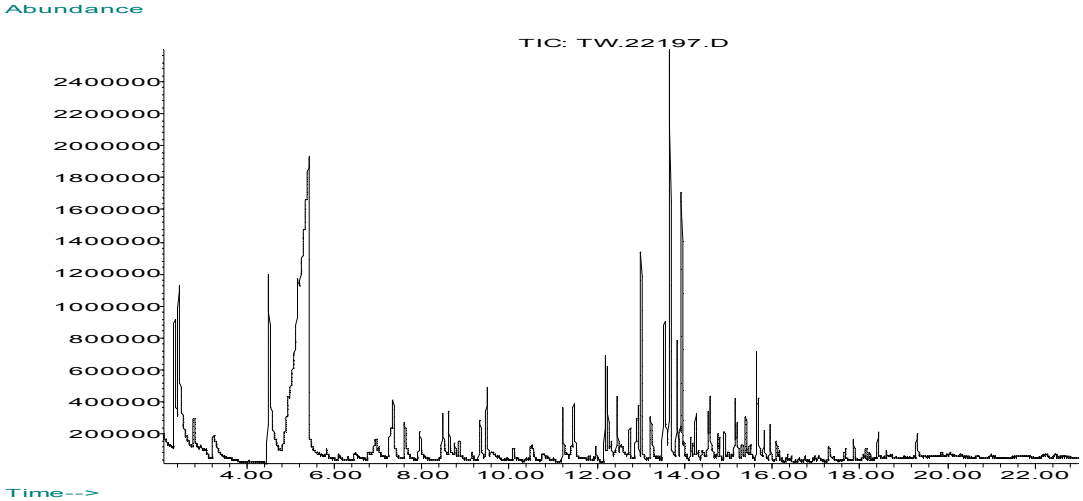

**19612**



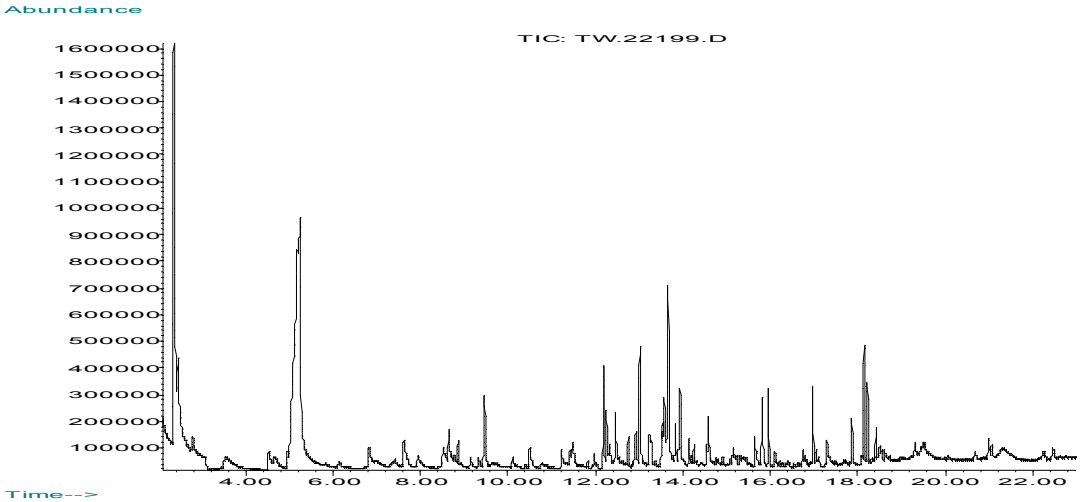

**21961**



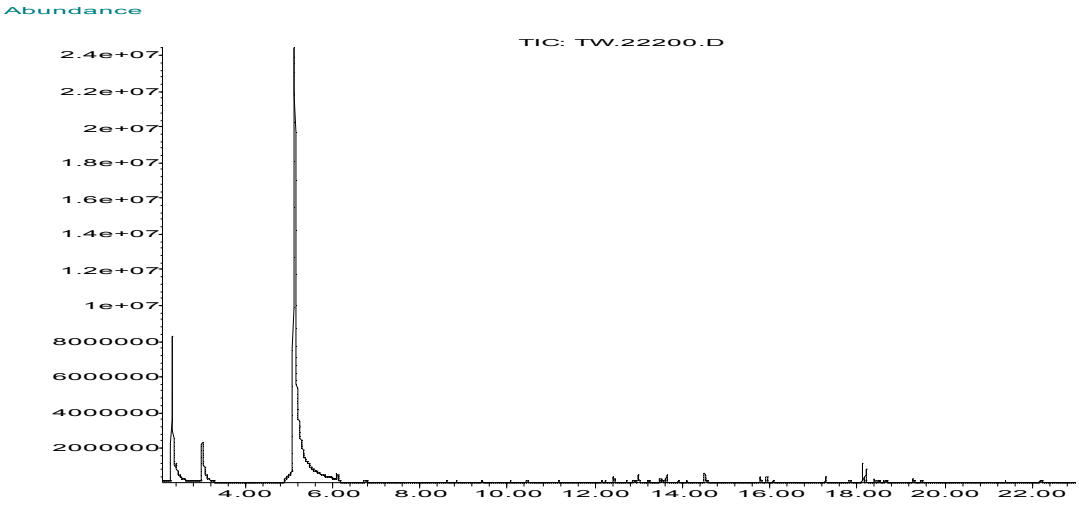

**22195**



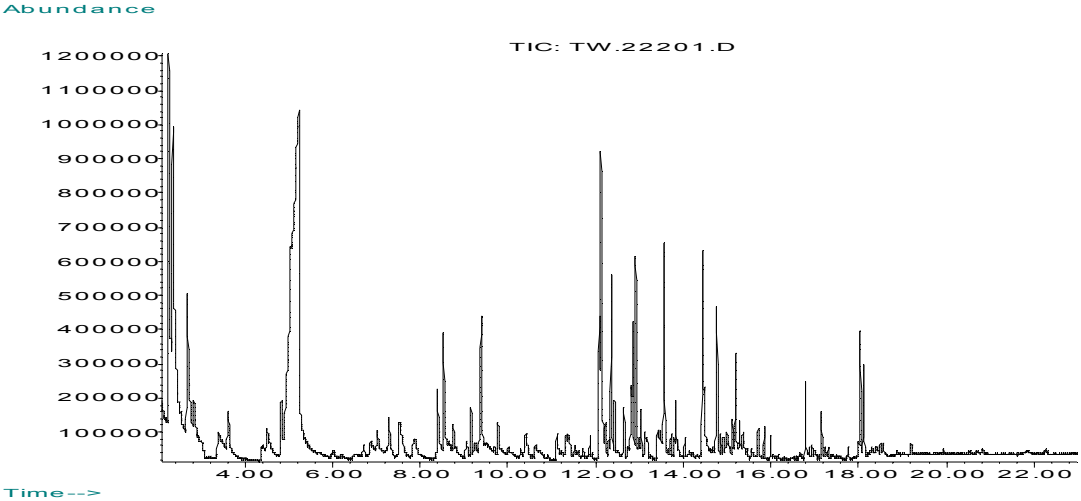

**22197**



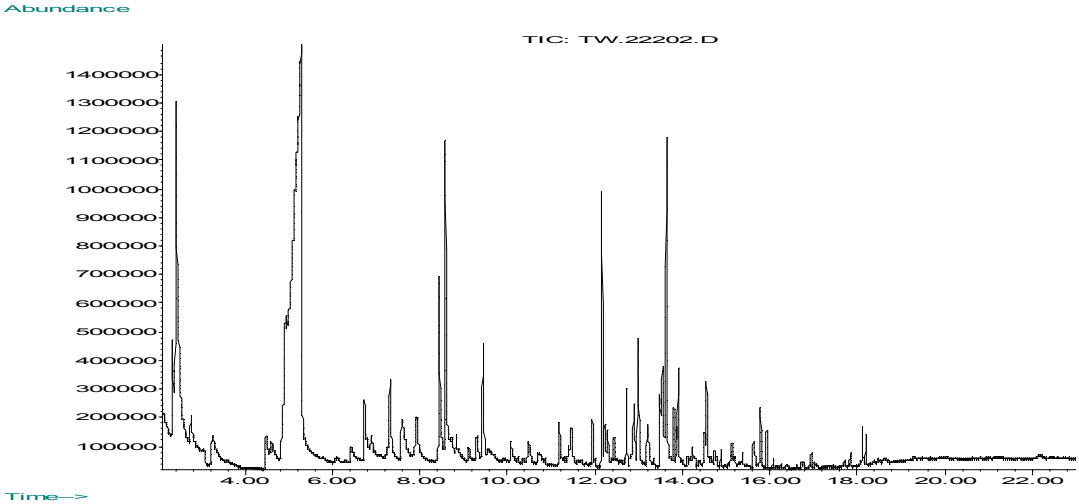

**22199**



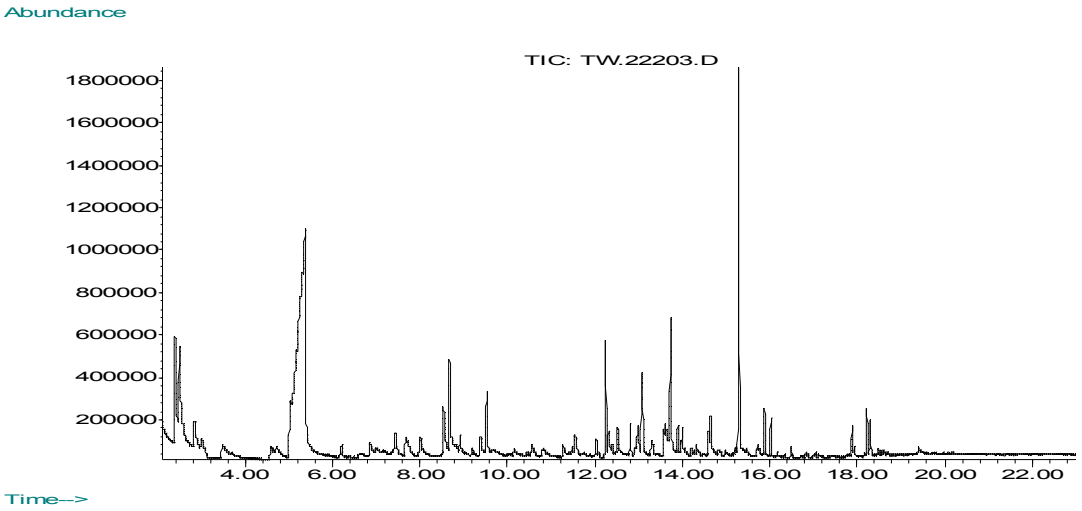

**22200**



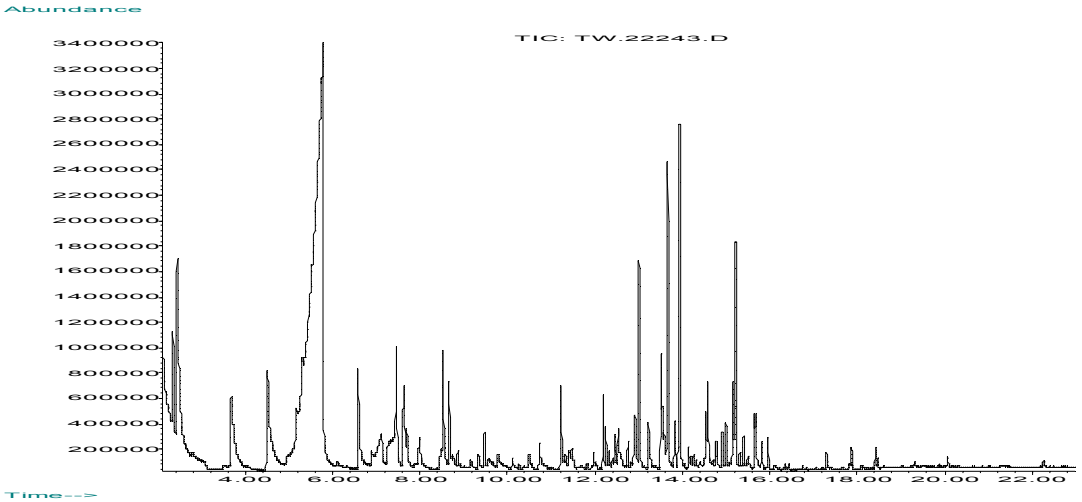

**22201**



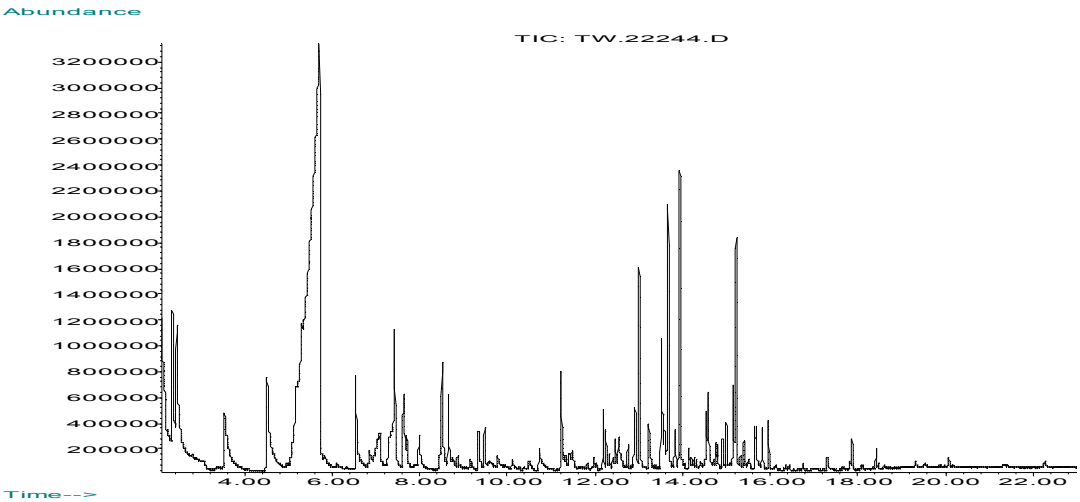

