Metabolomic studies of plant response to environmental cues and phytochemical analysis of medicinal plants using NMR spectroscopy

Sumit Mishra

A thesis submitted for the partial fulfillment of
the degree of Doctor of Philosophy

Department of Physical Sciences
Indian Institute of Science Education & Research Mohali
Knowledge city, Sector 81, SAS Nagar, Manauli PO, Mohali 140306, Punjab, India

April 2023

Declaration

The work presented in this thesis has been carried out by me under the guidance of Prof. Kavita Dorai at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, diploma or a fellowship to any other University or Institute. Whenever contributions of others are involved, every effort has been made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Sumit Mishra

Place :
Date :

In my capacity as supervisor of the candidate's PhD thesis work, I certify that the above statements by the candidate are true to the best of my knowledge.

Prof. Kavita Dorai
(Supervisor)
Professor
Department of Physical Sciences
IISER Mohali

Place :
Date :

Acknowledgments

Firstly, I would like to thank my supervisor Prof. Kavita Dorai for giving me the opportunity of starting a career in Research, for her supervision throughout this work, all the fruitful discussions and for her help, not only in Research matters but also providing her support with personal issues. I must admit that without her help and support this thesis would not have been possible.

I would like to thank my doctoral committee members Prof. Arvind, Prof. N.G. Prasad and Prof. Abhishek Chaudhuri for their useful advices. I am grateful to director IISER Mohali: Prof. J. Gowrishankar (Present Director), Prof. Arvind (Acting Director), Prof. Debi P. Sarkar (Former Director) and Prof. N. Sathyamurthy (Former Director) for providing all kind of help needed for my research work. I am thankful to the faculty of the departnent of physical sciences of IISER Mohali for providing their excellent guidance during my course work. I wish to express the warmest of thanks to Dr. Paramdeep Chandi and Balbir Singh Ji (Scientific Staff) for support during my Ph.D.

I would especially like to thank my parents, Ram Nivas sharma and Brijesh Sharma, as well as my brother Sunil Mishra, my bhavi Reena Mishra, and my nieces Garima Mishra and Divya Mishra. Thanks for their support and patience. This piece of work, is dedicated to my younger brother Sachin Mishra, who is no longer with us. I would like to thank my friend, Jyoti Lather. She was always there, in good and bad times. All your help when I was writing my thesis and the possibility to talk to you were very helpful. You really helped me through this adventure.

I also want to thank Dr. Nisha & Dr. Amandeep for support & motivation and specially all the cups of tea we had the last years. I would like to thank my friends Dr. Anita, Dr. Pushpendra, Dr. Munish, Dr. Manvendra, Dr. Samita, Akhila, Pallavi, Jaiveer, Dr. Shyam, Dr. Ramu, Dr. Shelendra, Dr. Ritesh, Dr. Anirban, Dr. Anzar, Vivek, Dr. Namrata, Dr. Aparajita, Gyanendra, Dr. Anubhav, Prabhat, Dr. Sumeeta, Dr. Gunjan. I would like to acknowledge my current labmates: Dr. Rajbinder kaur virk, Dr. Jyotsana, Akshay, Dr. Dileep, Akanksha, Vaishali, Krishna, Gayatri, Arshdeep and previous labmates: Dr. Navdeep, Dr. Amandeep, Dr. Rakesh, Omkar, Saptarishi, Himanshu, Divyanshi, Spandana, Ankit, Guru, Raj, Riyaz, Shivanshu, Saksham, Safana, Abhay, Fathima. I would like to thanks for funding provided by IISER Mohali (JRF-SRF fellowship), EUROISMAR-2019 (registrations fees waived off), CSIR-2022 (Travel grant), NMRS-2022 (Travel grant), AMPERE NMR-2022 (registrations fees waived off), STARS Project JRF-SRF fellowship during my Ph.D. work.

Sumit Mishra

Abstract

Metabolomics is the large-scale study of small molecules in cells, biofluids, tissues, or organisms. These small molecules are called metabolites. The most prevalent approaches in the field of metabolomics are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). These approaches are used to investigate the metabolic profile of a living organism. NMR spectroscopy is primarily used in therapeutics, traditional medicine, environmental monitoring, the diagnosis of human disease, and food quality management.

This thesis discusses the applications of NMR-based metabolomics analysis to provide a metabolic profile of phytomedicinal compounds in Momordica Charantia (Bitter melon), Phyllanthus emblica (Amla), and Tinospora cordifolia (Giloy). Various portions of the medicinal fruits have their own unique metabolic profile of beneficial minerals, proteins, and vitamins, in addition to a wide variety of vital bioactive chemicals. The later part of the thesis focuses on the applications of NMR-based metabolomics to study the plant metabolome. These studies include several elements of plant physiology, such as circadian rhythm, stress metabolism, and plants signaling. The last part of the thesis highlights the importance of NMR-based metabolomics in understanding several facets of insect physiology, including the evolution of insect immune response. The main objective is to obtain an understanding of the metabolic pathway that ultimately result in a series of chemical reactions occurring within a cell. The contents of the chapters of the thesis are briefly described below.

Chapter 1

The introductory chapter briefly introduces the basics of NMR spectroscopy, the various analytical techniques used for metabolomics and the role of metabolomics in system biology. The chapter also provides details about NMR pulse sequences, preprocessing of raw spectral data, univariate and multivariate statistical methods used in metabolomics.

Chapter 2

This chapter discusses the use of NMR in the analysis of phytomedicinal compounds identified in the pericarp, skin, and seeds of Momordica charantia (bitter melon). The Momordica charantia plant has a long history of usage as a medicinal plant, and has been included in a variety of different traditional medical systems. A multivariate statistical analysis revealed that the metabolic profiles of the seeds and pericarp were grouped together, and that these two groups were clearly differentiated from the metabolic profile of the skin part of Momordica charantia. A few phytosterols such as charantin and momordicine were identified which are known to correlate with antidiabetic action.

Chapter 3

This chapter details the metabolomic profiles of phytomedicinal compounds in Phyllanthus emblica and Tinospora cordifolia using one- and two-dimensional NMR spectroscopy. Phyllanthus emblica is an Ayurvedic herb which is used as a medication and tonic to restore vitality. Tinospora cordifolia has long been recognized as an important herb in traditional Ayurvedic medicine, where it has been utilized for the treatment of a wide range of conditions, including fever, diarrhoea, cancer and eye disorders. In this study, ^{1}H NMR spectroscopy coupled with multivariate statistical data analysis such as principal component analysis (PCA) was applied to investigate metabolomic variations among Phyllanthus emblica raw fruits and juice ((procured from Patanjali Ayurveda). The assignments of primary and secondary metabolites in Phyllanthus emblica fruit and Tinospora cordifolia stem have been carried out making use of ^{1}H and 2D NMR (^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC) spectroscopy.

Chapter 4

This chapter focuses on the utility of NMR-based metabolomics for metabolite profiling. The in vivo metabolic profile was constructed from the stems of young Helianthus annuus L. (sunflower) during the circadian cycle. This metabolic profile reveals metabolites and the functional periodicity of the circadian rhythm that they exhibit. One- and two-dimensional NMR spectroscopy methods were used to profile various metabolites, including sugars, amino acids, and tri-carboxylic acid (TCA) intermediates, of the immature sunflower stems.

Chapter 5

This chapter highlights the utility of NMR-based methods in studying plant metabolite response to air pollution stress. We used one- and two-dimensional NMR spectroscopy to perform metabolite fingerprinting on the leaves of Bougainvillea spectabilis, which is a plant that is known to be tolerant to different types of abiotic stresses. It was observed a few metabolites exhibit a consistent rhythmic pattern throughout the circadian cycle, which is evidence that circadian rhythms have a significant impact on the metabolic processes that occur in plants.

Chapter 6

This chapter emphasizes the utility of NMR-based metabolomics to study plant response against wounding stress. NMR metabolomic methods were used to study the metabolic alterations that occurred in the leaves of the Bougainvillea spectabilis plant following a wounding treatment that was designed to simulate a herbivore attack. It was observed that the content of amino acids, carbohydrates, and a few secondary metabolites varied in wounded plants.

Chapter 7

This chapter describes the application of NMR-based metabolomics approaches to the study of evolution in the fruit fly Drosophila melanogaster at different ages. Metabolic variations between a fly population chosen for increased immunity and the control population, reveal that populations develop diverse metabolomes in response to selection at three different ages namely young (5 days old), middle (20 days old) and old (35 days old) aged flies. Also, the metabolic response to prick injury/infection was examined between immune-selected and control populations. It was noted that the immune selected population continues to be able to mount a stronger immune response to prick injury/infection even after it matures in age.

Chapter 8

In the last chapter, a summary of all the studies that were carried out and the conclusions that were made from those research is presented. In addition to this, the chapter contains some observations about the role that NMR-based metabolomics will play in different disciplines.

List of Publications

1. **Mishra, S.**, Gogna, N., & Dorai, K. (2019). NMR-based investigation of the altered metabolic response of Bougainvillea spectabilis leaves exposed to air pollution stress during the circadian cycle. Environmental and Experimental Botany, 164, 58-70, **https://doi.org/10.1016/j.envexpbot.2019.04.019.**

2. **Mishra, S.**, Ankit, Sharma, R., Gogna, N., & Dorai, K. (2021). NMR-based metabolomic profiling of the differential concentration of phytomedicinal compounds in pericarp, skin and seeds of Momordica charantia (bitter melon). Natural Product Research, 36(1), 390-395, **https://doi.org/10.1080/14786419.2020.1762190.**

3. Deepika, V. B., Vohra, M., **Mishra, S.**, Dorai, K., Rai, P., Satyamoorthy, K., & Murali, T. S. (2020). DNA demethylation overcomes attenuation of colchicine biosynthesis in an endophytic fungus Diaporthe. Journal of Biotechnology, 323, 33-41, **https://doi.org/10.1016/j.jbiotec.2020.07.019.**

4. **Mishra, S.**, & Dorai, K. NMR metabolomic profiling of phytomedicinal compounds in Phyllanthus emblica and Tinospora cordifolia. (Manuscript under preparation).

5. **Mishra, S.**, Gogna N. & Dorai, K. NMR-based investigation of cycling metabolites in Helianthus annuus L. (sunflower) stems during circadian cycle. (Manuscript under preparation).

6. **Mishra, S.**, & Dorai, K. NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to wounding stress. (Manuscript under preparation).

7. **Mishra, S.**, Gogna N., Spandana B. S., Ahlawat N., Dorai K., & Prasad N. G. Studying immunity across ages in a population of Drosophila melanogaster evolved for increased immunity using NMR-based metabolomics. (Manuscript under preparation).

Contents

List of Figures

List of Tables

Abbreviations used in the Thesis

NMR	:	Nuclear Magnetic Resonance
MS	:	Mass spectroscopy
COSY	:	Correlation spectroscopy
TOCSY	:	Total correlation spectroscopy
HSQC	:	Heteronuclear coherence spectroscopy
CPMG	:	Carr-Purcell-Meiboom-Gill
CD_3OD	:	Methanol-d4
D_2O	:	Deuterium oxide
TMSP	:	Trimethylsilane propionic acid sodium salt
1D	:	One dimensional
2D	:	Two dimensional
ANOVA	:	One-way analysis of variance
PCA	:	Principal component analysis
PLS-DA	:	Partial least squares-discriminant analysis
OPLS-DA	:	Orthogonal projections to latent structures discriminant analysis
FDR	:	False discovery rate
DPPH	:	2,2-diphenyl-1-picrylhydrazyl
ABTS+	:	2,2′-azino- bis(3-ethylbenzothiazoline-6-sulphonic acid) diamonium salt
Trolox	:	6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
p-NPG	:	p-Nitrophenyl-α-D-glucopyranoside
UPLC-MS	:	Ultra-performance liquid chromatography-mass spectrometry
ESI	:	Electrospray ionization
QXI	:	Quadrupolar resonance probe
mg QE	:	milligrams of quercetin equivalent
mg GAE	:	milligrams of gallic acid equivalent

Chapter 1

Introduction

1.1 Basics of NMR Spectroscopy

Nuclear magnetic resonance spectroscopy, more often known as NMR, has emerged as the method of choice for evaluating the structure of natural products [1], study of tautomerism [2], understanding of disease [3], plants physiology [4], environmental sciences [5] and in drug discovery [6]. NMR and Mass spectrometry (MS) - based metabolomics [7], [8], [9], [10], [11], [12], [13] are the most commonly used techniques to examine a metabolic profile within cells, biofluids, tissues or organisms. Both of these approaches come with their own set of benefits and drawbacks [14]. NMR spectrometry is a non-destructive [15], highly reproducible technique that can measure hundreds of metabolites. MS is more sensitive [15] and can identify more metabolites in a sample, but it requires various chromatography techniques for different types of metabolites.
NMR is a spectroscopic technique based on the magnetic properties of nuclei. Rabi made the initial discovery that a magnetic field has an influence on the nuclei by performing an experiment that involved passing a stream of hydrogen atoms through a homogeneous magnetic field while also exposing them to a radiofrequency field [16]. It was discovered that a specific frequency of energy was absorbed by the molecules, resulting in a small but observable deflection of the beam [16]. This was the first NMR observation, and Rabi was awarded the Nobel Prize in 1944 for his role in the invention and advancement of the technology. In 1946, Bloch, Hansen, and Packard at Stanford [17] and Purcell, Torrey, and Pound at Harvard [18] published the first reports on NMR in bulk materials. In 1952, Bloch and Purcell awarded the Physics Nobel Prize, demonstrating the significance of their discovery. In 1966, Ernst and a colleague found that NMR spectroscopy could be improved by using short, strong pulses instead of slow sweeping radio waves [19]. The major innovation occurred when Nagayama,

Wuthrich, Bachmann, and Ernst demonstrated that biopolymers could be studied using 2D NMR approaches [20]. In 1973, Lauterbur [21] demonstrated that 2D images could be generated by applying magnetic field gradients across a sample, kicking off a different line of development that would have far-reaching implications for in vivo NMR spectroscopy.

When a nuclear spin with a magnetic dipole moment μ is placed simultaneously in the presence of a static magnetic field B_0 and an electromagnetic field oscillating with the appropriate frequency, resonant absorption/emission can occur in the radio frequency range as shown in Fig 1.1 (A). This phenomena, known as magnetic resonance, can be seen in a variety of closely related approaches, including electron spin resonance (ESR), nuclear magnetic resonance (NMR), ferromagnetic resonance (FMR), and nuclear quadrupole resonance (NQR) spectroscopy [22]. Nuclear magnetic resonance can be observed in general for nuclei with non-vanishing total angular momentum [23].

The nuclear total angular momentum is commonly known to as nuclear spin. In the formalism of quantum mechanics, it is a vector operator that is usually represented by $\hbar\mathbf{I}$, where $\mathbf{I}$ is a dimensionless operator that stands for the total angular momentum of the nucleus. In quantum mechanics, the nuclear spin possesses the same general properties as any other angular momentum operator; it is defined by a quantum number $\mathbf{I}$, which is also known to as the nuclear spin quantum number. In addition to nuclear spin, the nuclei also possess nuclear magnetic moment μ which is related to the nuclear spin angular momentum as

$$\mu = \gamma_n \hbar \mathbf{I} \tag{1.1}$$

where γ_n is the gyromagnetic ratio and which is a fundamental property of the nucleus.

An atomic nuclei with total angular momentum $\mathbf{I} \neq 0$ interact with the electromagnetic fields in their surroundings via nuclear magnetic dipole moments and, in the case of nuclei with $\mathbf{I} >$1/2 through nuclear electric quadrupole moments. A nuclear spin with $\mathbf{I} \neq 0$ is placed in an external magnetic field $\mathbf{B_0}$ that is applied in the direction of the z-axis. Depending on the orientation of the nuclear spins with respect to the direction of $\mathbf{B_0}$, the nuclear states have different energy values. This splitting is known as nuclear Zeeman effect and the Hamiltonian of interaction between the nuclear spins and the external magnetic field is given by;

$$H_z = -\mu.\mathbf{B_o} = -\gamma_n \hbar \mathbf{I}.\mathbf{B_o} = -\gamma_n \hbar B_o I_z = -\hbar \omega_L I_z \tag{1.2}$$

where ω_L is a positive number (for $\gamma_n > 0$) and I_z being the z-component of the spin angular momentum. The nuclear magnetic moment precesses about the z-axis

with a characteristic frequency is called the Larmor frequency $\omega_L = -\gamma_n \mathbf{B}_o$ (in rad/s) as shown in Fig 1.1 (B).

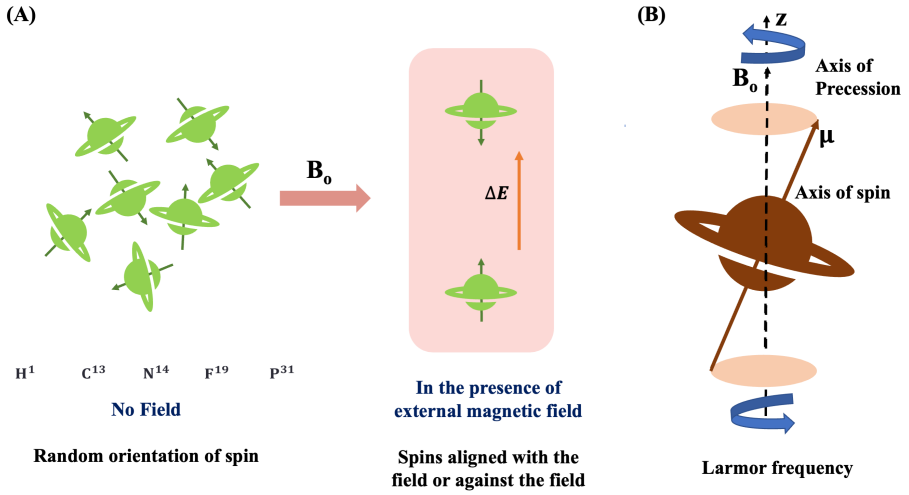

Figure 1.1: (A) Nuclear spins orientation in relation to the magnetic field strength, (B) Larmor frequency.

The z-direction corresponds to the magnetic field $\mathbf{B_0}$ axis, and all quantum operators work in the subspace defined by $|m\rangle$ where, m = -I, -I+1, -I+2, ...I-2, I-1, I. Under the action of such Hamiltonian, the expectation values of the Cartesian components of the nuclear spin operator in the plane perpendicular to the z-direction are: $\langle I_x \rangle$ and $\langle I_y \rangle$ show an oscillatory behavior with time, with a frequency given by $\omega_L = -\gamma_n \mathbf{B}_o$ (in rad/s), whereas $\langle I_z \rangle$ is stationary. In this sense the nuclear magnetic moment is analogous to the classical magnetic dipole [23]. The energy eigenvalues of the Hamiltonian are given by:

$$\mathbf{E}_m = -\mathbf{m}\hbar\omega_\mathbf{L} \tag{1.3}$$

As a result, a nucleus with spin $\mathbf{I}$ has $2\mathbf{I}$ + 1 energy levels that are equally spaced by the quantity $\hbar\omega_\mathbf{L}$.

For an ensemble of identical nuclei in thermal equilibrium, the Boltzmann distribution describes the population of each energy level [24].

For a two level system $I = \frac{1}{2}$ & $m = \pm\frac{1}{2}$, with the population of the energy levels n_+ $(m = \frac{1}{2})$ and n_- $(m = -\frac{1}{2})$ is governed by Boltzmann factor as

$$\frac{n_-}{n_+} = e^{-(E_{-I} - E_{+I})/k_B T} = e^{-\hbar\omega_L/k_B T} \tag{1.4}$$

where k_B is the Boltzmann constant and T is the absolute temperature of the spin ensemble. For ^{1}H protons ensemble placed inside a magnetic field of 14.1 Tesla, $\hbar\omega_L$ is around 2.47×10^{-6} eV, whereas, at room temperature, $k_BT \cong 2.5 \times 10^{-2}$ eV, so the Boltzmann factor $e^{-\frac{\hbar\omega_L}{k_BT}}$ is very close to the unity. This imply that the fractional difference of populations in this case is about 1 part in 10^5, which shows the low sensitivity of such experiments in NMR. This slight difference between the populations of the energy levels $m = \frac{1}{2}$ and $m = -\frac{1}{2}$ gives rise to a total thermal equilibrium magnetization along the z-direction given by

$$M_z = \frac{\mu_o\gamma_n^2\hbar^2B_o}{4k_BT} \tag{1.5}$$

where μ_o is the number of nuclei per unit volume (should not be confused for the magnetic moment μ). The dependency of the magnetization, which increases linearly with field intensity and is inversely proportional to temperature.