**22202**



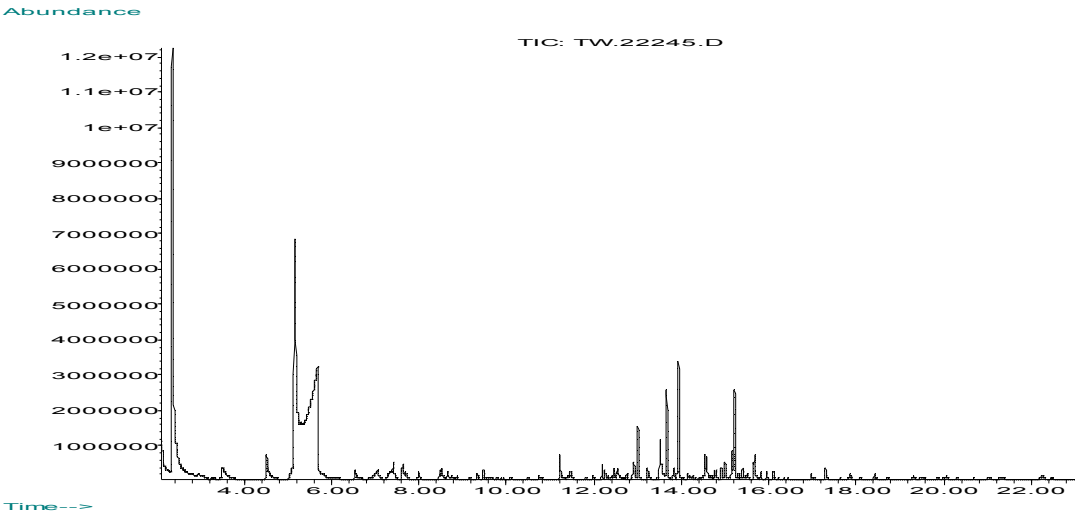

**22203**



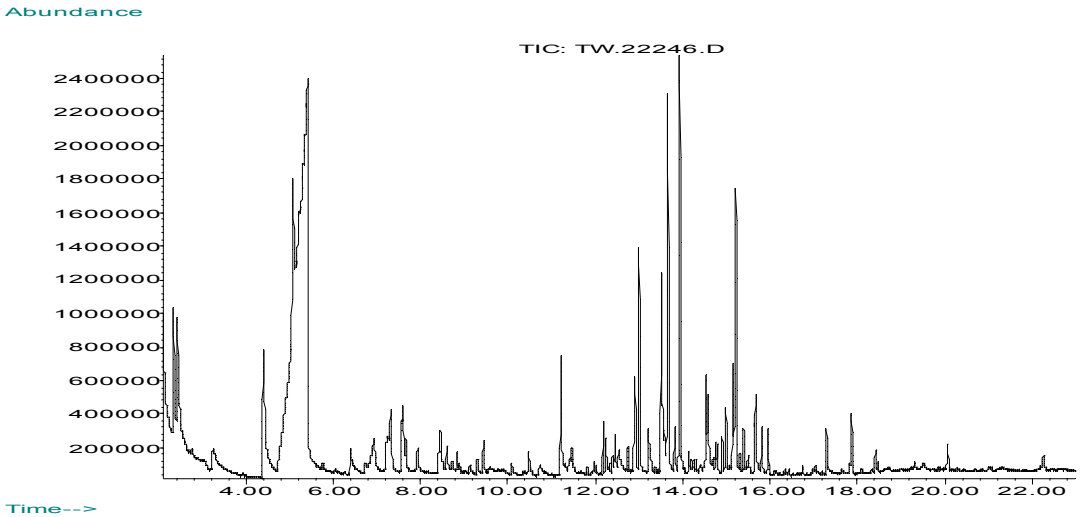

**22243**



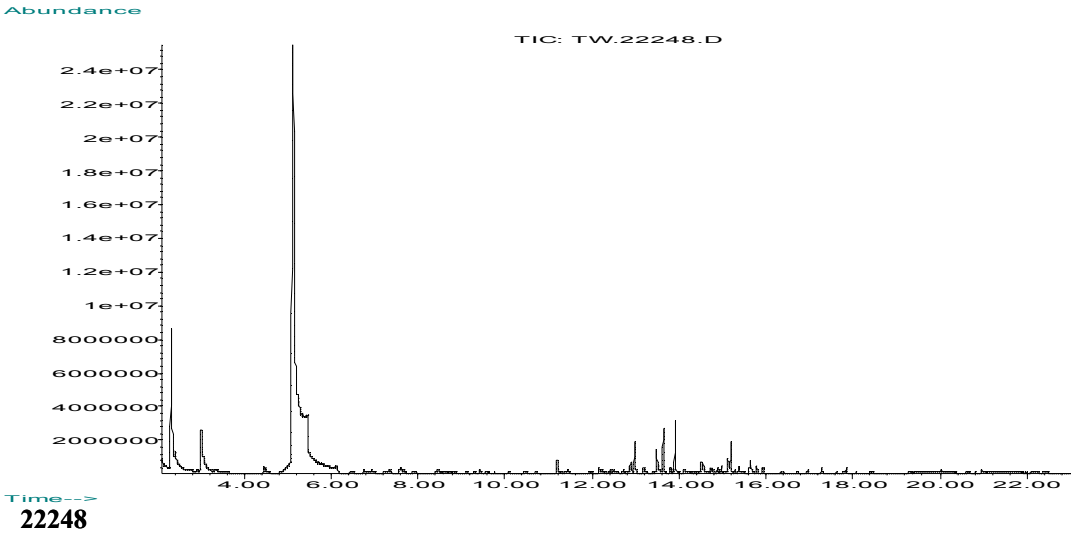

**22244**



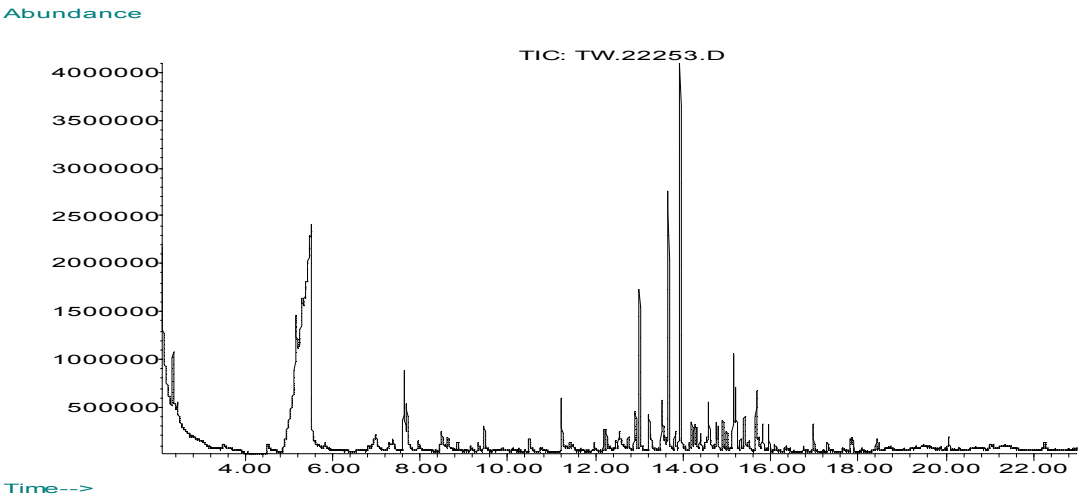

**22245**



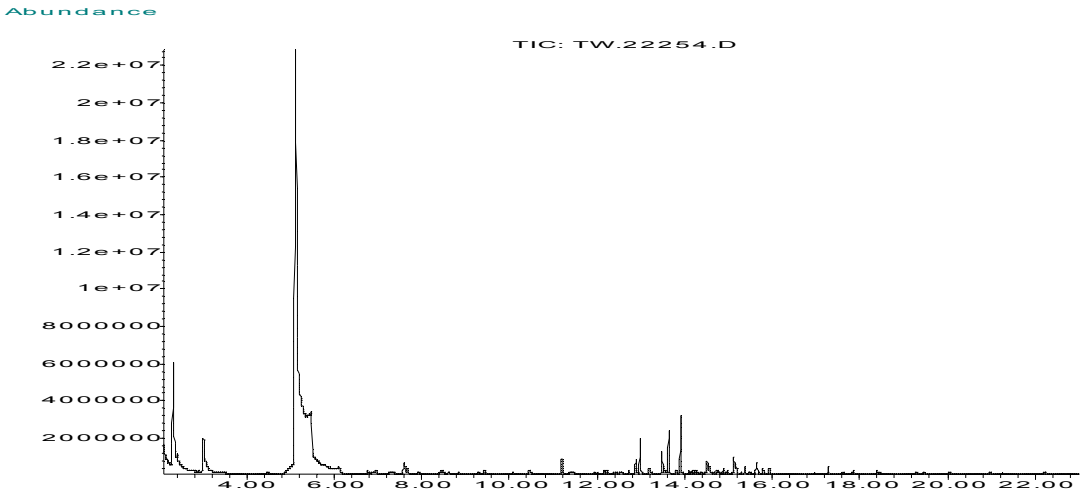

**22246**



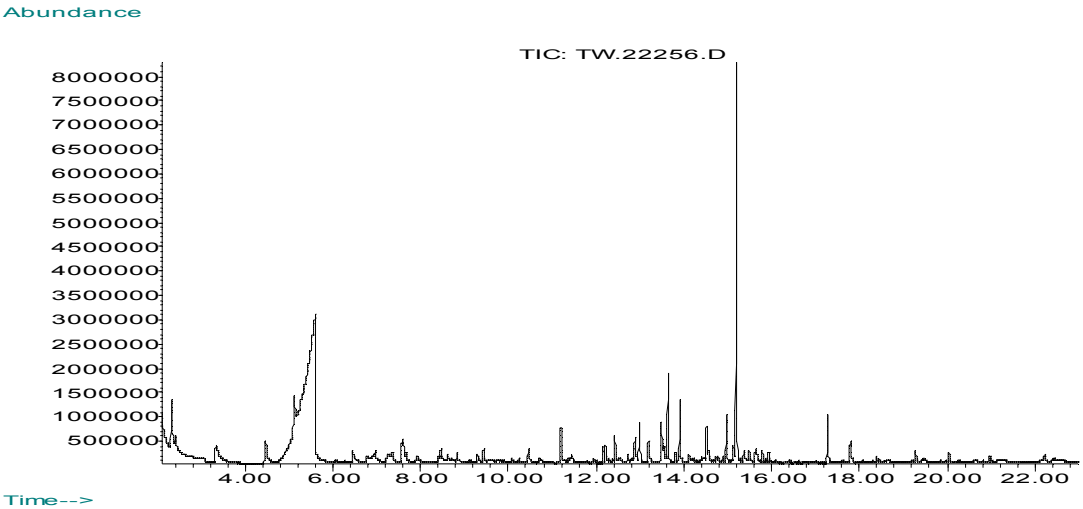

**22248**



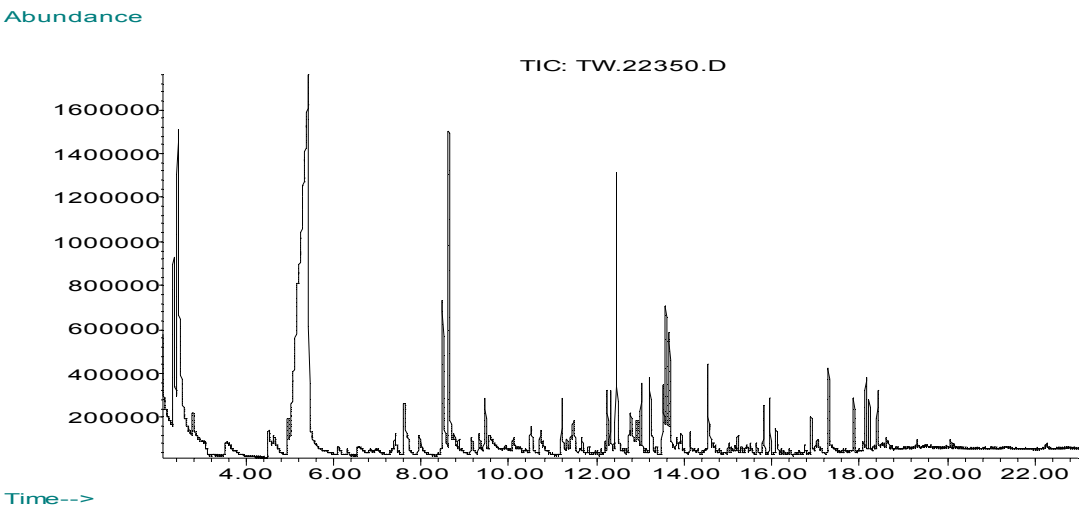

**22253**



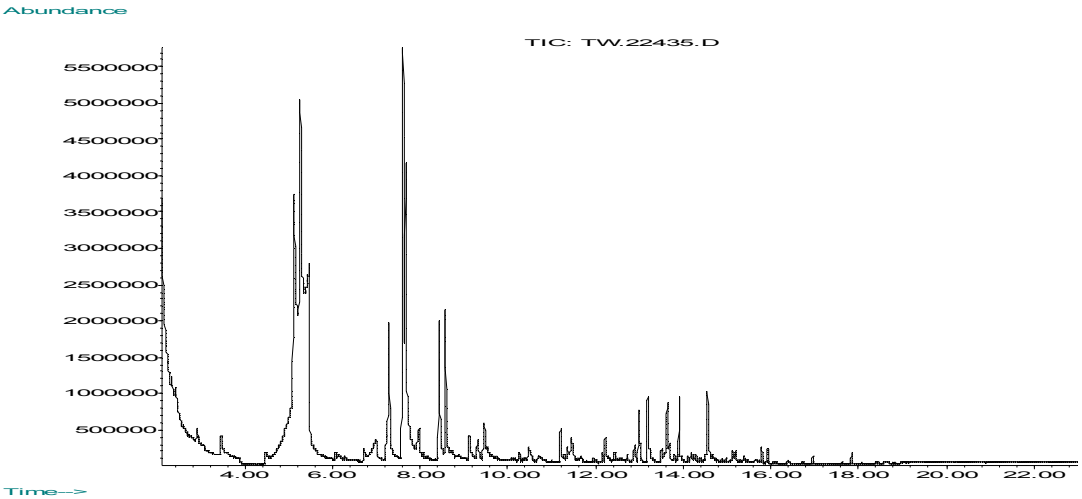