1.1.1 Interaction with a radio frequency (R.F.) field

Transitions between the energy levels can be induced by the application of oscillating magnetic fields with the appropriate frequency; $\omega_L = -\gamma_n\mathbf{B}_o$ (in rad/s). For nuclear spins, the Larmor frequencies are of the order of MHz, so they are excited by a radio frequency (R.F.) field.
The excitation of the nuclear spins system can be induced by a radio frequency (R.F.) oscillating magnetic field $\mathbf{B}_1(t)$, applied perpendicularly to the static magnetic field $\mathbf{B}_o(t)$, along the x-direction as given by

$$\mathbf{B}_1(t) = 2B_1cos(\Omega t + \phi)\hat{i} \tag{1.6}$$

where Ω and ϕ are the frequency and phase of the R.F. magnetic field and $\hat{i}$ is the unit vector in x-direction. The interaction Hamiltonian between nuclear spin and the R.F. magnetic field can be written as

$$H_{R.F.} = -\mu.\mathbf{B}_1(t) = -\gamma_n\hbar I_x[2B_1cos(\Omega t + \phi)] \tag{1.7}$$

The R.F. Hamiltonian can be assumed as a perturbation to the Zeeman Hamiltonian, as the magnitude of $\mathbf{B}_1(t)$ field is a few Gauss and is much smaller compared to the $\mathbf{B}_o$ field magnitude. As a result, H_z continues to play the predominant role, and the effect of $H_{R.F.}$ can be calculated by making use of the usual time-dependent perturbation theory [25]. A semiclassical interpretation for the excitation of nuclear spins is obtained by assuming that the linearly oscillating magnetic field $\mathbf{B}_1(t)$ is composed

of two circularly polarized fields, with same amplitude and phase as that of $\mathbf{B}_1(t)$, precessing about z-axis in opposite direction *i.e.*

$$\begin{aligned} \mathbf{B}_1(t) &= \mathbf{B}_1^+(t) + \mathbf{B}_1^-(t) \\ \mathbf{B}_1^+(t) &= B_1[cos(\Omega t + \phi)\hat{i} + sin(\Omega t + \phi)\hat{j}] \\ \mathbf{B}_1^-(t) &= B_1[cos(\Omega t + \phi)\hat{i} - sin(\Omega t + \phi)\hat{j}] \end{aligned} \tag{1.8}$$

For $\Omega = \omega_L$, i.e., on resonance, the $\mathbf{B}_1^-(t)$ component rotates around z-axis in sync with the nuclear Larmor precession, whereas $\mathbf{B}_1^+(t)$ rotates in the opposite. In a coordinate system rotating with angular velocity $\mathbf{\Omega} = -\Omega\hat{k}$, i.e. *rotating frame* the component $\mathbf{B}_1^-(t)$ will appear stationary to the nuclear spins, while $\mathbf{B}_1^+(t)$ rotates with twice the Larmor frequency. Therefore, only the field $\mathbf{B}_1^-(t)$ will have an effect on the nuclear spins, given that the magnitude of both fields is significantly less than that of the static field $\mathbf{B}_o(t)$. By controlling R.F. exposure time, spins can be excited from low to high energy eigenstates, which creates the NMR signal [26].

1.1.2 Nuclear spin interactions

The nuclear spin Hamiltonian $\mathrm{H}_{nuclear}$ describes interactions between the nucleus and electromagnetic fields present in its environment (including the interactions with electrons, other nuclei, other ions, etc) and it is given by

$$H_{nuclear} = H_{ext} + H_{int} \tag{1.9}$$

where H_{ext} denotes the nucleus interactions with external electromagnetic fields (external interactions), and H_{int} represents the internal interactions with the local environment of the nucleus. The two contributions to H_{ext} are the Zeeman and the R.F. Hamiltonians. The Zeeman and R.F. Hamiltonians are the two contributors to H_{ext}, and they have both been covered in previous section 1.1.1. It is convenient to write the external interaction Hamiltonian H_{ext} in the form:

$$H_{ext} = H_z + H_{R.F.} \tag{1.10}$$

The effect of the RF pulse is better explained in the rotating frame, where as H_{RF} plays the dominant role. There are multiple contributions to the internal Hamiltonian H_{int}, and these contributions are dependent on the physical properties of the material that was examined. The main interactions are often categorized in accordance with:

$$H_{int} = H_{CS} + H_D + H_J + H_Q \tag{1.11}$$

where H_{CS} represents the chemical-shift interaction of the nucleus with the orbital motion of the surrounding electrons; H_D denote the direct (through space) dipolar interaction between nuclei; H_J is the electron-mediated interaction between nuclei; and H_Q is the quadrupolar interaction between a nucleus with spin $>$ 1/2 and the electric field gradient at the nuclear position.

1.1.2.1 Chemical shift

The magnetic field that is actually experienced at the nuclear position is not the same as the external magnetic field. A local magnetic field is created when the orbital motion of neighboring electrons is disturbed, which results in an induced magnetic field. A local magnetic field is given by the equation:

$$\mathbf{B}_{loc} = (1 - \bar{\sigma})\mathbf{B}_o \tag{1.12}$$

where $\bar{\sigma}$ is known as the chemical shielding tensor. The tensorial nature of $\bar{\sigma}$ suggests that B_{loc} is in general in a direction that is different from that of B_o. This difference in direction reflects the anisotropy of the molecular environment of the nucleus that is being considered [23]. The Hamiltonian H_{CS} is given by:

$$H_{CS} = -\mu.(-\bar{\sigma}.\mathbf{B}_o) \cong \gamma_n \hbar \sigma_{zz} B_o I_z \tag{1.13}$$

Here, the component σ_{zz} in this case is determined by the relative orientation of the electron cloud in the molecule with respect to the external magnetic field. For an isotropic liquid substance, the average of all possible molecular orientations give an average value for the chemical shift that is called the isotropical chemical shift (σ_{iso}) [27].

$$H_{CS} \cong \gamma_n \hbar \sigma_{iso} B_o I_z \tag{1.14}$$

where, the quantity σ_{iso} is the trace of the tensor $\bar{\sigma}$. This is commonly represented in a molecular reference frame in which this tensor is diagonal, known as the principal axis system (PAS) of the tensor $\bar{\sigma}$:

$$\sigma_{iso} = (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})/3 \tag{1.15}$$

where σ_{xx}, σ_{yy}, and σ_{zz} denote the Cartesian components of the tensor $\bar{\sigma}$ in the PAS.

1.1.2.2 Dipolar coupling

Any two magnetic dipole moments can directly interact with one another due to the magnetic fields each one of them creates on the other's position. The magnetic field that is produced by a classical point dipole at a point that is located by the vector $\mathbf{r}$ with its origin on the dipole is (in SI units) [28].

$$\mathbf{B}_{dip} = \frac{\mu_o}{4\pi}\frac{3(\mu.\mathbf{e})\mathbf{e} - \mu}{r^3} \tag{1.16}$$

where μ is the magnetic dipole moment, r is the magnitude of the vector, e is the unit vector in its direction, and μ_o is the magnetic permeability of free space.

1.1.2.3 J -coupling

The J -coupling also known as indirect or scalar coupling is an interaction between the nuclear magnetic dipole moments of neighboring nuclei. In this case the interaction is not direct but it is mediated via the electron cloud involved in the chemical bonds between the atoms. Hamiltonian that is used to describe the J -coupling between two nuclear spins I_1 and I_2 is expressed as:

$$\mathbf{H}_J = 2\pi\hbar\mathbf{I_1}.\tilde{J}.\mathbf{I_2} \tag{1.17}$$

where $\tilde{J}$ represents the tensor and give non-vanishing trace.

1.1.2.4 Quadrupolar coupling

All nuclei with spin I $>$ 1/2 have a non-spherical charge distribution, they are electrostatically interacting with neighboring electrons and ions. The nucleus interacts with the electric field gradient (EFG) at the nuclear position via its electric quadrupole moment. Quadrupolar coupling is defined by the following Hamiltonian:

$$H_Q = \hbar\omega_Q(3I_z^2 - \mathbf{I^2}) \tag{1.18}$$

This basic form is only applicable when first-order perturbation theory can be applied safely.

1.1.3 NMR spectrometer

Fig 1.2 shows a schematic NMR instrument containing the most expansive part: a magnet, radio frequency oscillator, and a detector. In a brief, a spectrometer shown in Fig 1.3, is made up of the following parts:

- **Superconducting magnet**: Most NMR experiments require a homogeneous magnetic field. Modern NMR spectrometers use persistent superconducting magnets to generate the B_o magnetic field. A typical magnet is made up of a coil through which a current flows and generate the magnetic field. The superconducting magnet coil windings are commonly built of a Nb/Sn alloy [29]. The coil wire is kept at a low temperature (less than 6 K) to become superconducting, meaning that its resistance approaches to zero. Therefore, once the current is set to running in the coil, it will continue for all of time, thereby producing a magnetic field without requiring any additional electrical power [30].

- **The transmitter**: NMR spectrometer irradiate samples with R.F. fields at the Larmor frequency and detect nuclear radio signals. The transmitter section of the spectrometer is capable of generating R.F. pulses. Typically, the transmitter is divided into multiple sections mainly: the synthesizer, the pulse gate, radio-frequency amplifier. Each of which is responsible for producing R.F. signals at frequencies near to the Larmor frequencies of certain isotopes [29].

- **The duplexer**: The duplexer receives the amplified R.F. pulse via a cable. A signal coming from the pulse programmer causes the duplexer to rapidly flip between the transmit mode and the receive mode [29]. Two more cables come out of the duplexer, one cable leads to the probe and another one goes to receiver section which is used to detect the weak R.F. signals generated by the nuclear spins.

- **The probe**: The probe consists of a metal tube in the shape of a cylinder that is put into the bore of the magnet [30]. The small coil used both to excite the nuclear spins and detect the NMR signal. In some instances, the probe contains a device for rotating the sample to narrow the line-widths of NMR signals.

- **The receiver**: The probe is producing very weak NMR signals. These signals can be amplified in the receiver section. Pre-amplifier is usually placed near to the probe to amplify all the NMR signals. This is a low-noise R.F. amplifier that increases the voltage level of the tiny signal [29]. It is necessary to input the NMR signal into the computer so that it can be interpreted and presented. The NMR signal, which is an oscillating electrical current, must therefore be converted into digital form using analogue-digital converter.

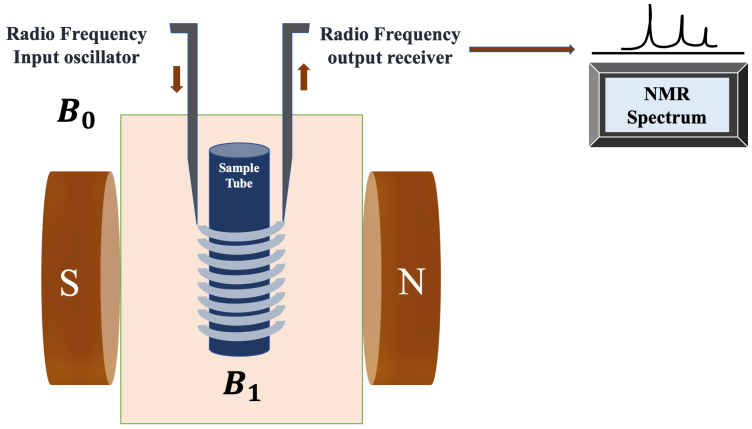

Figure 1.2: Schematic of a typical NMR spectrometer.