**22254**



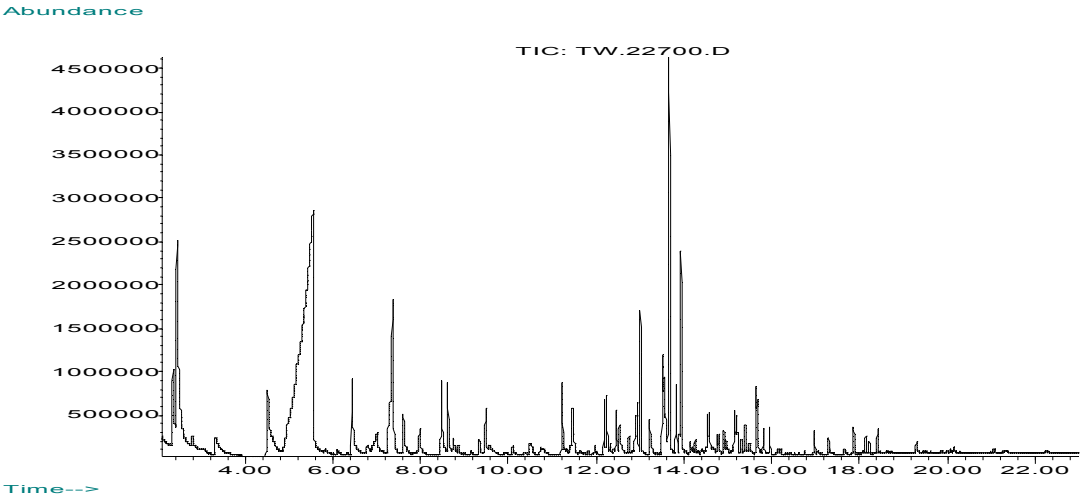

**22256**



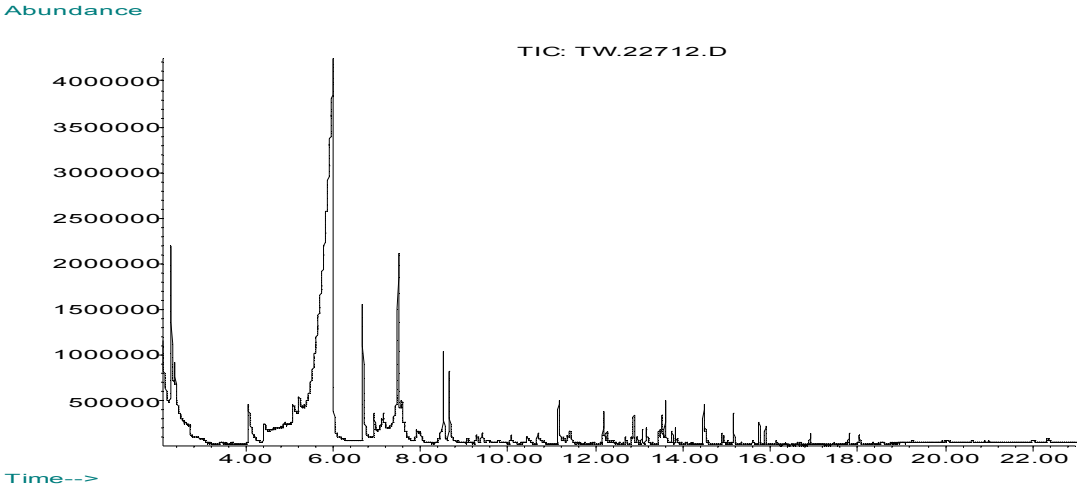

**22350**




**22435**

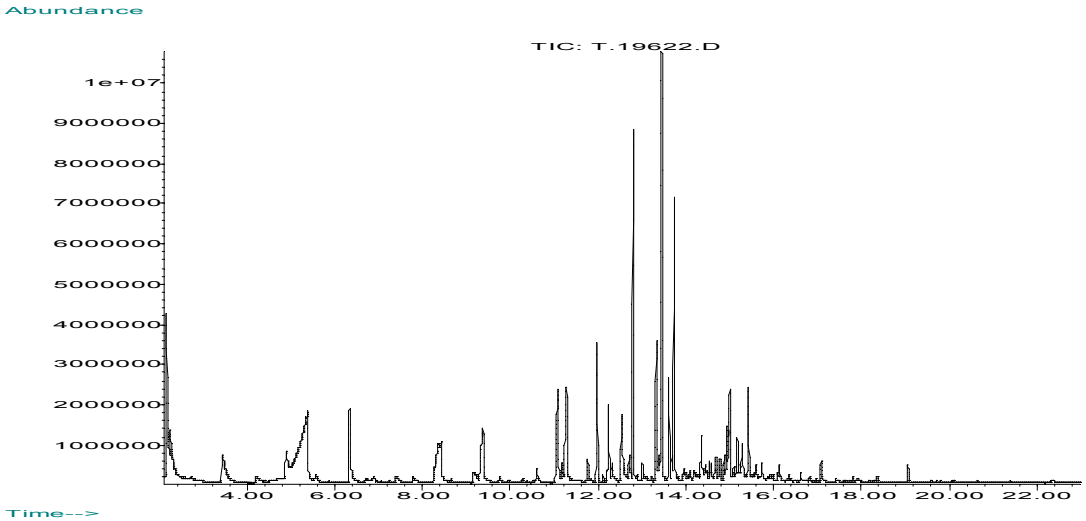



**22700**

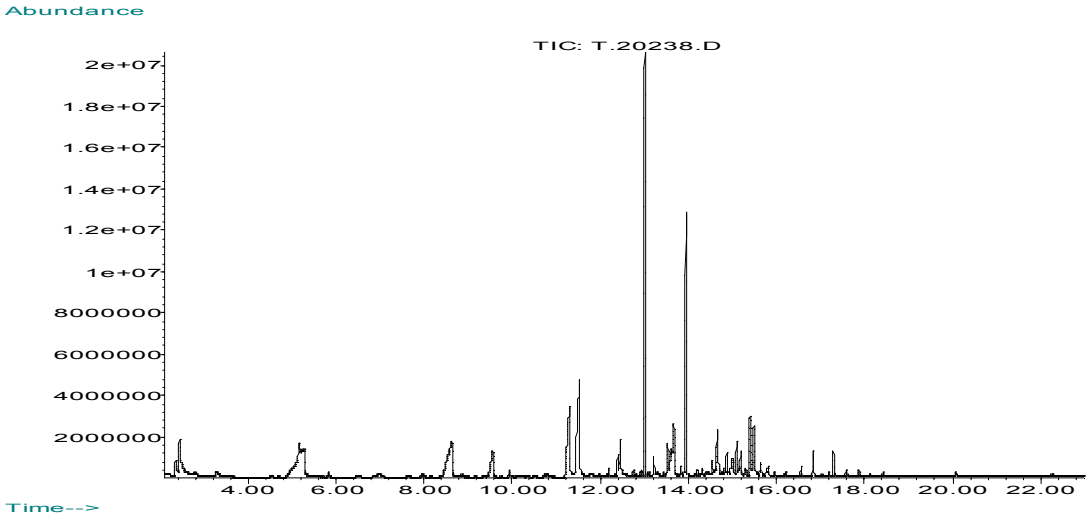

Abundance
TIC: TW.22712.D
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**22712**

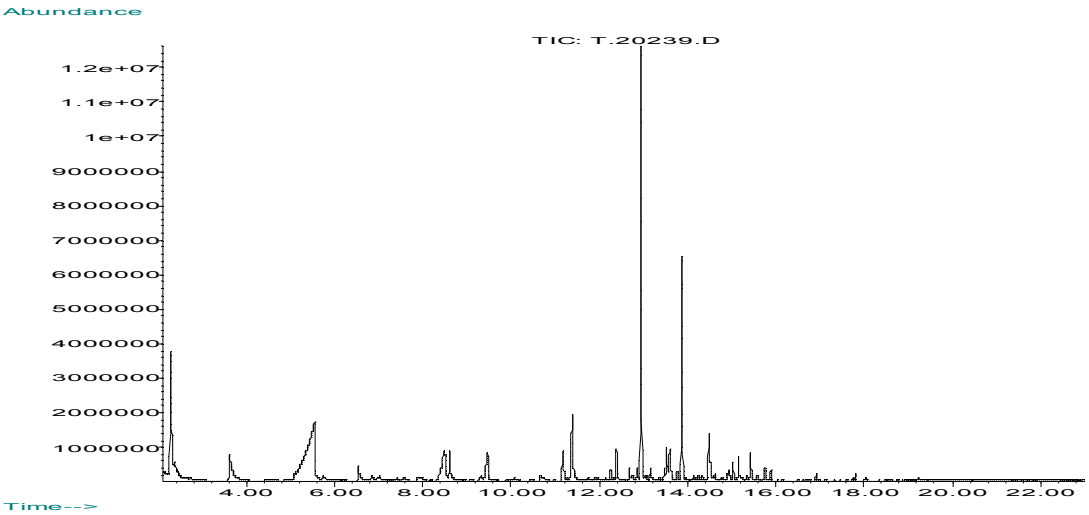
# Appendix 10: Chromatograms of Thyme Honey Samples