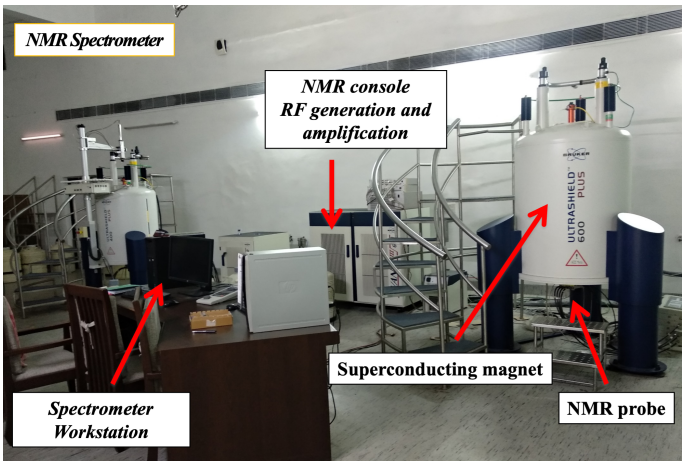

Figure 1.3: Photograph of IISER Mohali NMR Research Facility.

1.1.4 Pulse sequences

Typically, a variety of one-dimensional (1D) and two-dimensional (2D) NMR experiments are needed to identify chemical compounds. The 1D ^{1}H NMR experiment is the one that is used in the vast majority of NMR-based metabolomics research. This is due to the fact that 1D ^{1}H NMR is more sensitive than other nuclei (such as ^{13}C and ^{15}N), with high isotopic natural abundance (99%). More importantly, protons are present in the majority of metabolites.
The one-dimensional ^{1}H CPMG (Carr-Purcell-Meiboom-Gill) sequence and ^{1}H NOESY

(Nuclear Overhauser effect spectroscopy) with water presaturation are the two pulse sequences that are most frequently utilized. The CPMG pulse sequence removes broad signals from macromolecules such as proteins and lipids [31], [32], [33]. NOESY spectra provide a quantitative and comprehensive profile of the detected metabolites by suppressing the water peak without affecting the strength of the other peaks [31], [32]. In one-dimensional NMR spectra, signals from several metabolites overlap substantially. Utilizing two-dimensional NMR experiments is one approach that can be used to resolve this issue. For identification, ^{1}H-^{1}H homonuclear correlation experiments are frequently utilized. COSY (correlation spectroscopy) identifies proton spin-spin interaction [31], [32], [33]. TOCSY (total correlation spectroscopy) shows the correlation between all protons in metabolites [32], [33]. Another experiment is known as ^{1}H J-resolved, and it separates the influence of chemical shift and J-coupling into two distinct directions [34]. When identifying new metabolites, it can be advantageous to employ ^{1}H-^{13}C heteronuclear correlation experiments. This is because these experiments provide a correlation between the two nuclei [32], [33]. Two dimensional ^{1}H-^{13}C heteronuclear single quantum coherence experiments measure a chemical shift correlation between directly bonded ^{13}C and ^{1}H which give one cross-peak for each C-H pair in the spectra. The 2D ^{1}H-^{13}C Heteronuclear multiple bond correlation (HMBC) experiment compares the chemical shifts of two different types of nuclei (^{1}H and ^{13}C) that are separated from one another by two or more bonds [35].

1.2 Introduction to Metabolomics

The overview of four key omics families, from genomes to metabolome, is shown in the Fig 1.4. At first, there is DNA, which stands for deoxyribonucleic acid. DNA is what stores all of an organism's genomic information, and the field of study that is concerned with genomes is called genomics. Transcription is the next process that happens after DNA. The term "transcriptome" represents a set of a cell's RNA (ribonucleic acid) composition, and the study of transcriptomes is referred to as "transcriptomics". Then after transcriptions, here comes the process of translations in which proteins are made. The study related to these kind of activities is called proteomics. The next process is metabolomics, which comes right after proteomics.
The metabolomic processes within an organism produce metabolites, and all of these small molecules including sugars, nucleotides, amino acids, and lipids combine to form the metabolome. In brief, metabolomics is the large-scale study of tiny molecules, known as metabolites, within cells, biofluids, tissues, and organisms [36]. As a result, metabolomics clearly reflects the molecular phenotype (an individual observable traits, such as height, eye color, blood type and thumb print). There are two major categories

of metabolites: primary and secondary. Primary metabolites have an important part to play in all stages of an organism's life cycle, including growth, development, and production. Secondary metabolites have a significant role in ecological processes such the formation of pigments, antibiotics, and defense mechanisms. Fatty acids, amino acids, vitamins, carbohydrates, and organic acids are examples of primary metabolites. Secondary metabolites include alkaloids, flavonoids, steroids, phenolics, essential oils etc.

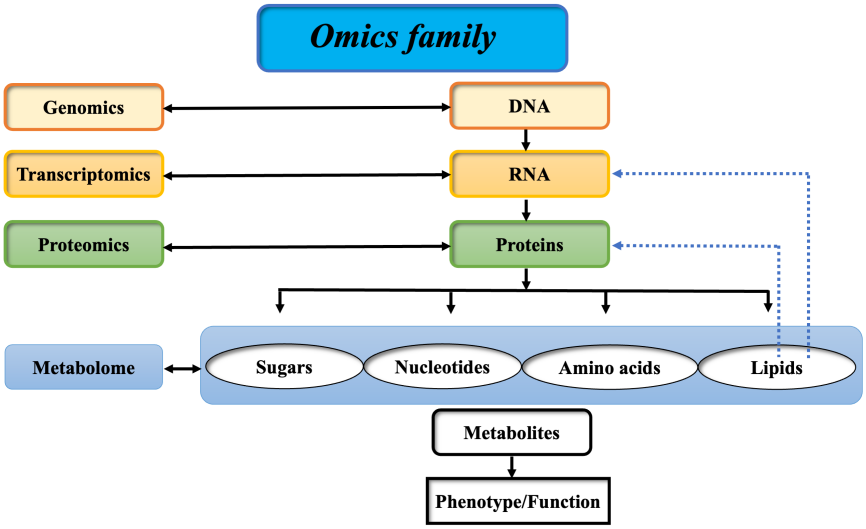

Figure 1.4: Connection networks among several omics families.

1.3 NMR-Based Metabolomics

Over the past two decades, a number of outstanding reviews have been written that offer a great introduction to the concepts & methods of NMR-based metabolomics [7], [37], [38], [39] and on quantitative metabolomics [40], [41]. The term "metabolomics" refers to the study of all the metabolites found in a given organism, cell, or tissue which have a low molecular weight (MW $<$1000 g mol^{-1}) and are produced naturally in the living organism [42], [43], [44]. Metabolomics is a powerful method because metabolite patterns directly represent underlying metabolic activity and tissue condition. Metabolomic studies mostly use NMR or MS as an analytical technique, followed by multivariate statistical analysis [45], [46]. Additionally, it is a reliable method for investigating several primary and secondary metabolites in targeted or non-targeted metabolomic analysis [47]. High-resolution NMR analysis provides a detailed

and particular perspective of cellular metabolic processes under normal and disturbed (disease-related) conditions. The study of metabolites under varying conditions aids to a better knowledge of various aspects of biology [48], including biochemical mechanisms, diseases, and genomic characteristics. Hence, NMR-based metabolomics is a well-established and powerful tool for investigations in different fields of science, including medicine & drug discovery [42], food science & food chemistry [49], plant science [50], nutraceutical research [51], stress response [52], phytochemical analysis, [53], and in defense signaling [54]. NMR metabolomics is being utilized to identify disease biomarkers for a wide range of human diseases including cancer [55], [56], [57], Parkinson's disease [58], neurological disorders [59], infectious diseases like tuberculosis [60], malaria [61], and pneumonia [62]. The use of ^{1}H-NMR metabolomic profiling has been described in several studies to demonstrate unique metabolic recovery responses in shoots and roots of momentarily drought-stressed sugar beets [63], as well as metabolic disturbances in potato after cold storage [64]. Fig 1.5 represents the several applications in NMR-based metabolomics. Fig 1.6 shows the NMR-based metabolic workflow. The workflow begins with the collection of leaves/fruits from plants in order to prepare a sample, which is then followed by performing NMR experiments. The NMR raw data are pre-processed and reduced by binning of the spectra. Later, 1D, 2D NMR, pure compounds spectra & databases were used to identify and assign the various metabolites. Univariate & multivariate statistical analysis is executed to identify the outliers, examine the sample variation, and find the potential biomarkers. Finally, the metabolic pathway is used to interpret the changes in metabolic fluxes of potential bio-markers. General procedure for plant/fruit sample preparation is shown in Fig 1.7.

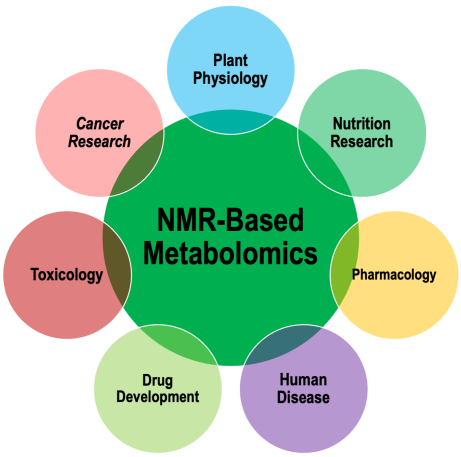

Figure 1.5: A Schematic representation of NMR-based metabolomic applications.

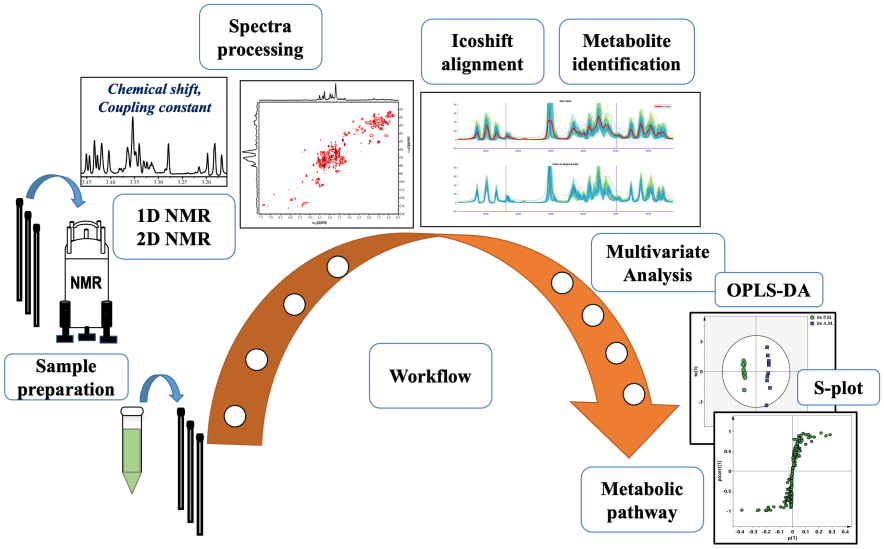

Figure 1.6: NMR-based metabolomics workflow.

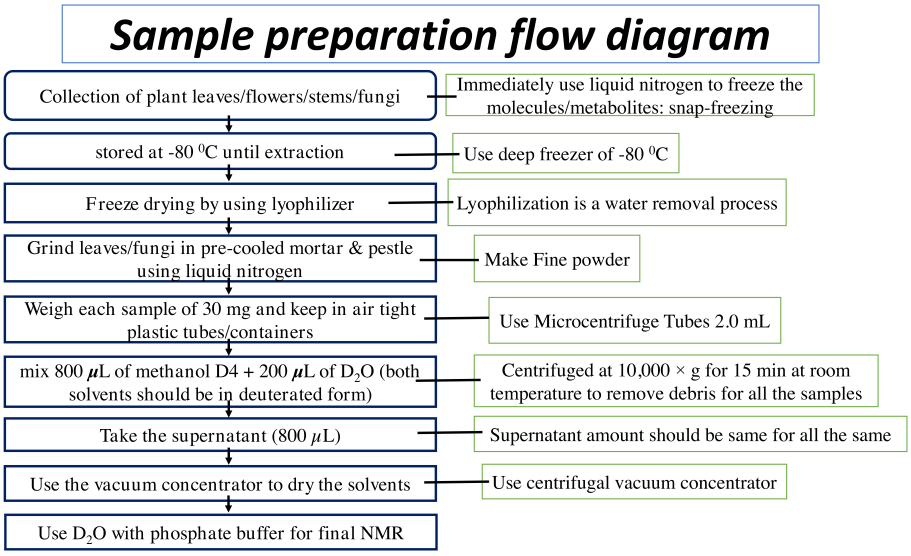

Figure 1.7: Sample preparation procedure.

1.4 Univariate & Multivariate Statistical Methods

Univariate analysis is the most basic type of data analysis and refers to statistical testing that involves only one variable [65]. The choice of the procedure is determined according to the statistical characteristics of the feature distribution [66]. For metabolomic data analysis, several univariate analysis approaches are available. Some examples of univariate statistics are the t-test and the analysis of variance (ANOVA). Multivariate analysis is recognized as one of the most useful methods for evaluating metabolomics data [67]. Multivariate data analysis can be used to look for systematic variations in the data set by analyzing the selected observations and variables [68]. Additionally, it may provide a thorough overview, recognize patterns, predict models, and classify groupings [69]. The three most prevalent models that are utilized in the process of analyzing metabolomics data sets are: principal component analysis (PCA) [70], [71], partial least squares discriminant analysis (PLS-DA) [72], [73], [74], and orthogonal partial least squares discriminant analysis (OPLS-DA) [75], [76]. PCA is an unsupervised model that depicts systematic trends and patterns in the data matrix (X) that consists of observations and variables in the rows and columns, respectively. This model does not use any predetermined groupings to represent the trends and patterns. The principal components (PCs) of the dataset, which are linear combinations of the variables, are used to depict the variation in the data. The plot of PCA scores displays the responsible observations as a function of the principal components. The corresponding PCA loadings plot is used in conjunction with PCA scores plot to identify the significant variables that are responsible for the observed differences in the datasets. The connection between the variables is contributed by the vectors (p) in the loadings plot. PCA model validity is indicated by the goodness-of-fit (R^2) (elucidated variation) and predictive ability (Q^2) values [68], [69]. In addition, PLS-DA is a supervised model that generates Y variables and dummy Y variables which adds discriminating information and makes an effort to locate the discriminatory line that separates two groups [72]. The main advantage of utilizing OPLS-DA model in the field of metabolomics is the identification of spectral features (metabolite fluctuations) that identify the separations between groups. OPLS-DA model is the one that is utilized the most. The OPLS-DA model is helpful for determining why two groups are distinct from one another. In essence, the changes in the metabolome that are biologically significant are found using OPLS-DA model [76]. Thus, the fundamental advantage of OPLS-DA over PLS-DA in terms of interpretation is its ability to separate between orthogonal component (non-predictive) and predictive component [77].

1.4.1 Principal Component Analysis (PCA)

For metabolomics, (PCA) is likely the most popular and practical multivariate statistic [78]. The primary goals of PCA are to identify patterns in huge datasets with a focus on data similarities and differences in the NMR via principal component “clusterings", and to help users identify outliers that are statistically significant [79]. The data will cluster, and the classification of each group can be assessed [79]. Its unsupervised nature and consequent impartiality make it useful for non-targeted metabolomic research [80]. As a result, the method seeks to identify the largest variation in a set of data (X) without taking the class into account (Y) [80]. PCA examines the relationships between the variables to find a new structure for a group of variables, such as principal components (PCs) or scores with a weight or loading [79], [80]. A linear transformation of the starting variables is carried out using the proper weighting factors in PCA, and the tool makes an effort to preserve as much variance as is practicable by ensuring that each PC is uncorrelated with the others throughout the process [78], [79]. The principal component analysis (PCA) will keep the most variance in its first component, whereas successive components will contain less and less until there is no variance [79]. The PCA plot is constructed by determining the eigenvectors and variance based on the data matrix that is provided, which in the case of metabolomics is typically an NMR dataset matrix. The non-singular portion of the sample covariance matrix (S) that corresponds to the column is used to generate the eigenvectors. The number of variables that make up the principal component will be shown by the eigenvalue. In the field of metabolomics, a principal component (PC) is formed by a group of metabolites that has either a negative or positive correlation. A second PC will be formed by a second group of metabolites that are correlated, and PCA will compare these two sets of loadings.

1.4.2 Orthogonal partial least squares discriminant analysis (OPLS-DA)

In order to differentiate between two or more groups (classes) by utilising multivariate information, OPLS-DA was developed as an enhancement of the PLS-DA approach [77], [81]. A regression model is computed in OPLS-DA between the multivariate data and a response variable that solely contains information about classes. In OPLS-DA, as compared to PLS-DA, only one component is utilised as a predictor for the class, while the remaining components characterise the variance in a way that is orthogonal to the first predictive component. This gives OPLS-DA a significant advantage over PLS-DA.

1.5 Applications of NMR-Based Metabolomics to Plants

NMR combined multivariate analysis is a powerful approach to study plant metabolomics. NMR has been used extensively to investigate the effects of biotic [82], [83] and abiotic stress [84], [85] on the metabolic profile of plants. Examples include, the effects of abiotic stress on wheat [86], soybean plant [87], legume crop [88] as well as the effects of biotic stress on cereals [82], rice [83]. Exposure to various stress situations in plants causes changes in metabolite composition and changed metabolic interactions, which can improve plant defense mechanisms. The targeted biological system may experience stress as a result of changes in environmental factors like light, temperature, and water [89], as well as an attack by other living organisms [90] such as viruses, bacteria, and fungi. The metabolome of plants can be investigated with the help of NMR-based metabolomics. These investigations include a wide range of topics pertaining to plant physiology [91], including the circadian rhythm [92], the metabolic response to stress [93], and plant signaling [94], [95]. NMR-based metabolomics is used in herbal extracts as medicinal agents and in the quality control of herbal medicine [96]. Fig 1.8 shows the medicinal properties of flavonoids rich foods.

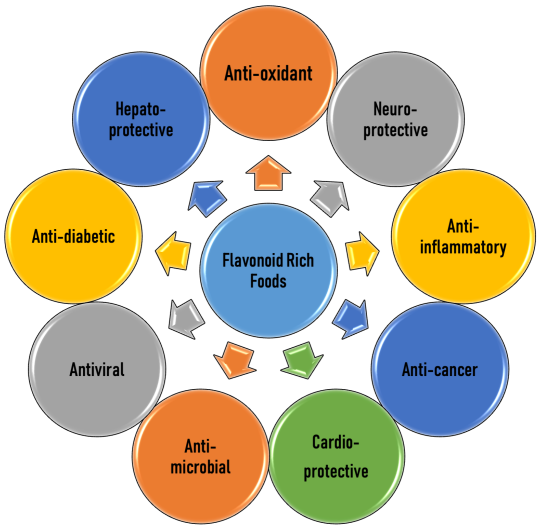

Figure 1.8: Medicinal properties of flavonoid rich foods.

1.6 Applications of NMR-Based Metabolomics to Drosophila Melanogaster metabolism

The application of NMR-based metabolomics highlights the significance in understanding numerous aspects of insect physiology, such as the evolution of insect immune response. Metabolomics is a cutting-edge tool for studying and exploring the physiology of Drosophila melanogaster. This method allows researchers to detect hundreds of metabolites in a single spectra which can represent the combined contributions of genes, proteins and various environmental exposures (lifestyle, temperature, diet etc.) of Drosophila melanogaster. Various studies on Drosophila melanogaster reported are: the effect of heat stress [97], hypoxia [98], cold shock [99], hypoxia in the heart [100] and circadian regulation of metabolites [101].

1.7 Organization of this thesis

The work done towards this thesis has been divided into eight chapters, briefly described below: Chapter 1 gives an overview of NMR spectroscopy and metabolomics. Chapter 2 describes NMR-based metabolomics application to study phytomedicinal properties in Momordica charantia fruit. The aim of this study was to identify & quantify the phytomedicinal compounds in the pericarp, skin, and seeds of mature Momordica charantia fruit. Chapter 3 highlights the role of NMR-based metabolomics in profiling the phytomedicinal constituents in Phyllanthus emblica and Tinospora cordifolia. Chapter 4 examines the utility of NMR-based metabolomics for metabolite fingerprinting of young Helianthus annuus L. (sunflower) stems throughout the circadian cycle. Chapter 5 deals with the study of effect of air pollution stress on Bougainvillea spectabilis leaves using NMR-based metabolomic approach. Chapter 6 describes the ability of NMR-based methods to study plant response against wounding stress. NMR metabolomic approaches were utilized to analyze metabolic changes in Bougainvillea spectabilis leaves after a wounding treatment to simulate a herbivore attack. Chapter 7 addresses the use of NMR-based metabolomics techniques to explore evolution in the fruit fly Drosophila melanogaster at various ages. Chapter 8 provides a summary of the application of NMR-based metabolomics and possible future directions are outlined.