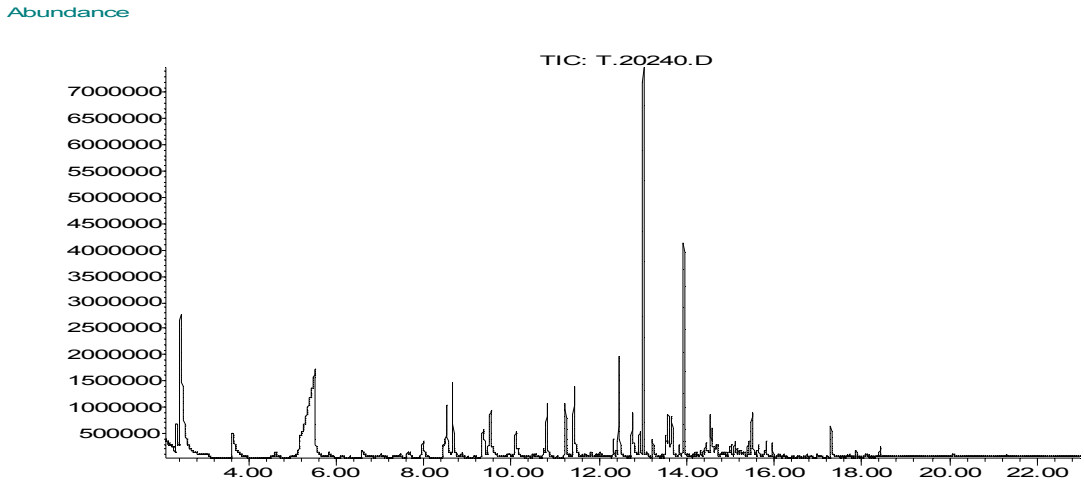

**19622**



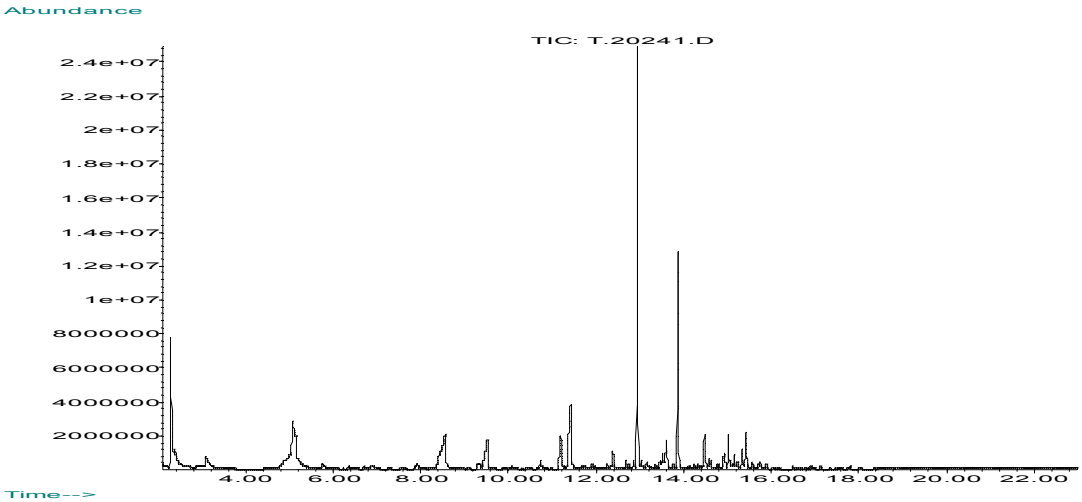

**20238**



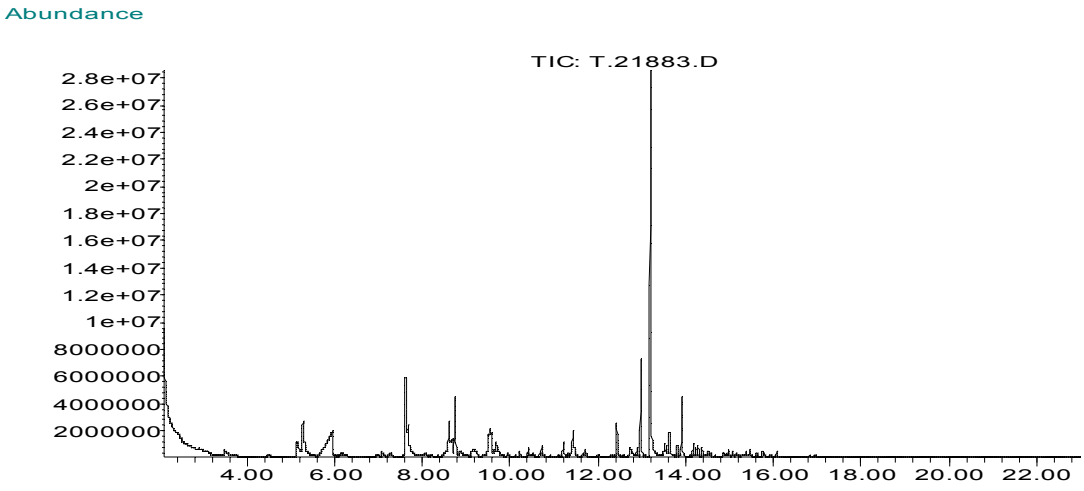

**20239**



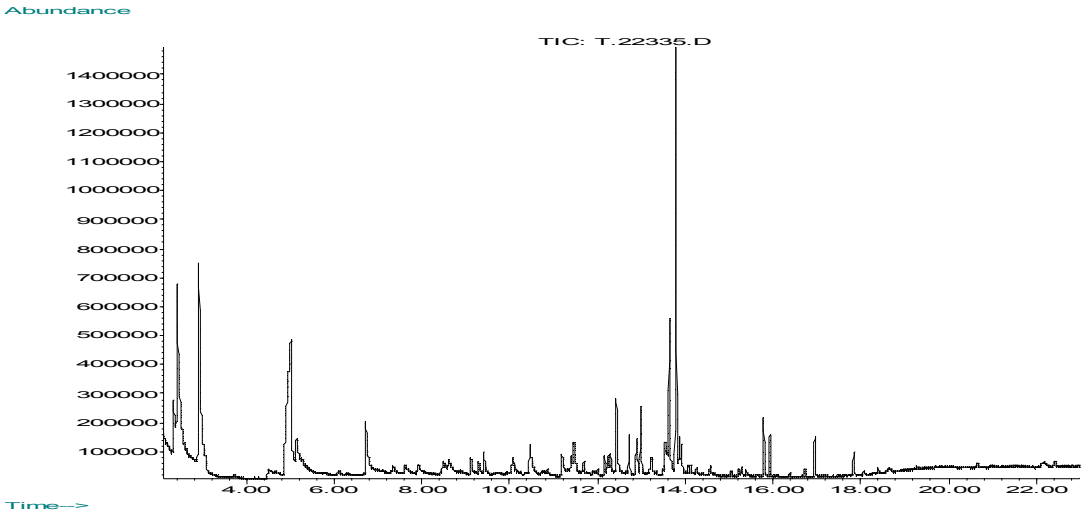

**20240**



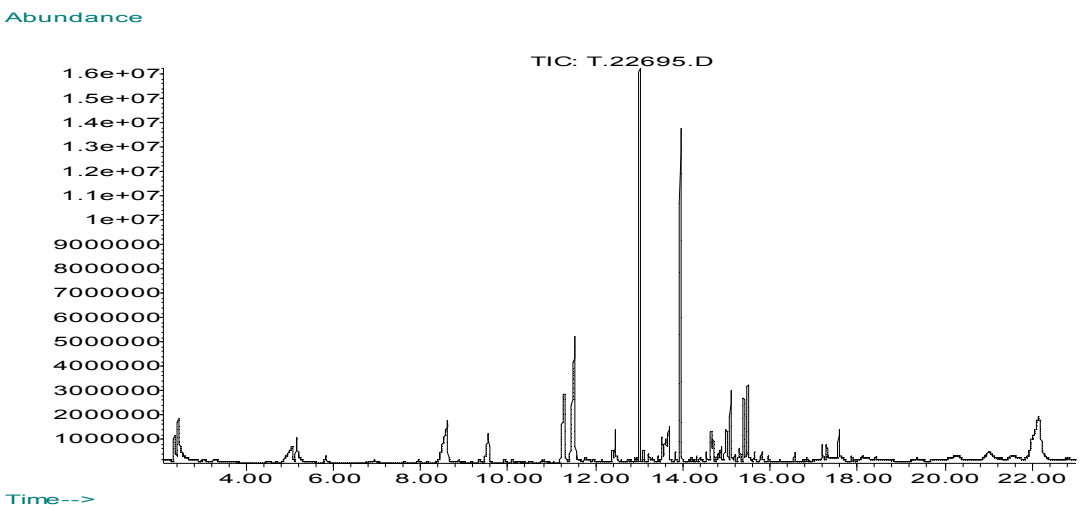

**20241**



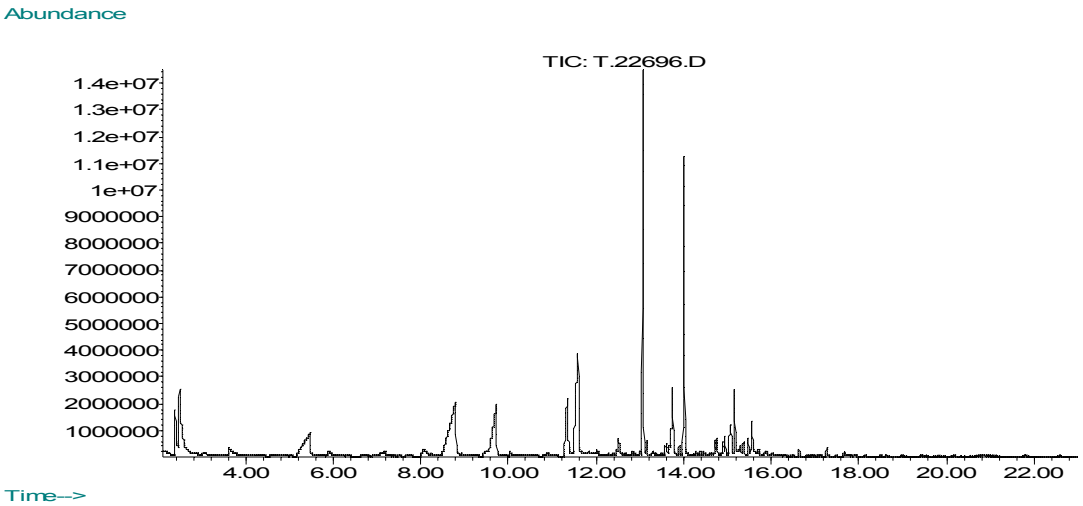

**21883**



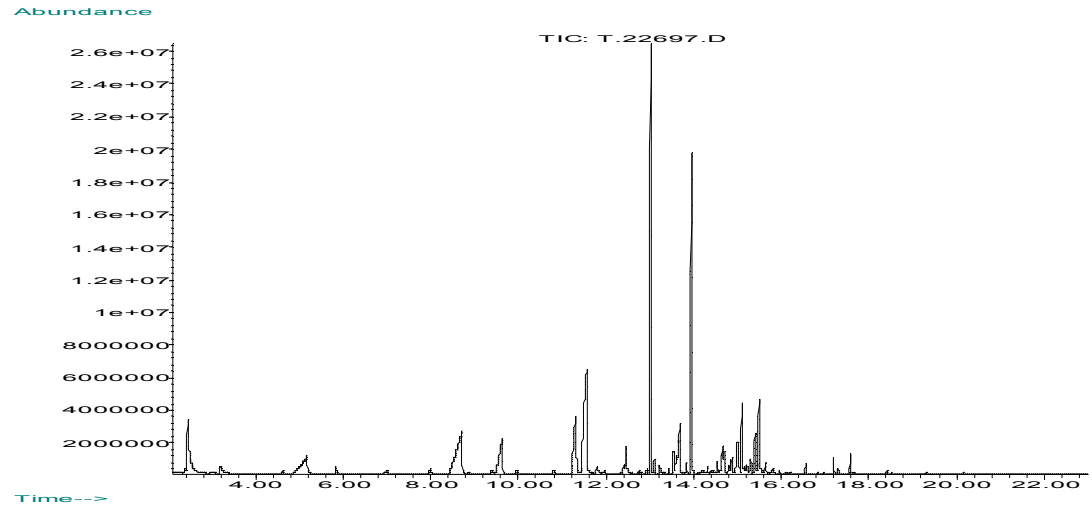

**22335**



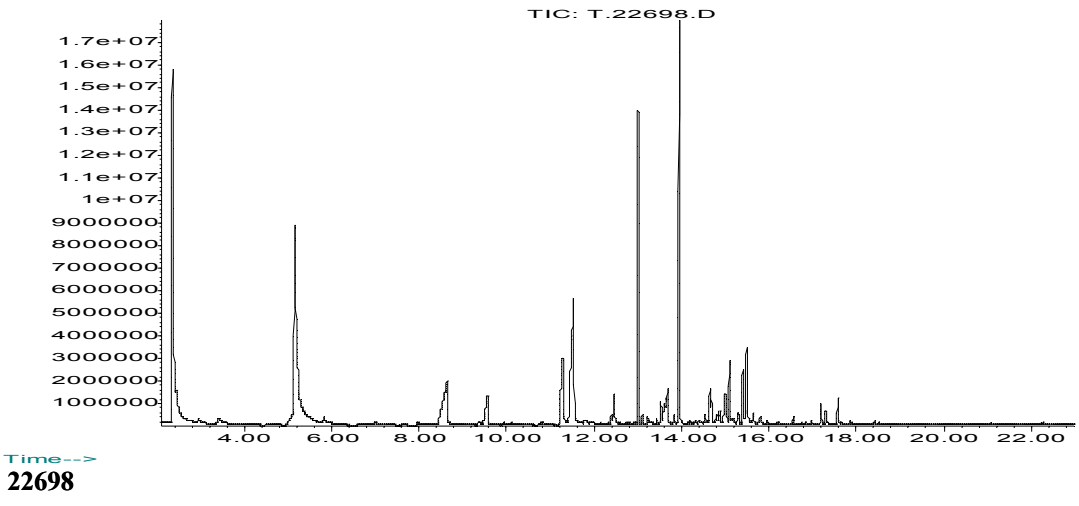

**22695**



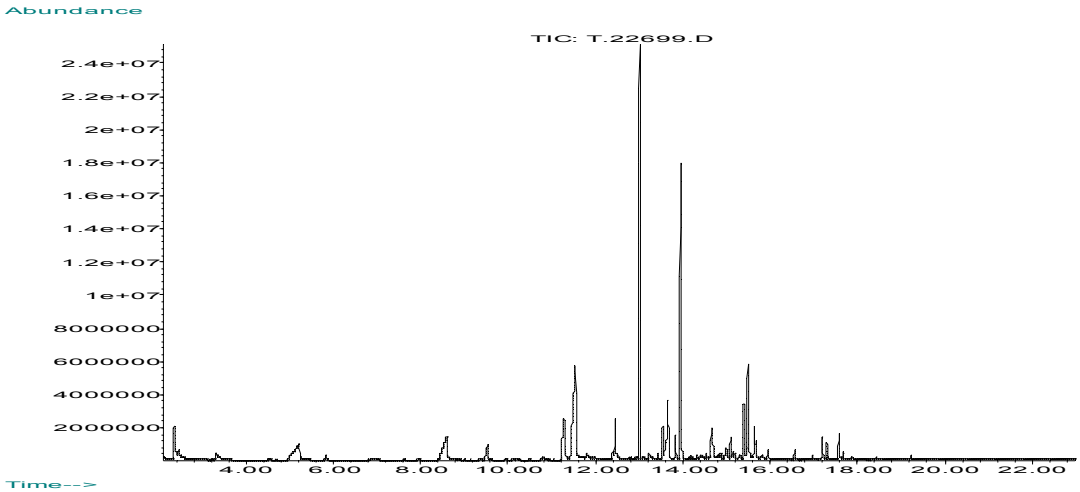

**22696**




**22697**

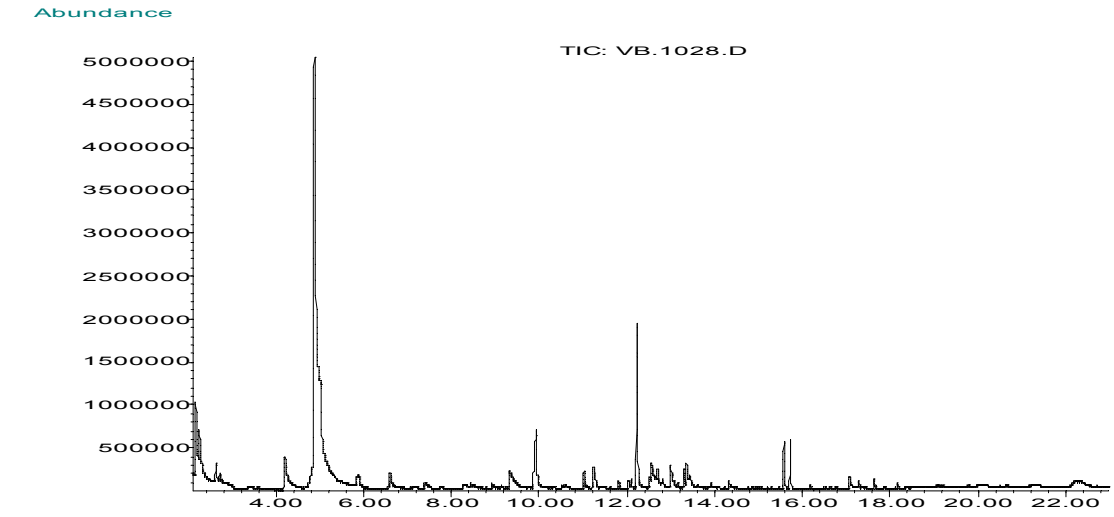



**22698**

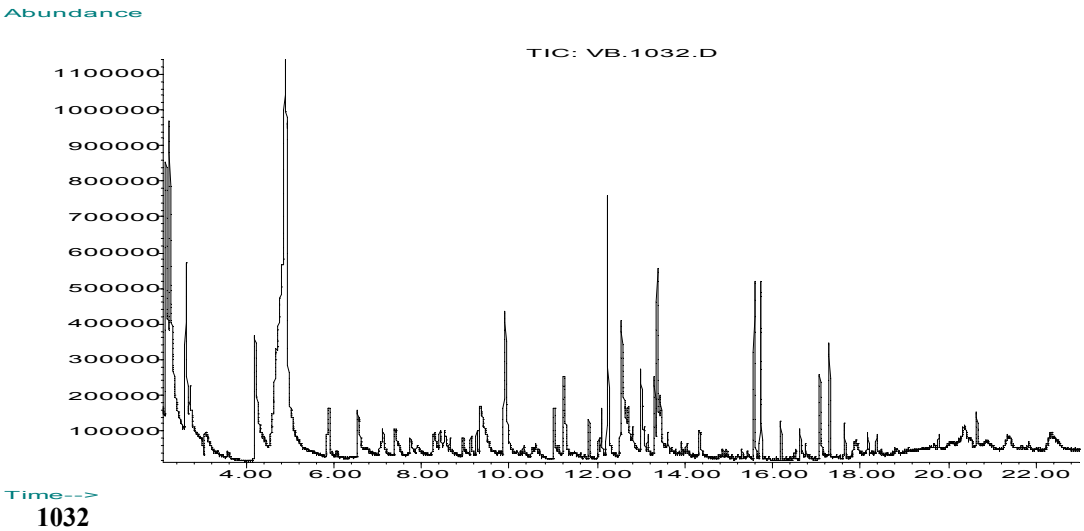

Abundance
TIC: T.22699.D
2.4e+07
2.2e+07
2e+07
1.8e+07
1.6e+07
1.4e+07
1.2e+07
1e+07
8000000
6000000
4000000
2000000
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Time-->

**22699**

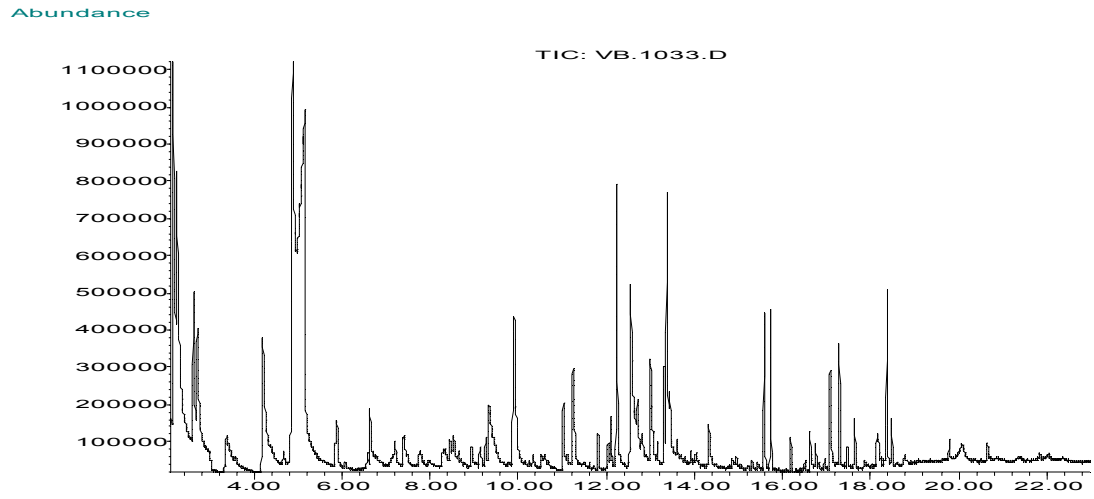
# Appendix 11: Chromatograms of Viper's Bugloss Honey Samples