Chapter 2

NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

2.1 Introduction

Momordica charantia, also known as bitter melon, is a plant widely utilized in traditional Ayurvedic medicine. Its antioxidant [102], [103], anticancer [104], [105], antiosteoporosis [106], and antibacterial [107] properties have been extensively researched. The various bioactive components of Momordica charantia, including polysaccharides, saponins, steroids, triterpenoids, alkaloids, amino acids, phenolics, and flavonoids, have been credited with the plant's medical efficacy [108], [109]. There is currently a lot of interest in creating α-glucosidase inhibitors from plants that do not have any negative side effects and could be utilized as anti-diabetic medications [110]. The potential of Momordica charantia to treat type-2 diabetes mellitus has been thoroughly researched [111]. Steroidal saponins such charantins, alkaloids, and insulin-mimetic peptides have been discovered as the bioactive components of Momordica charantia that are responsible for its hypoglycemic effects [112], [113], according to several studies on the bioactive ingredients of Momordica charantia. There has been a lot of research done on the process of extracting the triterpenoids charantin and momordicine from the fruit of the Momordica charantia plant [114], [115]. In diabetic mice induced with alloxan, the saponin fraction found in Momordica charantia demonstrated hypoglycemic effect [116]. An improvement in the functioning of β-cells as well as an increase in insulin levels has been found in neonatal streptozotocin-induced type-2 diabetic rats treated with an ethanolic extract of Momordica charantia [117]. NMR

spectroscopy-based plant metabolomics investigations are frequently utilized to determine whether plant-based therapeutic medicines are able to rectify disrupted metabolic pathways in diseased organisms [118], [119], [120], [121], [122]. In this study, we employed an NMR-based metabolomics approach to determine the types of metabolites that are found in the pericarp, skin, and seeds of mature Momordica charantia fruit. These are the three major portions of the fruit that were examined. We used multivariate statistical analysis to identify metabolites that differ significantly in different parts of the fruit. We determined the total phenolic and flavonoid content of the pericarp, seeds, and skin of the fruit by conducting a quantitative analysis [123]. In order to determine the antioxidant abilities [124] of the Momordica charantia fruit, we subjected it to both the 2,2-diphenyl-1-picrylhydrazyl (DPPH) and 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS+) free radical scavenging tests. Using an α-glucosidase inhibitory activity bioassay, we looked for signs of antidiabetic efficacy [125].

2.2 Experimental Methods

2.2.1 Materials

Organic Momordica charantia mature fruits (picked approximately 3-4 weeks post-flowering) were bought from the market. Ten Momordica charantia mature fruits approximately 8×3 cm were chosen for the study. The skin, pericarp and seeds of the mature fruit was green, white and yellowish in color respectively, and the seeds were completely developed. The fruits of Momordica charantia used in this study was collected by Sumit Mishra from local market (latitude 30.68110 N & longitude 76.74660 E) of Mohali (Panjab) in October 2019.

2.2.2 Chemicals and reagents

All extraction solvents, reagents, NMR reference standards and deuterated solvents were of analytical grade with purity greater than 90% and were procured from Sigma Aldrich (India). Deuterium oxide (D_2O, 99.9%) and deuterated methanol-D_4 (CD_3OD-D_4, 99.80%) were used as the deuterated solvents. Trimethylsilane propionic acid sodium salt (TMSP) was added as an internal standard for all ^1H NMR measurements. Sodium azide (NaN_3, 99.5%) and phosphate buffer solution (2.0 mM, pH 7.4) were used in this study. Metabolite extraction from Momordica charantia was performed at room temperature using CD_3OD and D_2O as extraction solvents via standard protocols [120].

2.2.3 NMR sample preparation

Ten replicates from the individual fruits were analyzed for each type of sample: skin, pericarp and seeds. The fruits were surface cleaned by washing them with distilled water to remove all debris and dried with tissue paper. The pericarp, seeds were removed, and the skin was cut into thin pieces. Then the samples were ground in liquid nitrogen using a mortar and pestle and kept in separate air-tight plastic containers. These samples were then stored overnight at -80 ^{0}C. All the samples were then lyophilized using a freeze dryer at a pressure of 30 Pa for 16 h at -40 ^{0}C. NMR samples were prepared by taking 25 mg of dried skin, pericarp and seed powder mixed thoroughly with the solvents containing 320 μL of methanol-D_4, 180 μL of D_2O (phosphate buffered saline, pH 7.4 made in D_2O), 100 μL of sodium azide (to prevent bacterial growth), and 2mM trimethylsilane propionic acid sodium salt (TMSP) as an internal reference. Then 600 μL solution was vortexed and centrifuged at 5000 g for 5 min and 500 μL of the filtered supernatant was used for the NMR experiments.

2.2.4 NMR spectroscopy

All NMR experiments were performed at 298 K on a Bruker Biospin 600 MHz Avance III spectrometer operating at a proton resonance frequency of 600.219 MHz, equipped with a 5 mm QXI quadrupolar resonance probe. Methanol-d_4 was used as an internal lock and gradient shimming was performed prior to signal acquisition. For 1H NMR spectra acquisition, a water suppressed Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequence optimized with a spin-echo delay τ of 300 μs and loop counter n = 400 and a total spin-spin relaxation delay time ($2n\tau$) of 240 ms was used to attenuate the background broad signals from large molecules. The proton spectra were collected with a 90-degree pulse width of 9.95 μs, a relaxation delay of 4 s, 8 scans, 64 K data points and a spectral width of 12 ppm. Data were zero-filled by a factor of 2 and the FID's were multiplied by an exponential weighting function equivalent to a line broadening of 0.3 Hz prior to Fourier transformation. The NMR spectra were phase and baseline-corrected and trimethylsilane propionic acid sodium salt (TMSP) was added to provide an internal NMR reference at 0.00 ppm. To confirm the peak assignments, 2D NMR experiments were recorded, including homonuclear ^{1}H-^{1}H correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and heteronuclear ^{1}H-^{13}C coherence spectroscopy (HSQC). 2D COSY and TOCSY spectra were recorded with a spectral width of 12 ppm in both the proton F1 and F2 dimensions, 2K data points, 16 scans and 128 t_1 increments. 2D HSQC spectra were acquired with a spectral width of 12 ppm and 200 ppm in proton and carbon dimensions, respectively, 1K data points, 32 scans and 128 t_1 increments.

2.2.5 Metabolite identification

The metabolites were identified using both 1D and 2D NMR experiments and also from comparing the obtained chemical shift and coupling constant values with standard NMR metabolite peaks from databases such as Biological Magnetic Resonance Data Bank (BMRB) (http://www.bmrb.wisc.edu), the Madison Metabolomics Consortium Database (MMCD) (http://mmcd.nmrfam.wisc.edu/) and the Human Metabolite Data Base (HMDB) (http://www.hmdb.ca/). Pure beta-sitosterol and stigmasterol were recorded in deuterated chloroform and used as a reference to identify peaks in the NMR spectra of Momordica charantia. We tried to record the NMR spectra of as many pure compounds as possible such as chlorogenic acid, gamma-aminobutyric acid (GABA), quinic acid, sucrose, tryptophan and vanillic acid (recorded in a mixture of deuterated methanol-D_4 and D_2O). We used the metabolomics standards initiative (MSI) [118] to classify the identified metabolites: as belonging to MSI level 1 if two or more independent and orthogonal parameters are identical to those of a reference compound interpreted under the same conditions (for instance through 1D ^{1}H and 2D HSQC NMR spectra). If this information is not available the metabolite is putatively annotated from publications, databases or libraries and categorized as belonging to MSI level 2.

2.2.6 Multivariate statistics

The NMR data were converted to ASCII format and aligned using the icoshift (interval correlation shifting) MATLAB algorithm [119]. Further NMR data analysis was performed using MestReNova software version 10.0.2-15465 (Mestrelab Research, Spain). The spectra were subdivided in the range 0.02 to 10.0 ppm and binned into regions of equal spectral width (0.01 ppm), referenced to a peak of TMSP. Data were normalized to a total integral of 100. The spectral region δ 4.60-5.12 ppm contained residual water peaks and was not considered for the analysis. Prior to the chemometric analysis, all the data sets were Pareto-scaled. Principal component analysis (PCA), Orthogonal projections to latent structures discriminant analysis (OPLS-DA), one-way analysis of variance (ANOVA), statistical Student's t-tests, and hierarchical clustering heat map analysis were performed using SIMCA version 14.1.0-2047 software (Umetrics, Umea, Sweden) and Metaboanalyst software (http://www.metaboanalyst.ca/). For detection of outliers, PCA was first performed and data points located outside the 95 % confidence region of the Hotelling's ellipse in the PCA score plot, were removed from the subsequent OPLS-DA analysis. The quality of the OPLS-DA model was assessed from the R2Y (variance predicted) and Q2 (variance explained) values. The separation between the groups were observed using the OPLS-DA scores plot, having one predictive and one orthogonal component. Significant metabolites, responsible for group separation were identified from the S-plot and Student's t-test and p-values $\geqslant$

0.05 and 0.01 respectively, were considered significant. Metabolomic pathway analysis was constructed using the MetPA web-based tool (http://www.metaboanalyst.ca). The highest impact value was obtained for the flavone and flavonol synthesis (impact value = 0.8, FDR p-value < 0.05). Metabolic pathways including: flavonoid biosynthesis, phenylpropanoid biosynthesis, aminoacyl-tRNA biosynthesis, sulfur metabolism, cysteine and methionine metabolism, butanoate metabolism, alanine, aspartate and glutamate metabolism were statistically significant (impact value $\geqslant 0.005$, FDR p-value $<$ 0.05).

2.2.7 Quantifying total phenolic content

The total phenolic content in pericarp, skin and seeds of Momordica charantia was quantified using the Folin-Ciocalteu spectrophotometric method [121]. The fruit extract (1.0 mL of methanol extraction) was mixed in a vial with 5 mL of Folin-Ciocalteu phenol reagent (diluted 10 times with distilled water). After 5 minutes, 5 mL of 7.5% Na_2CO_3 solution was mixed and the vials were kept for 20 minutes at 25 ^{0}C to complete the reaction. The absorbance was measured on the UV-Vis spectrophotometer at 760 nm at room temperature. The standard curve was prepared using gallic acid. The total phenolic content for each type of extract (pericarp, skin and seed) was expressed as milligrams of gallic acid equivalent (mg GAE) per g of dried weight sample (g DW). All measurements were performed thrice and the results were averaged.