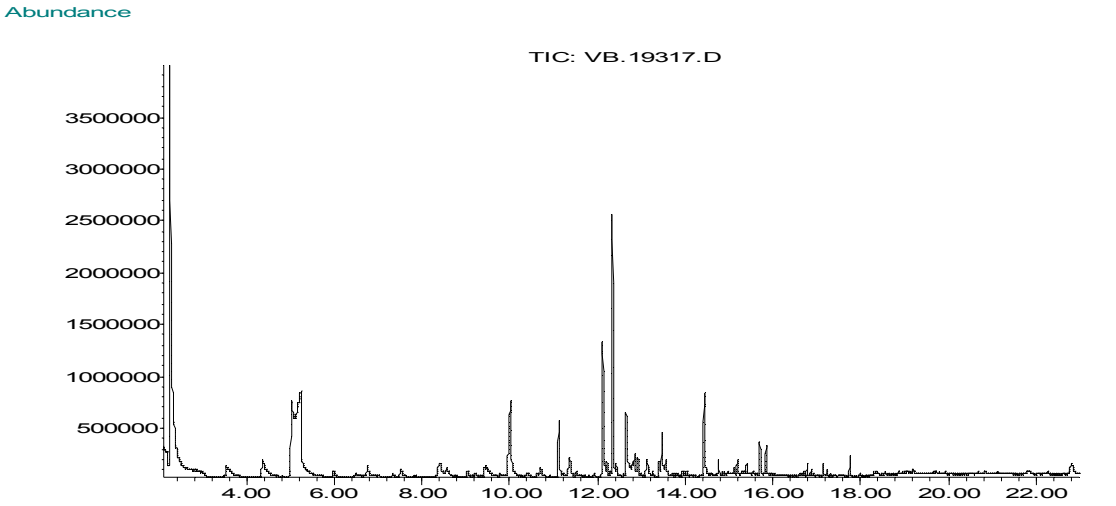

**1028**



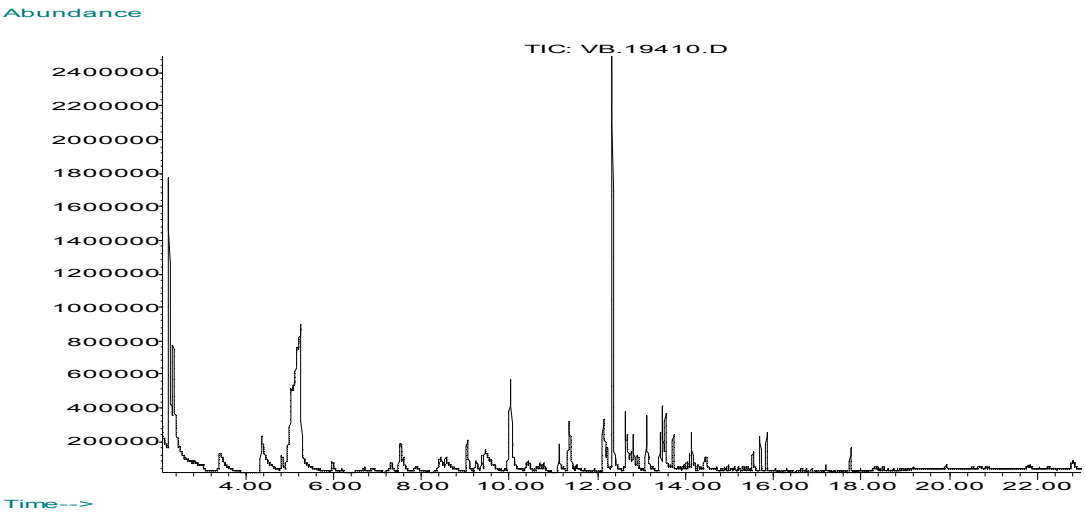

**1032**



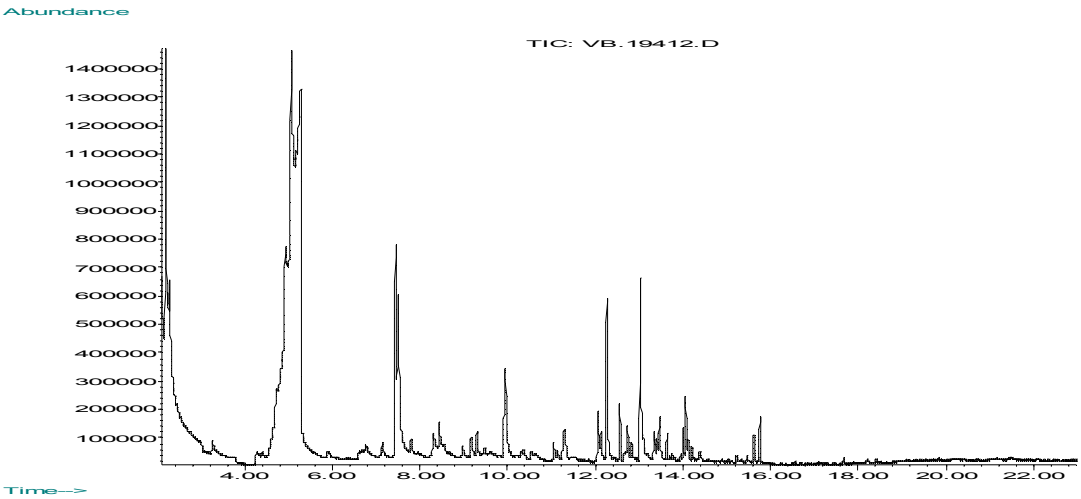

**1033**



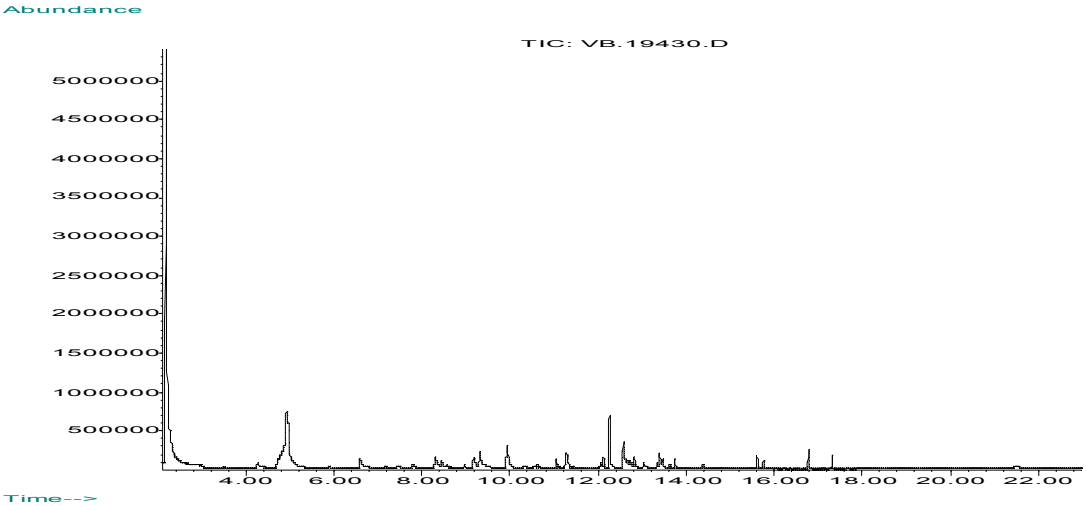

**19317**



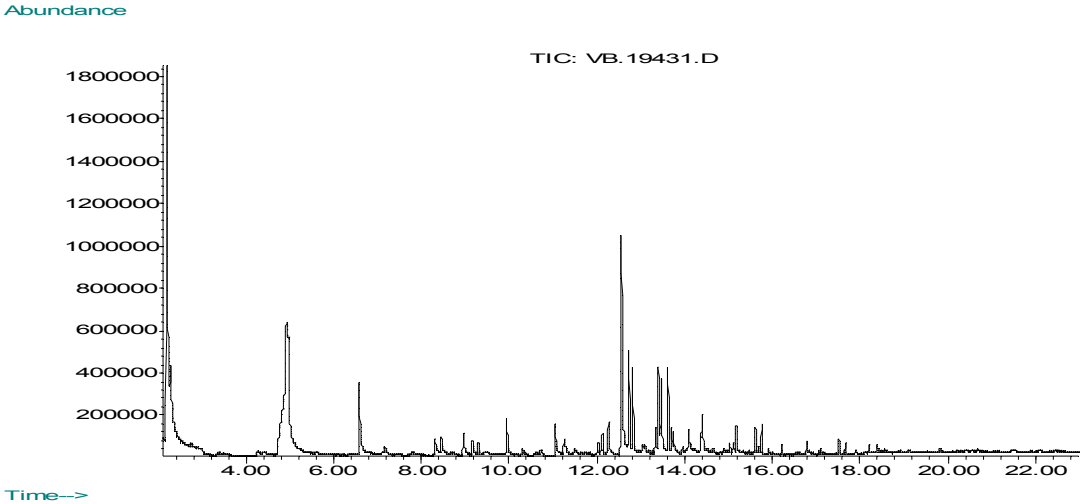

**19410**



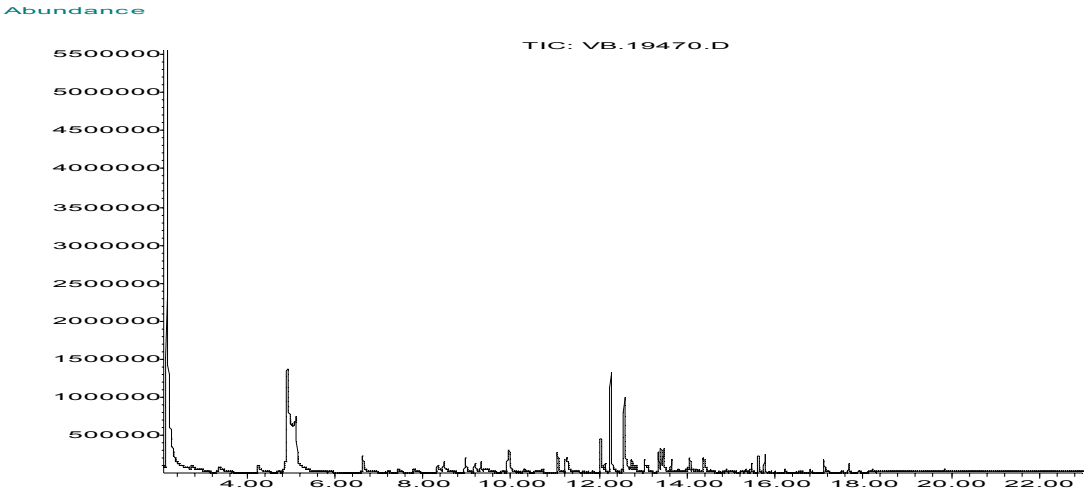

**19412**



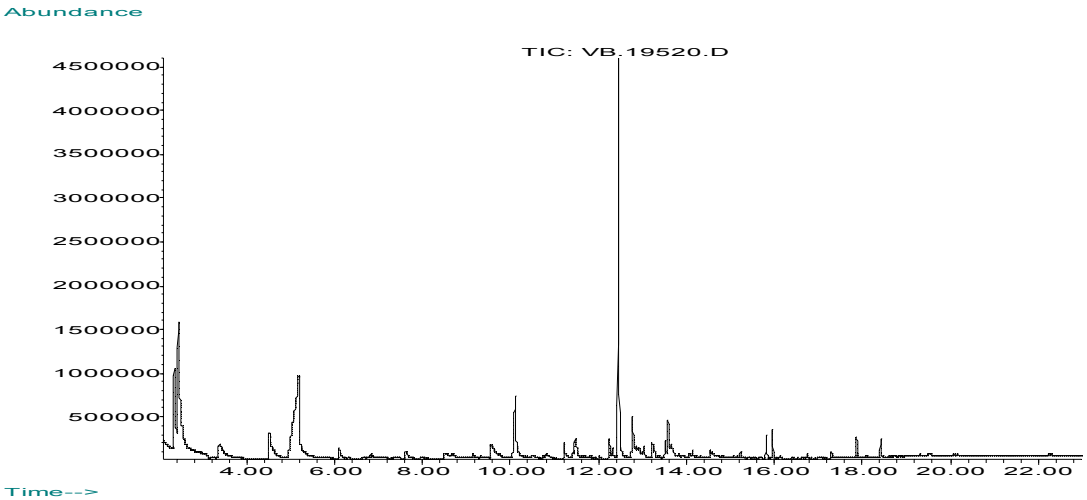

**19430**



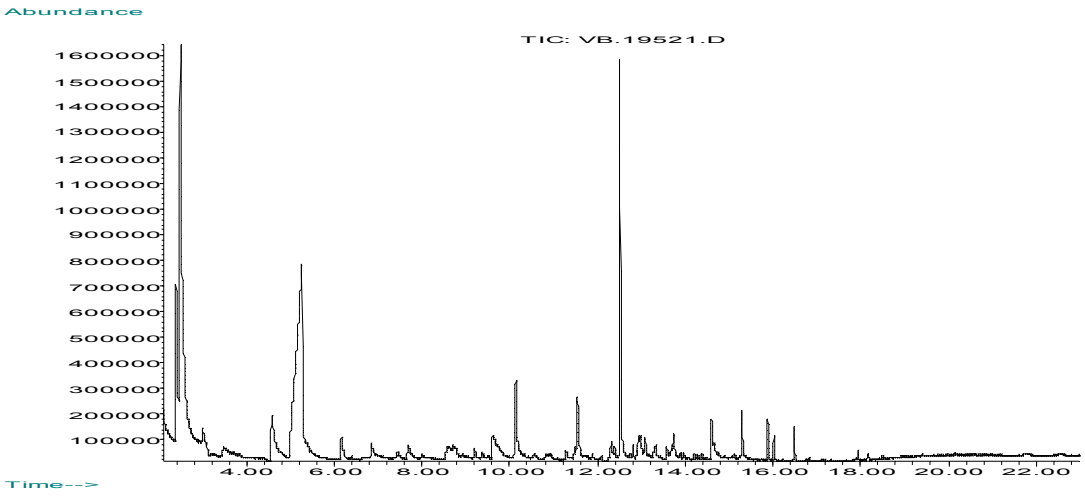

**19431**



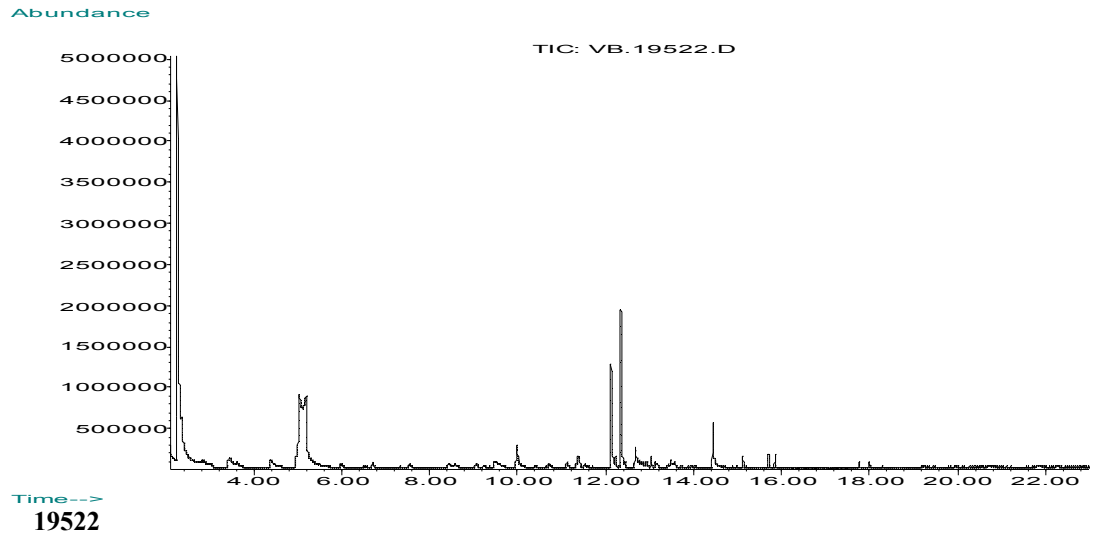

**19470**



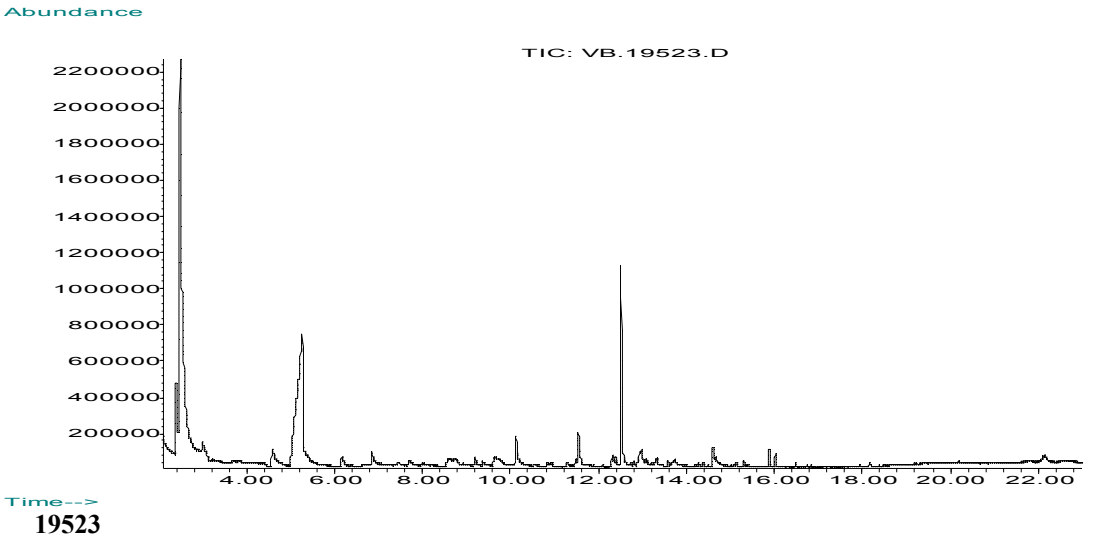

**19520**



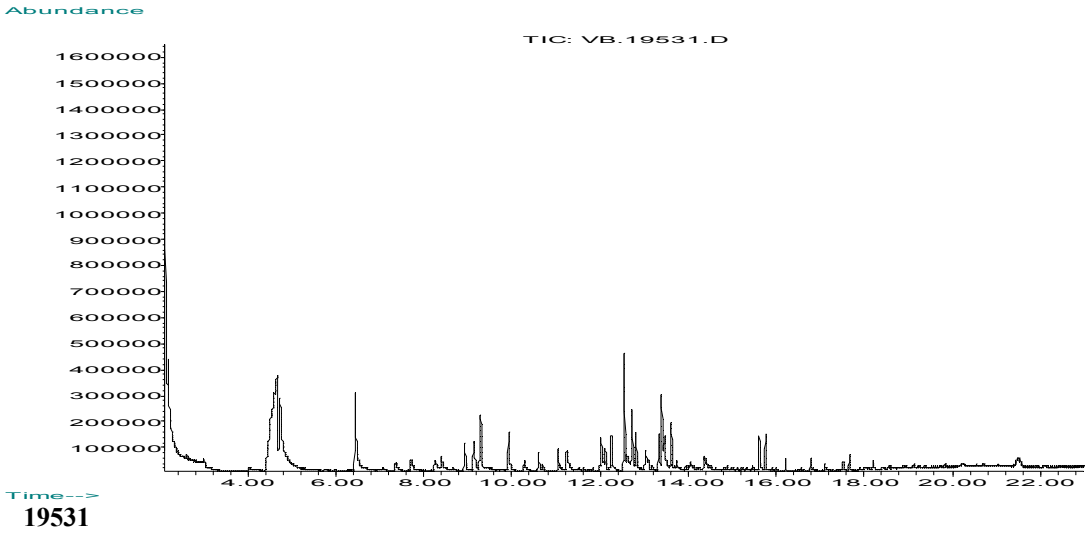

**19521**



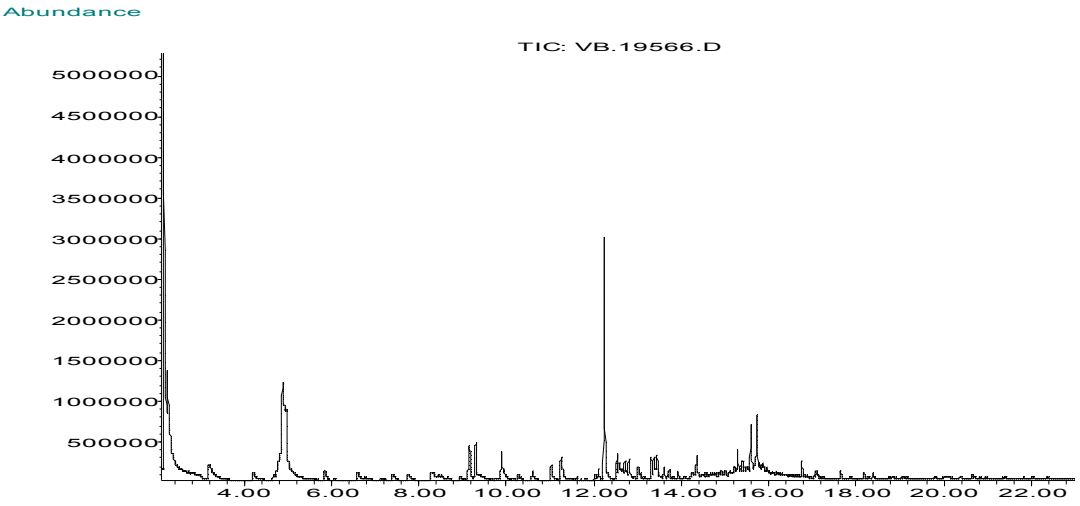

**19522**



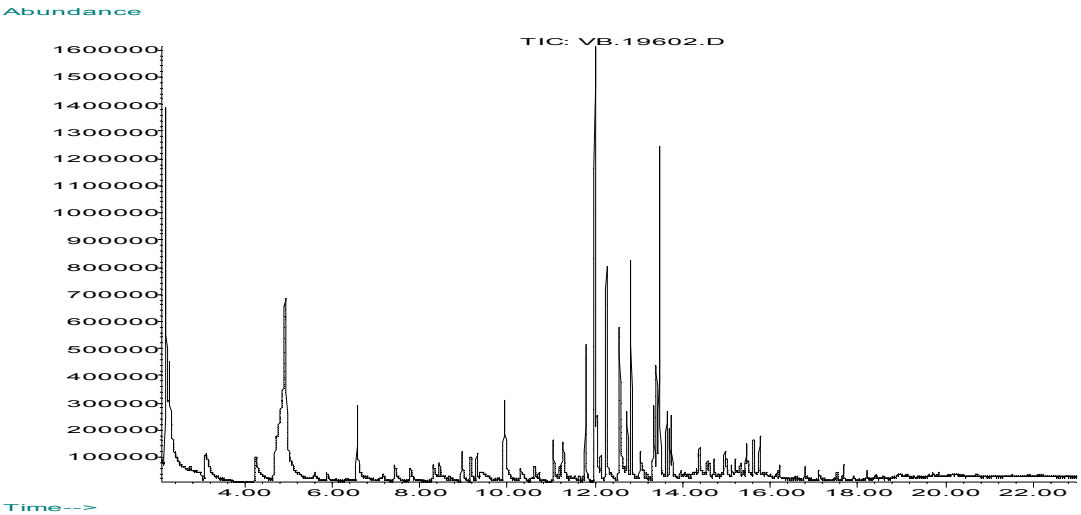

**19523**



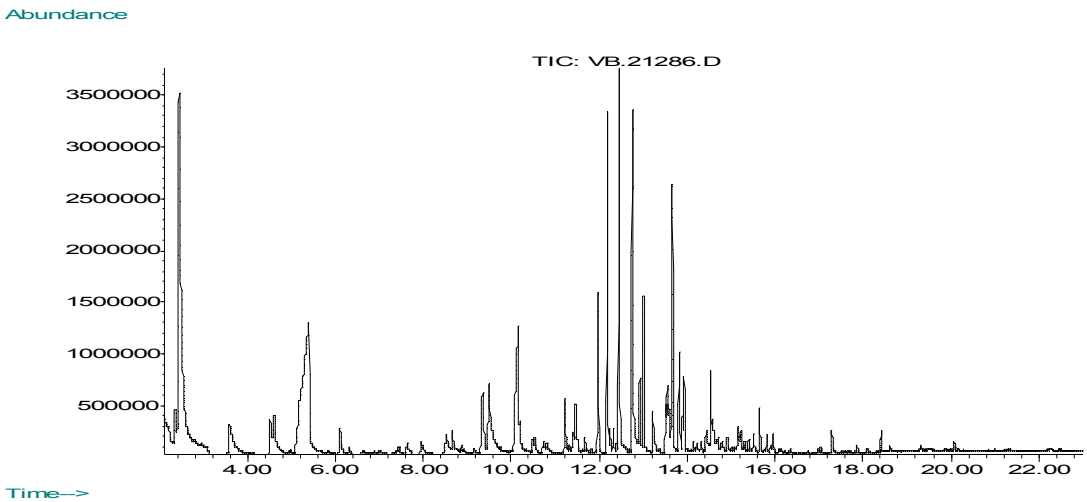

**19531**



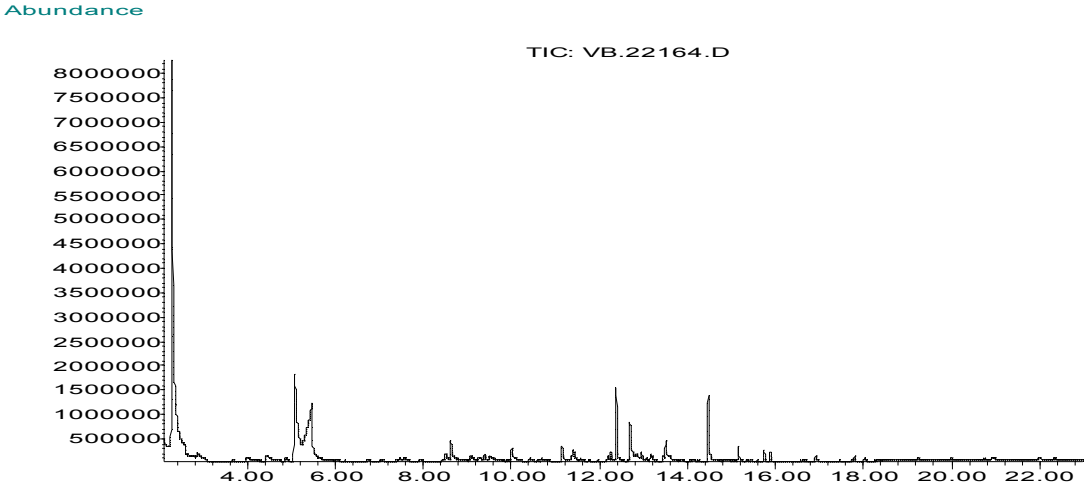

**19566**



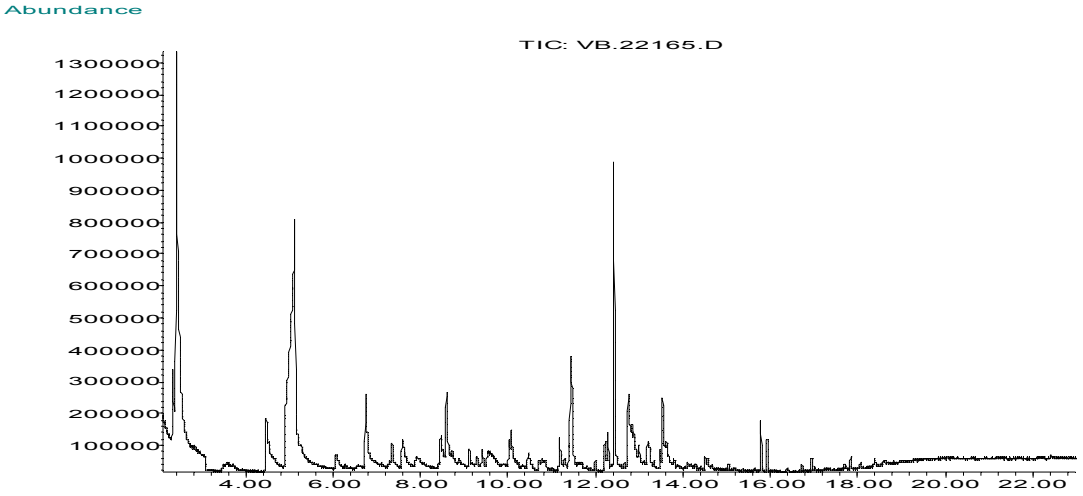

**19602**



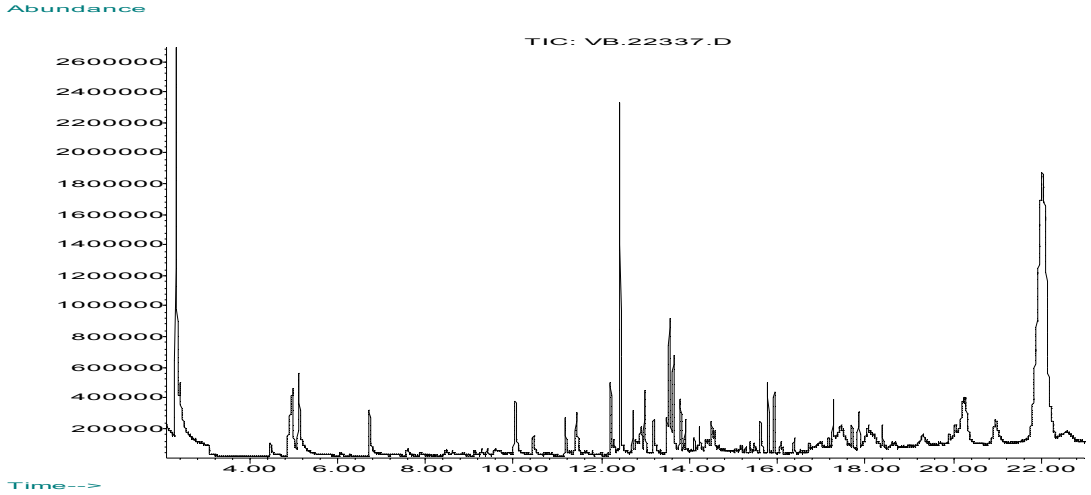

**21286**



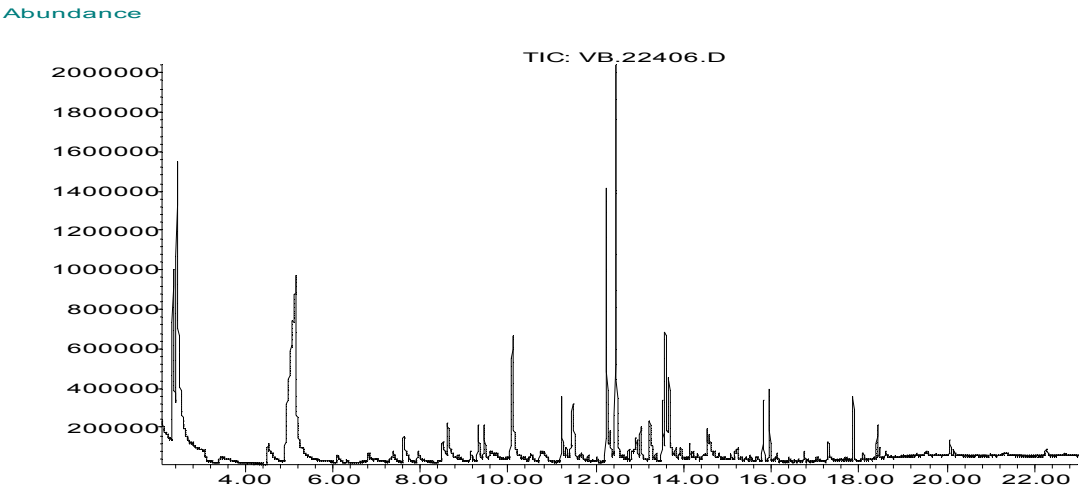

**22164**



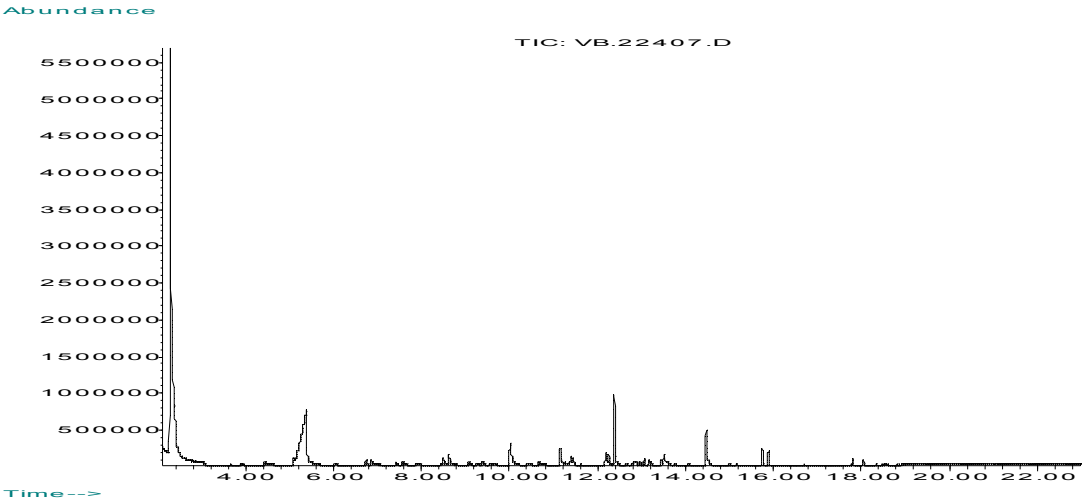

**22165**



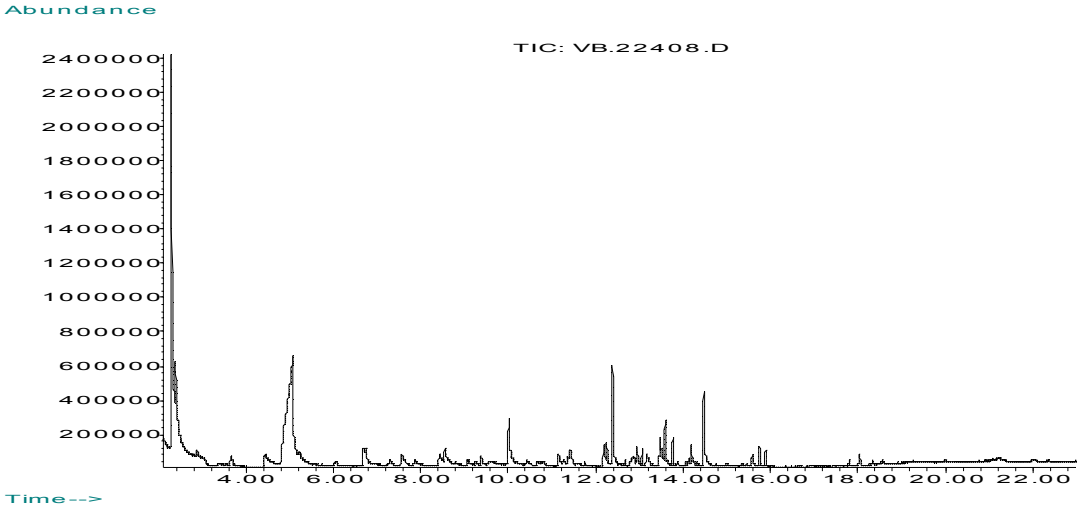

**22337**




**22406**

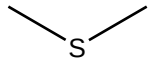
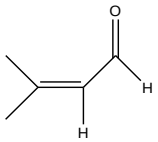
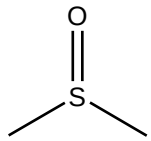
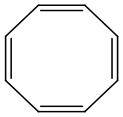
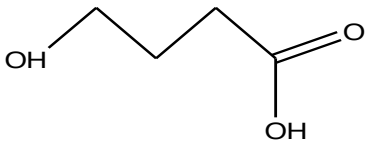
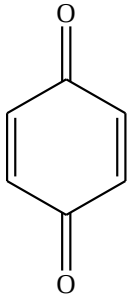
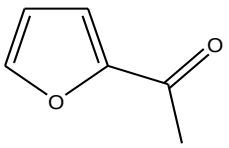
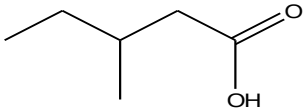


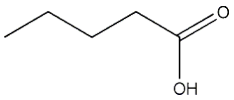
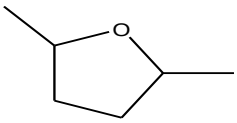
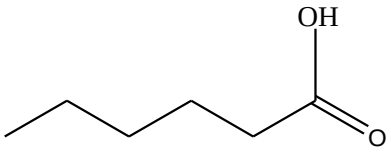
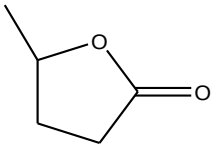
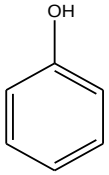
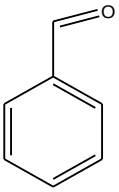
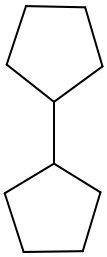
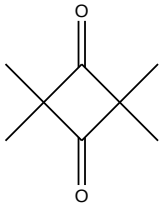
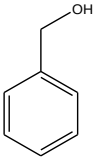

**22407**

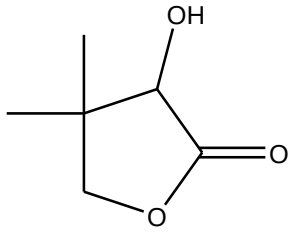
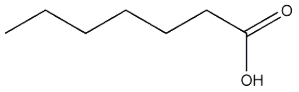
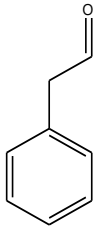
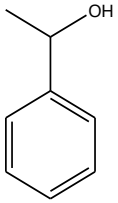
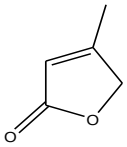
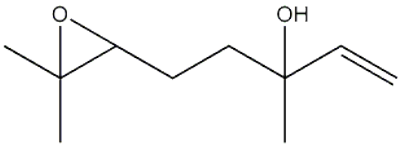
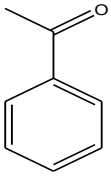
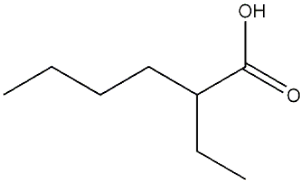


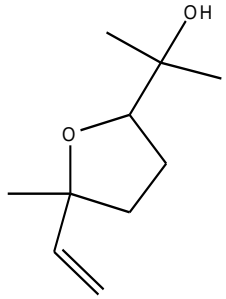
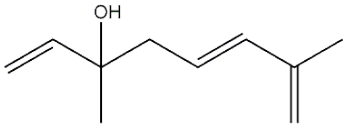
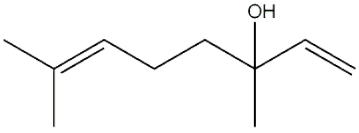
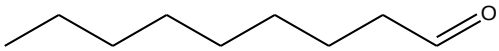
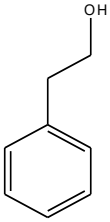
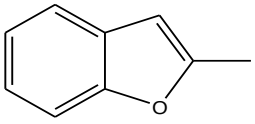
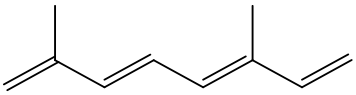
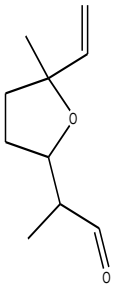
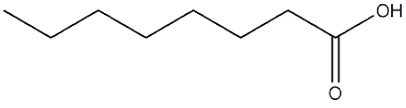

**22408**

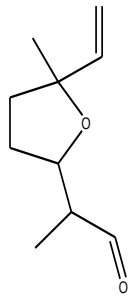
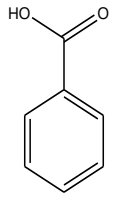
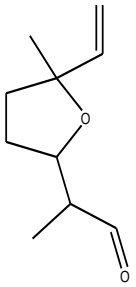
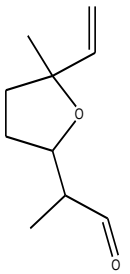
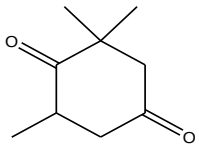
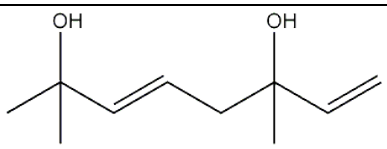
# Appendix 12: Compounds of Interest in Honey Samples

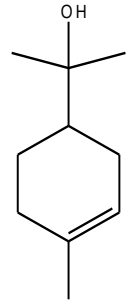
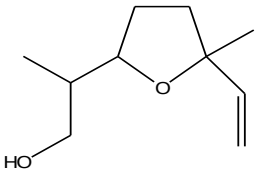
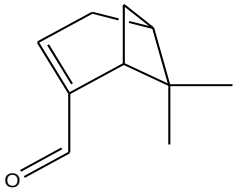
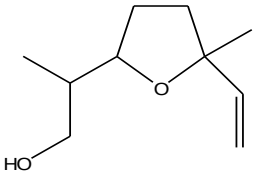
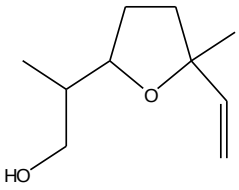
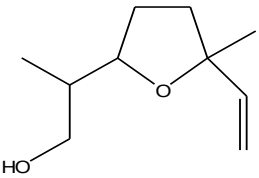
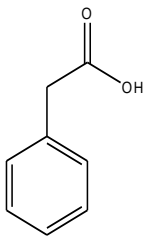
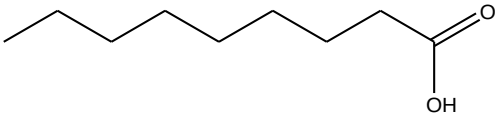
| **Name** | **RT (min)** | **Alternative name(s)** | **Structure** |
|---|---|---|---|
| Dimethyl sulfide | 2.69 | |  |
| 3-Methylbut-2-enal | 7.73 | |  |
| Dimethyl sulfoxide | 9.43 | |  |
| 1,3,5,7-Cyclooctatetraene | 10.49 | |  |
| 4-Hydroxybutanoic acid | 10.62 | |  |
| *p*-Benzoquinone | 10.63 | Cyclohexa-2,5-diene-1,4-dione |  |
| Furan-2-carbaldehyde | 10.67 | 1-(2-Furanyl)-ethanone |  |
| 3-Methylpentanoic acid | 10.72 | |  |

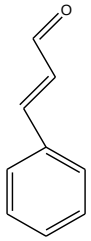
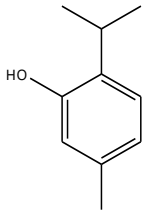
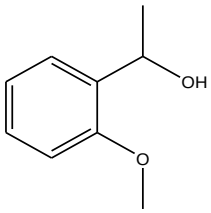
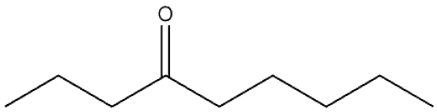
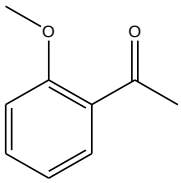
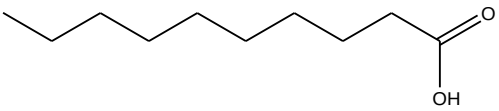
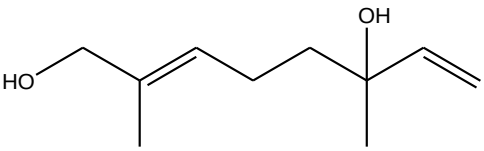
| | | | |
|---|---|---|---|
| Pentanoic acid | 11.22 | |  |
| Tetrahydro-2,5-dimethyl furan | 11.28 | 2,5-Dimethyloxolan |  |
| Hexanoic acid | 11.30 | |  |
| 5-Methyloxolan-2-one | 11.36 | Dihydro-4-methyl-2(3H)-furanone/ 4-Valerolactone |  |
| Phenol | 11.40 | |  |
| Benzaldehyde | 11.47 | |  |
| 1,1'-Bicyclopentyl | 11.97 | Cyclopentylcyclopentane |  |
| 2,2,4,4-Tetramethylcyclobutane-1,3-dione | 12.19 | |  |
| Benzyl alcohol | 12.24 | Phenylmethanol |  |

| | | | |
|---|---|---|---|
| Pantoyl lactone | 12.26 | 3-Hydroxy-4,4-dimethyloxolan-2-one/ Dihydro-3-hydroxy-4,4-dimethyl-2-(3H)-furanone |  |
| Heptanoic acid | 12.37 | |  |
| Benzeneacetaldehyde | 12.46 | 2-Phenylacetaldehyde |  |
| 1-Phenylethanol | 12.47 | Methylbenzenemeth-anol |  |