2.2.8 Quantifying total flavonoid content

The total flavonoid content in different parts of the Momordica charantia was quantified by the method described by [103]. 1 mL of fruit extract (methanol extraction) was mixed in a vial with 6 mL of distilled water, 200 μL of aluminum chloride solution and 200 μL of potassium acetate solution. The mixture was allowed to stand for 30 min. at 25 ^{0}C to complete the reaction. The absorbance was measured at 510 nm using a UV-Vis spectrophotometer. The total flavonoid content was calculated from a calibration curve obtained using a standard solution of quercetin, and the result was expressed as mg quercetin equivalent (mg QE) per g of dry weight sample (g DW). All measurements were performed thrice and the results were averaged.

2.2.9 Quantifying free radical scavenging activity

Free radical scavenging activity was measured from a 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay using the method described by [123], wherein 4 mg of fruit extract was

dissolved in 10 mL of methanol to prepare a 400 μg/mL solution and then serial dilution was performed to prepare the required concentrations. As a positive control, 2 mg of quercetin (12.5 - 400 μg/ml) was dissolved in 2.5 mL distilled water. 1 mL of methanol solution of fruit extract/standard of different concentrations was mixed in a vial with 3 mL of methanol solution of DPPH and incubated at room temperature for 30 minutes in the dark. Absorbance of the solution was recorded at 517 nm using a UV-Vis spectrophotometer. The DPPH radical scavenging capacity (SC %) was calculated as:

$$\text{SC } (\%) = \frac{(A_0 - A_s)}{A_0} \times 100$$

where, A_0 is the absorbance of the reagent blank and A_s is the absorbance of the samples. The results were expressed as EC_{50} (the half maximal effective concentration), which denotes the concentration of the sample required to scavenge 50% of the DPPH free radicals. A stoichiometric change in color occurs during the conversion from a DPPH free radical to reduced DPPH-H

$$\text{DPPH}^{'} + \text{A-H} \rightarrow \text{DPPH-H} + \text{A}^{'}$$

It was observed that the purple color of DPPH was reduced upon treatment with the Momordica charantia extracts at all the tested concentrations. The scavenging activity increases with the increase in concentrations up to 400 μg/mL for the extracts, after which a plateau is reached. The antioxidant activity of plants has been attributed to the presence of phenolic compounds which have the ability to lose protons and can hence contribute to radical scavenging. The results of the DPPH assay corroborate those obtained from evaluating the total phenolic content. Seed extracts of Momordica charantia, which had the highest total phenolic content also showed the highest antioxidant activity.

2.2.10 Quantifying radical cation scavenging activity

Antioxidant activity of Momordica charantia extracts fruit extracts with the 2,2$^{'}$-azinobis(3- ethylbenzothiazoline-6-sulphonic acid) diamonium salt (ABTS+) cation radical was measured using the method reported by Rajurkar et. al. with some modifications [124]. A stock solution of ABTS+ radical was prepared by mixing ABTS (7 mM) with potassium persulfate (2.45 mM) in distilled water. The solution was kept in the dark at room temperature for 12-16 hours before use, and then diluted with methanol to obtain an absorbance of 0.70$\pm$0.02 units at 734 nm on the UV-Vis spectrophotometer. Methanol fruit extracts (10 μL of each sample) were mixed thoroughly

with 2.99 mL of the ABTS+ stock solution in a cuvette in the dark, and the decrease in the absorbance was measured after 1 hour. The standard curve was obtained using 2 mg Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) dissolved in 2.5 mL distilled water and then serial dilutions performed of different concentrations of 31.25–500 μg/mL. ABTS scavenging ability was expressed as half-maximal effective concentration (EC_{50} in μg/mL).

2.2.11 α-glucosidase inhibition assay

p-Nitrophenyl-α-D-glucopyranoside (p-NPG), yeast α-glucosidase, sodium carbonate and acarbose were purchased from Sigma Aldrich and the α-glucosidase inhibition assay was performed using the procedure as reported by [125]. Each sample was prepared by adding of 500 μl phosphate buffer (67mM, pH=6.9), 100 μl fruit extract/acarbose (0.25-4.0 mg/ml) with 100 μl of α-glucosidase (0.5 U/ml). The mixture was incubated at room temperature for 15 minutes. 250 μl of pNPG (50 mM in phosphate buffer) was used as substrate and then added to the sample and the mixture was re-incubated at room temperature. The reaction was stopped by using of 200 μl (0.1 M) sodium carbonate. The absorbance was measured at the wavelength of 405 nm using a spectrophotometer. Acarbose and distilled water were used as positive and negative controls, respectively. The percentage of α-glucosidase enzyme activity was calculated as follows:

$$\text{Inhibition of Sample (\%)} = \frac{(A_0 - A_s)}{A_0} \times 100$$

where, A_0 is the absorbance of the control and A_s is the absorbance of the sample.

2.2.12 UPLC-ESI-MS experiments

For mass spectrometric detection, a high-resolution mass spectrometer (Waters Synapt G2-S) was used. The separation was analyzed using reverse phase chromatography on a Waters Acquity UPLC BEH C18 column (2.1×50 mm column size, 1.7 μm particle size). The system is equipped with a binary solvent (mobile phase), a column controller and sample chamber. The mobile phase solvent consisted of water containing 0.1% formic acid (elution buffer A) and acetonitrile containing 0.1% formic acid (elution buffer B). The initial conditions were 95% A at a flow rate of 0.2 mL/min. After that it was processed by multiple linear gradients to 5% A and then again to 95% A at 22 min. The polarity ES (in negative ionization mode) was used for mass analysis, with a mass range m/z between 50 to 1500. The optimal conditions for analysis were

achieved with an injection volume of 10 μL, a capillary voltage of 2 kV, and sampling cone voltage of 40 eV. The source temperature was 150 ^{0}C and the desolvation temperature was set at 350 ^{0}C. The standard mass of Leucine Enkephalon (554.2615 Da in negative mode) was used as the reference for mass calibration. The spectra were processed and analyzed using Waters Masslynx^{TM} v4.1 software. We performed targeted mass-spectrometric analysis of methanol extracts of different parts of the Momordica charantia fruit, in order to validate the presence of the steroids and phenolic compounds that were identified via NMR fingerprinting. These targeted metabolites underwent the same chromatographic conditions and were separated and eluted off the C18 column between 0 to 22 min. The identified metabolites present in the skin, pericarp and seed extracts based on the detection of [M-H]$^-$ ions are reported in Table 2.3. The corresponding chromatogram and mass spectra of the plant extracts are given in Fig 2.14, Fig 4.5, Fig 4.6, Fig 2.17. The values of the first three peaks to be identified were those of epicatechin, isorhamnetin and beta-sitosterol, corresponding to retention time R_t = 0.67 min and m/z values of 289, 315 and 413, respectively. The next peaks to be identified were those of catechin, syringic acid and stigmasterol at R_t = 5.909 min and m/z values of 289, 197 and 411, respectively.

2.3 Results and Discussions

The metabolites were identified by the use of both 1D and 2D NMR studies, as well as through the comparison of the resulting chemical shift, coupling constant and integration values with those of typical NMR metabolite peaks obtained from databases and pure compounds. In order to double check the compound peaks, further 2D NMR experiments, including COSY, TOCSY, HSQC, and HMBC, were carried out in all the three samples (seeds, skin and pericarp) obtained from Momordica charantia fruit Fig 2.1. All the molecule peak assignments were made manually. Fig 2.2 shows representative 1D ^{1}H NMR spectra for the pericarp, skin and seed of Momordica charantia fruit, recorded at 600 MHz.

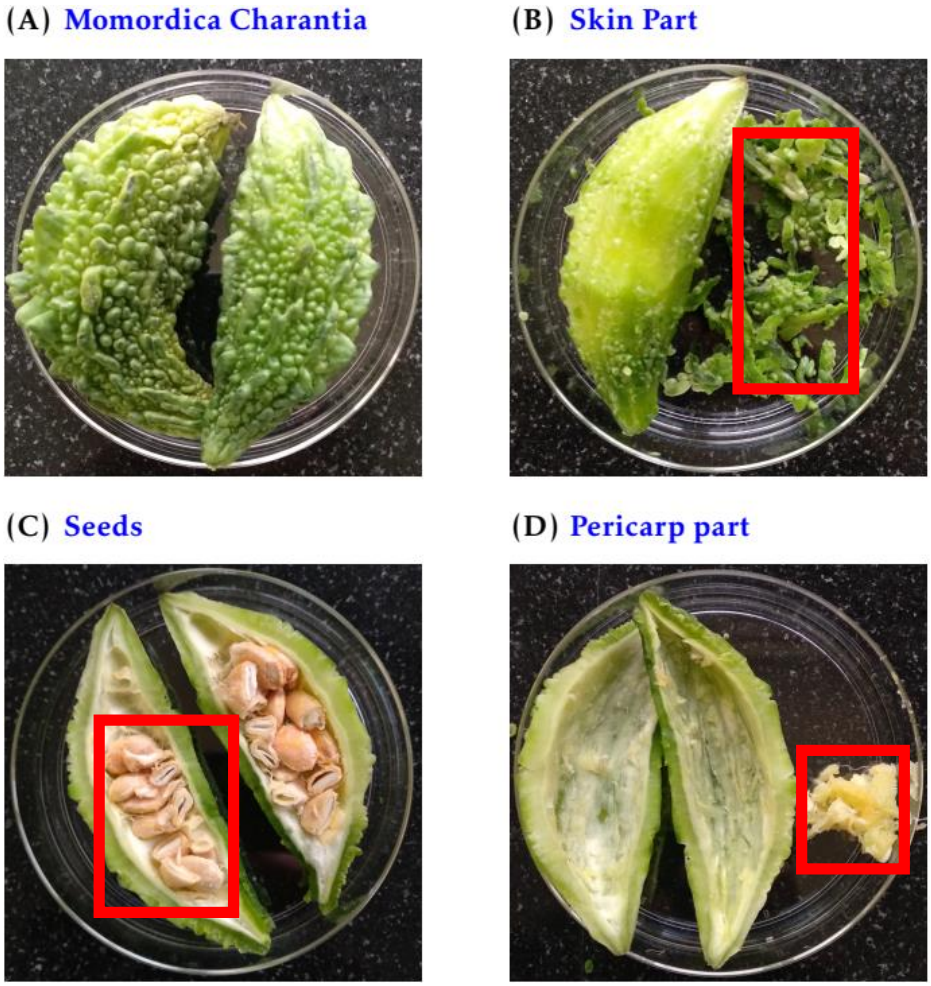

Figure 2.1: Different parts of Momordica charantia. (A) Fruit, (B) Removal of skin part, (C) Mature seeds, and (D) Removal of pericarp inside of the fruit.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