| 4-Methyl-5H-furan-2-one | 12.51 | |  |
| *cis*-Linaloloxide | 12.68 | 5-(3,3-Dimethyloxiran-2-yl)-3-methyl-pent-1-en-3-ol |  |
| Acetophenone | 12.79 | 1-Phenylethanone |  |
| 2-Ethylhexanoic acid | 12.89 | |  |

| | | | |
|---|---|---|---|
| Linalool oxide | 12.90 | 5-Ethenyltetrahydro-α,α-5-trimethyl-*cis*-2-furanmethanol/ 2-(5-Ethenyl-5-methyloxolan-2-yl)propan-2-ol |  |
| Hortrienol | 12.91 | 3,7-Dimethylocta-1,5,7-trien-3-ol |  |
| Linalool | 12.92 | 3,7-Dimethylocta-1,6-dien-3-ol |  |
| Nonanal | 13.03 | |  |
| 2-Phenylethanol | 13.23 | Phenylethyl alcohol |  |
| 2-Methylbenzofuran | 13.37 | |  |
| 2,6-Dimethyl-1,3,5,7-octatetraene | 13.44 | |  |
| Lilac aldehyde A | 13.48 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propanal |  |
| Octanoic acid | 13.52 | |  |

| | | | |
|---|---|---|---|
| Lilac aldehyde B | 13.55 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propanal |  |
| Benzoic acid | 13.58 | |  |
| Lilac aldehyde C | 13.60 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propanal |  |
| Lilac aldehyde D | 13.83 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propanal |  |
| 2,2,6-Trimethylcyclohexane-1,4-dione | 13.87 | |  |
| 2,6-Dimethyl-3,7-octadiene-2,6-diol | 13.94 | |  |

| | | | |
|---|---|---|---|
| α-Terpineol | 14.11 | 2-(4-Methyl-1-cyclohex-3-enyl)propan-2-ol/ α,α-4-Trimethyl-3-cyclohexene-1-methanol |  |
| Lilac alcohol B | 14.14 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propan-1-ol |  |
| Myrtenal | 14.24 | 6,6-Dimethyl-bicyclo[3.1.1]hept-2-ene-2-carboxaldehyde |  |
| Lilac alcohol C | 14.25 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propan-1-ol |  |
| Lilac alcohol A | 14.32 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propan-1-ol |  |
| Lilac alcohol D | 14.33 | 2-(5-Ethenyl-5-methyl-oxolan-2-yl)propan-1-ol |  |
| Benzeneacetic acid | 14.44 | Phenylacetic acid |  |
| Nonanoic acid | 14.54 | |  |

| | | | |
|---|---|---|---|
| (*E*)-Cinnamaldehyde | 14.86 | (*E*)-3-Phenylprop-2-enal |  |
| Thymol | 14.96 | 5-Methyl-2-propan-2-yl-phenol |  |
| 1-(2-Methoxyphenyl)ethanol | 15.00 | O-methoxy-α-methylbenzyl alcohol |  |
| Nonan-4-one | 15.16 | |  |
| *ortho*-Methoxyacetophenone | 15.23 | 1-(2-Methoxyphenyl)ethanone |  |
| Decanoic acid | 15.29 | |  |
| 2,6-Dimethyl-2,7-octadiene-1,6-diol | 15.43 | |  |

# Appendix 13: Summary of Selected Learning Algorithm Outputs

## Manually Selected data: AdaBoostM1 and J48 algorithms

- Correctly classified instances: 88.0%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
41  1  0  0  0  0  0  1  0  0 |  a = Clover
 1 24  0  1  0  0  0  0  0  0 |  b = Honeydew
 0  0 16  1  0  0  1  0  0  0 |  c = Kamahi
 0  0  1 26  0  0  1  0  0  0 |  d = Manuka
 0  0  0  0 15  1  0  1  0  0 |  e = Pohutukawa
 0  0  0  0  0 22  1  0  0  0 |  f = Rata
 0  0  0  0  0  0 14  2  0  0 |  g = Rewarewa
 5  1  1  1  0  0  1 19  0  0 |  h = Tawari
 1  0  0  0  0  0  0  0 11  0 |  i = Thyme
 4  0  0  0  1  0  0  0  0 18 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.953 | 0.058 | 0.788 | 0.953 | 0.863 | 0.978 | Clover |
| 0.923 | 0.01 | 0.923 | 0.923 | 0.923 | 0.997 | Honeydew |
| 0.889 | 0.009 | 0.889 | 0.889 | 0.889 | 0.998 | Kamahi |
| 0.929 | 0.015 | 0.897 | 0.929 | 0.912 | 0.998 | Manuka |
| 0.882 | 0.005 | 0.938 | 0.882 | 0.909 | 0.98 | Pohutukawa |
| 0.957 | 0.005 | 0.957 | 0.957 | 0.957 | 0.998 | Rata |
| 0.875 | 0.018 | 0.778 | 0.875 | 0.824 | 0.993 | Rewarewa |
| 0.679 | 0.019 | 0.826 | 0.679 | 0.745 | 0.932 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.966 | Thyme |
| 0.783 | 0 | 1 | 0.783 | 0.878 | 0.971 | Viper’s Bugloss |

# Manually selected data: AdaboostM1 and LMT algorithms

- Correctly classified instances: 88.0%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
37  1  0  0  0  0  0  0  1  4 |  a = Clover
 0 25  0  0  0  0  0  1  0  0 |  b = Honeydew
 0  0 17  0  0  0  1  0  0  0 |  c = Kamahi
 1  0  1 25  0  0  1  0  0  0 |  d = Manuka
 0  0  0  0 15  2  0  0  0  0 |  e = Pohutukawa
 0  0  0  0  0 22  0  0  0  1 |  f = Rata
 2  0  1  0  0  0 11  2  0  0 |  g = Rewarewa
 0  0  0  0  0  0  1 26  1  0 |  h = Tawari
 0  0  1  0  0  0  0  0 10  1 |  i = Thyme
 5  0  0  0  0  0  0  0  0 18 |  j = Viper’s Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.86 | 0.042 | 0.822 | 0.86 | 0.841 | 0.974 | Clover |
| 0.962 | 0.005 | 0.962 | 0.962 | 0.962 | 0.998 | Honeydew |
| 0.944 | 0.014 | 0.85 | 0.944 | 0.895 | 0.949 | Kamahi |
| 0.893 | 0 | 1 | 0.893 | 0.943 | 0.994 | Manuka |
| 0.882 | 0 | 1 | 0.882 | 0.938 | 0.972 | Pohutukawa |
| 0.957 | 0.009 | 0.917 | 0.957 | 0.936 | 0.996 | Rata |
| 0.688 | 0.014 | 0.786 | 0.688 | 0.733 | 0.974 | Rewarewa |
| 0.929 | 0.015 | 0.897 | 0.929 | 0.912 | 0.995 | Tawari |
| 0.833 | 0.009 | 0.833 | 0.833 | 0.833 | 0.979 | Thyme |
| 0.783 | 0.028 | 0.75 | 0.783 | 0.766 | 0.906 | Viper's Bugloss |

# Manually selected data: AdaBoostM1 and classification via regression algorithms

- Correctly classified instances: 89.3%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
40  0  0  0  0  0  0  1  0  2 |  a = Clover
 0 26  0  0  0  0  0  0  0  0 |  b = Honeydew
 0  0 16  0  0  1  1  0  0  0 |  c = Kamahi
 0  0  0 28  0  0  0  0  0  0 |  d = Manuka
 0  0  0  0 16  1  0  0  0  0 |  e = Pohutukawa
 0  0  0  0  0 20  0  2  0  1 |  f = Rata
 0  0  0  0  0  1 14  1  0  0 |  g = Rewarewa
 2  2  0  1  0  1  0 21  0  1 |  h = Tawari
 0  0  0  0  0  0  0  1 11  0 |  i = Thyme
 6  0  0  0  0  0  0  0  0 17 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.93 | 0.042 | 0.833 | 0.93 | 0.879 | 0.979 | Clover |
| 1 | 0.01 | 0.929 | 1 | 0.963 | 0.998 | Honeydew |
| 0.889 | 0 | 1 | 0.889 | 0.941 | 0.999 | Kamahi |
| 1 | 0.005 | 0.966 | 1 | 0.982 | 0.999 | Manuka |
| 0.941 | 0 | 1 | 0.941 | 0.97 | 0.944 | Pohutukawa |
| 0.87 | 0.019 | 0.833 | 0.87 | 0.851 | 0.986 | Rata |
| 0.875 | 0.005 | 0.933 | 0.875 | 0.903 | 0.949 | Rewarewa |
| 0.75 | 0.024 | 0.808 | 0.75 | 0.778 | 0.961 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.968 | Thyme |
| 0.739 | 0.019 | 0.81 | 0.739 | 0.773 | 0.953 | Viper's Bugloss |

# Probability plot data: LMT algorithm

- Correctly classified instances: 88.5%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
39  1  0  1  0  0  0  2  0  0 |  a = Clover
 1 24  0  0  0  0  0  1  0  0 |  b = Honeydew
 0  0 17  0  0  0  1  0  0  0 |  c = Kamahi
 0  0  0 26  0  0  1  1  0  0 |  d = Manuka
 0  0  0  0 15  1  1  0  0  0 |  e = Pohutukawa
 0  0  1  0  0 22  0  0  0  0 |  f = Rata
 0  0  1  0  0  0 13  2  0  0 |  g = Rewarewa
 1  1  0  0  0  0  5 21  0  0 |  h = Tawari
 1  0  0  0  0  0  0  0 11  0 |  i = Thyme
 4  0  0  0  0  0  0  0  0 19 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.907 | 0.037 | 0.848 | 0.907 | 0.876 | 0.978 | Clover |
| 0.923 | 0.01 | 0.923 | 0.923 | 0.923 | 0.997 | Honeydew |
| 0.944 | 0.009 | 0.895 | 0.944 | 0.919 | 0.993 | Kamahi |
| 0.929 | 0.005 | 0.963 | 0.929 | 0.945 | 0.998 | Manuka |
| 0.882 | 0 | 1 | 0.882 | 0.938 | 0.999 | Pohutukawa |
| 0.957 | 0.005 | 0.957 | 0.957 | 0.957 | 0.992 | Rata |
| 0.813 | 0.037 | 0.619 | 0.813 | 0.703 | 0.921 | Rewarewa |
| 0.75 | 0.029 | 0.778 | 0.75 | 0.764 | 0.946 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.984 | Thyme |
| 0.826 | 0 | 1 | 0.826 | 0.905 | 0.975 | Viper's Bugloss |

# Probability plot data: Classification via regression algorithm

- Correctly classified instances: 88.0%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
39  0  0  0  0  0  0  4  0  0 |  a = Clover
 0 25  0  0  0  0  0  1  0  0 |  b = Honeydew
 0  0 18  0  0  0  0  0  0  0 |  c = Kamahi
 0  0  0 27  0  0  0  1  0  0 |  d = Manuka
 0  0  0  0 15  2  0  0  0  0 |  e = Pohutukawa
 1  0  0  0  0 20  1  1  0  0 |  f = Rata
 0  0  1  0  0  2 11  2  0  0 |  g = Rewarewa
 3  0  0  0  0  3  1 21  0  0 |  h = Tawari
 1  0  0  0  0  0  0  0 11  0 |  i = Thyme
 4  0  0  0  0  0  0  0  0 19 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.907 | 0.047 | 0.813 | 0.907 | 0.857 | 0.973 | Clover |
| 0.962 | 0 | 1 | 0.962 | 0.98 | 0.999 | Honeydew |
| 1 | 0.005 | 0.947 | 1 | 0.973 | 0.999 | Kamahi |
| 0.964 | 0 | 1 | 0.964 | 0.982 | 0.999 | Manuka |
| 0.882 | 0 | 1 | 0.882 | 0.938 | 0.994 | Pohutukawa |
| 0.87 | 0.033 | 0.741 | 0.87 | 0.8 | 0.985 | Rata |
| 0.688 | 0.009 | 0.846 | 0.688 | 0.759 | 0.951 | Rewarewa |
| 0.75 | 0.044 | 0.7 | 0.75 | 0.724 | 0.921 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.923 | Thyme |
| 0.826 | 0 | 1 | 0.826 | 0.905 | 0.941 | Viper's Bugloss |

# Probability plot data: AdaBoostM1 and J48 algorithms

- Correctly classified instances: 88.5%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
37  1  0  0  0  0  1  1  0  3 |  a = Clover
 1 25  0  0  0  0  0  0  0  0 |  b = Honeydew
 0  0 17  0  0  0  1  0  0  0 |  c = Kamahi
 0  0  0 27  0  0  1  0  0  0 |  d = Manuka
 0  0  1  0 13  3  0  0  0  0 |  e = Pohutukawa
 0  0  0  0  0 22  0  1  0  0 |  f = Rata
 0  0  0  0  0  0 14  2  0  0 |  g = Rewarewa
 3  1  0  1  0  0  1 21  0  1 |  h = Tawari
 1  0  0  0  0  0  0  0 11  0 |  i = Thyme
 2  0  0  0  0  1  0  0  0 20 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.86 | 0.037 | 0.841 | 0.86 | 0.851 | 0.973 | Clover |
| 0.962 | 0.01 | 0.926 | 0.962 | 0.943 | 0.999 | Honeydew |
| 0.944 | 0.005 | 0.944 | 0.944 | 0.944 | 0.999 | Kamahi |
| 0.964 | 0.005 | 0.964 | 0.964 | 0.964 | 0.999 | Manuka |
| 0.765 | 0 | 1 | 0.765 | 0.867 | 0.979 | Pohutukawa |
| 0.957 | 0.019 | 0.846 | 0.957 | 0.898 | 0.992 | Rata |
| 0.875 | 0.018 | 0.778 | 0.875 | 0.824 | 0.986 | Rewarewa |
| 0.75 | 0.019 | 0.84 | 0.75 | 0.792 | 0.97 | Tawari |
| 0.917 | 0 | 1 | 0.917 | 0.957 | 0.979 | Thyme |
| 0.87 | 0.019 | 0.833 | 0.87 | 0.851 | 0.959 | Viper's Bugloss |

# Probability plot data: AdaBoostM1 and LMT algorithms

- Correctly classified instances: 89.7%

- Confusion Matrix

```
 a  b  c  d  e  f  g  h  i  j   <-- classified as
36  0  0  0  0  0  0  4  2  1 |  a = Clover
 1 25  0  0  0  0  0  0  0  0 |  b = Honeydew
 0  0 17  0  0  0  1  0  0  0 |  c = Kamahi
 0  0  1 26  0  0  0  1  0  0 |  d = Manuka
 0  0  0  0 15  2  0  0  0  0 |  e = Pohutukawa
 0  0  0  0  0 22  0  1  0  0 |  f = Rata
 1  0  1  0  0  0 13  1  0  0 |  g = Rewarewa
 2  0  0  0  0  0  1 25  0  0 |  h = Tawari
 0  0  0  0  0  0  0  0 11  1 |  i = Thyme
 1  0  0  0  0  1  0  1  0 20 |  j = Viper's Bugloss
```

- Evaluation Statistics

| TP rate | FP rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|
| 0.837 | 0.026 | 0.878 | 0.837 | 0.857 | 0.977 | Clover |
| 0.962 | 0 | 1 | 0.962 | 0.98 | 1 | Honeydew |
| 0.944 | 0.009 | 0.895 | 0.944 | 0.919 | 0.996 | Kamahi |
| 0.929 | 0 | 1 | 0.929 | 0.963 | 0.996 | Manuka |
| 0.882 | 0 | 1 | 0.882 | 0.938 | 0.961 | Pohutukawa |
| 0.957 | 0.014 | 0.88 | 0.957 | 0.917 | 0.992 | Rata |
| 0.813 | 0.009 | 0.867 | 0.813 | 0.839 | 0.987 | Rewarewa |
| 0.893 | 0.039 | 0.758 | 0.893 | 0.82 | 0.961 | Tawari |
| 0.917 | 0.009 | 0.846 | 0.917 | 0.88 | 0.974 | Thyme |
| 0.87 | 0.009 | 0.909 | 0.87 | 0.889 | 0.981 | Viper’s Bugloss |