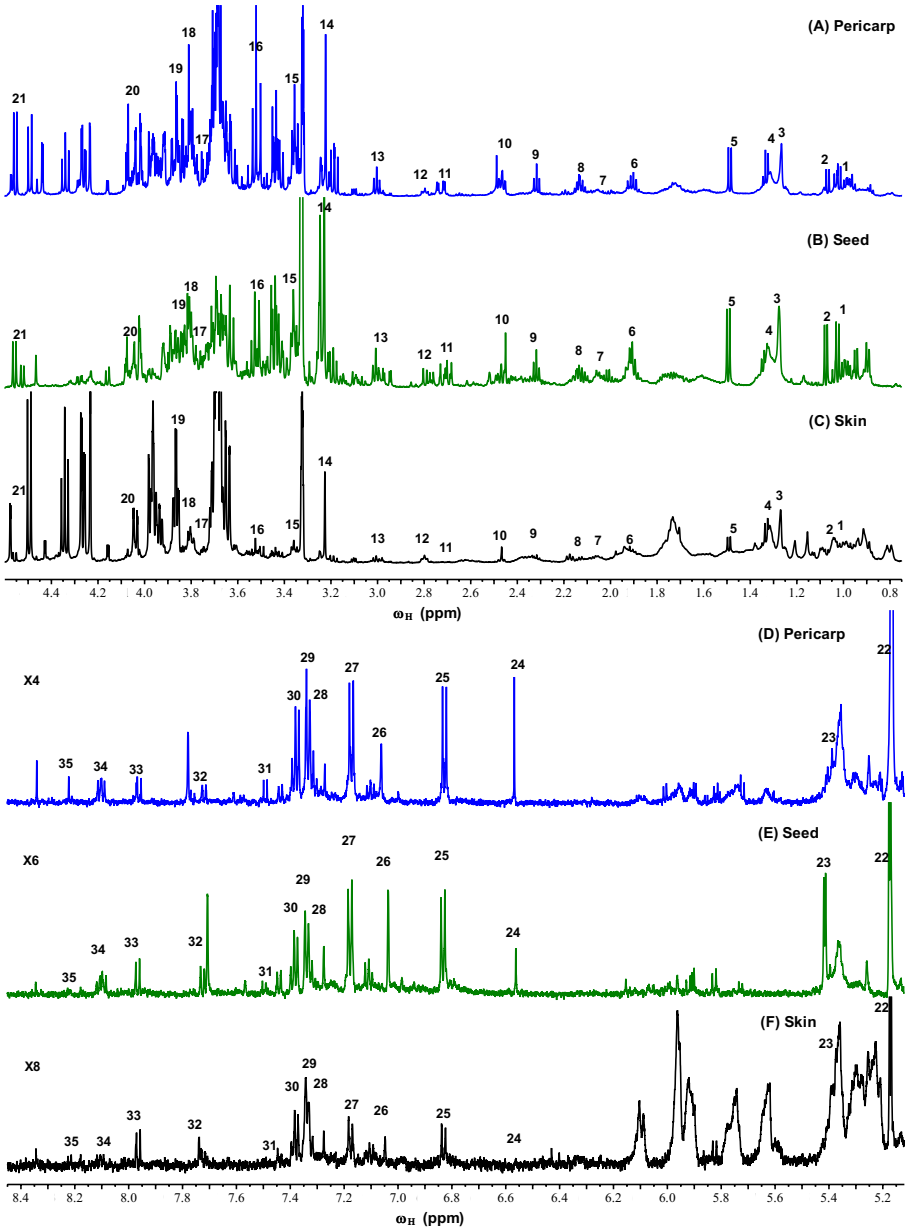

Figure 2.2: Representative ^{1}H NMR spectrum of Momordica charantia extracts: (A) & (D) pericarp, (B) & (E) seed and (C) & (F) skin, recorded at 600 MHz, showing identified metabolites. Peak labeling: 1, charantin; 2, valine; 3, lipid(-CH_2)n; 4, lactic acid; 5, alanine; 6, leucine; 7, chlorogenic acid; 8, glutamine/glutamic acid; 9, gamma-Aminobutyric acid (GABA); 10, epicatechin; 11, malic acid; 12, aspartic acid; 13, lysine; 14, choline; 15, inositol; 16, glycine; 17, arginine; 18, syringic acid; 19, threonine; 20, proline; 21, catechin; 22, alpha-glucose; 23, sucrose; 24, luteolin; 25, gentisic acid; 26, gallic acid; 27, tyrosine; 28, naringenin; 29, myricetin; 30, protocatechuic acid; 31, tryptophan; 32, trans-Cinnamic acid; 33, daidzein; 34, kaempferol; 35, adenine.

Table 2.1 lists all the primary metabolites identified in the three samples including lipids, amino acids, carbohydrates and organic acids, while Table 2.2 lists all the secondary metabolites identified included steroids, phenolics and flavonoids.

Table 2.1: Primary plant metabolites identified from ^{1}H NMR and 2D NMR spectra of Momordica charantia, with chemical shifts in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)	MSI Level
Term. methyl group	0.86(t,6.95), 0.95(t,6.9)	Level 2
(CH_2)n	1.27(m)	Level 2
$CH_2CH_2COOH(C_3)$	1.64(m)	Level 2
CH_2COOH (C_2)	2.37(t,7.5)	Level 2
Fatty acid	2.03(m)	Level 2
Fatty acid	5.3(m)	Level 2
Leucine	0.95(t,6), 1.9(m)	Level 2
Valine	0.98(d,7), 1.06 (d,7), 2.29(m)	Level 2
Alanine	1.49 (d,7.3), 3.77(q,7.2)	Level 2
Glutamine	2.15(m), 2.47(m)	Level 2
Glutamic acid	2.15(m), 2.44(m)	Level 2
Phenylalanine	3.0(m), 7.32(d,6.9)	Level 2
Tyrosine	6.86(m), 7.19(m)	Level 2
Tryptophan	7.32(s), 7.54(d,7.54)	Level 1
GABA	1.90(m), 2.30(t,7.2)	Level 1
Aspartic acid	2.79(dd,17.6), 3.88(dd,8.5)	Level 2
Threonine	3.82(dd,5.74)	Level 2
Glycine	3.52(s)	Level 2
Methionine	2.62(t,7.49), 2.13(m)	Level 2
Isoleucine	1.01(d,7.01), 0.96(t,7.62)	Level 2
Histidine	3.19(dd,14.5), 7.77(d,1.04)	Level 2
Lysine	3.02(t), 3.74(t,6.09)	Level 2
Arginine	1.68(m), 3.73(t,6.5)	Level 2
Proline	4.07(dd,8.56,6.4)	Level 2
Cysteine	3.98(dd), 3.03(m)	Level 2
Beta-glucose	4.58(d,7.8), 5.2(d,3.8)	Level 2
Alpha-glucose	4.63(d,7.9), 5.18(d,3.8)	Level 2
Sucrose	4.17(d,8.5), 5.4(d,3.8)	Level 1
Malic acid	2.68(dd,16,4.1), 4.27(dd,10.1,2.7)	Level 2
Fumaric acid	6.56(s)	Level 2
Formic acid	8.46(s)	Level 2
Quinic acid	1.84(dd), 1.94(m)	Level 1
Succinic acid	2.59(s)	Level 2
Lactic acid	1.32(d), 4.1(q)	Level 2
Adenine	8.11(s), 8.21(s)	Level 2
Inositol	3.26(t,9.3), 4.00(t,2.8)	Level 2
Methyl salicylate	3.92(s), 10.74(s)	Level 2
Choline	3.22(s), 3.50(dd,5.8,4.1)	Level 2

Abbreviations: GABA (gamma-aminobutyric acid), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

Table 2.2: Secondary plant metabolites identified from ^{1}H NMR and 2D NMR spectra of Momordica charantia, with chemical shifts in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)	MSI
Charantin∗	0.85(s), 0.95(s), 1.01(s), 1.09(s), 1.25(s)	Level 1
Momordicine	0.86-0.91(m), 6.06 (m), 1.25-1.35(m)	Level 2
Diosgenin	0.97(d,6.4), 1.02(s), 1.99(m), 3.36(t,10.5)	Level 2
Gallic acid	7.06(s)	Level 2
Protocatechuic acid	6.92(d,8.2), 7.36(dd,8.77,3.14), 7.40(d,2.1)	Level 2
Gentisic acid	6.82(d,8.58), 6.98(dd,8.77,3.14), 7.29(d,3.14)	Level 2
Catechin	4.55(d,7.90), 2.5(dd,16.07,8.1)	Level 2
Vanillic acid	7.43(dd), 3.90(s), 6.94(d,8.24)	Level 1
Chlorogenic acid	7.17(d,1.8), 4.25(d,2.58), 3.89(dd), 2.13(m)	Level 1
Syringic acid	7.03(s), 3.81(s)	Level 2
Epicatechin	4.28(m), 2.47(m), 2.91(m), 7.04(d,1.24)	Level 2
Trans-ferulic acid	7.10(dd), 3.88(s), 6.37(d,15.96)	Level 2
Benzoic acid	7.34(dd), 7.84(d), 7.54(t,7.33)	Level 2
Trans-cinnamic acid	7.77(m), 7.34(m), 6.52(d,16.06)	Level 2
Luteolin	6.55(s), 7.38(m), 6.20(d),6.43(d)	Level 2
Kaempferol	6.18(d,2.09), 8.10(m), 6.39(d,2.09), 6.90(m)	Level 2
Quercetin	6.82(d,8.57), 6.19(s), 6.41(s), 7.68(d,2.02)	Level 2
Naringenin	6.80(m), 7.30(m), 2.68(dd), 3.08(dd)	Level 2
Myricetin	6.18(d,2.09), 7.34(s), 6.37(d,2.09)	Level 2
Daidzein	7.96(d,8.75), 8.29(s), 6.81(m), 6.86(d,2.2)	Level 2
Isorhamnetin	7.72(d,2), 6.47(d), 9.74(s), 6.19(d)	Level 2

The asterisk $*$ symbol is showing that the compound charantin is a mixture of beta-sitosterol and stigmasterol together, s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

The steroid charantin reported in the literature is a 1:1 mixture of two steroidal saponins namely, β - sitosteryl glucoside and stigmasteryl glucoside [126] and we were able to separately identify the peaks of β-sitosterol and stigamasterol in the NMR spectra. The presence of all the metabolites identified from 1D ^{1}H NMR spectra was confirmed and validated from 1D ^{1}H NMR experiments on pure compounds Fig 2.3, Fig 2.4, Fig 2.5, Fig 2.6, Fig 2.7, Fig 2.8, Fig 2.9, Fig 2.10, by 2D NMR experiments Fig 2.11, Fig 2.12, Fig 2.13 from mass spectrometric analysis Fig 2.14, Fig 4.5, Fig 4.6, Fig 2.17 and Table 2.3 and previously reported literature.

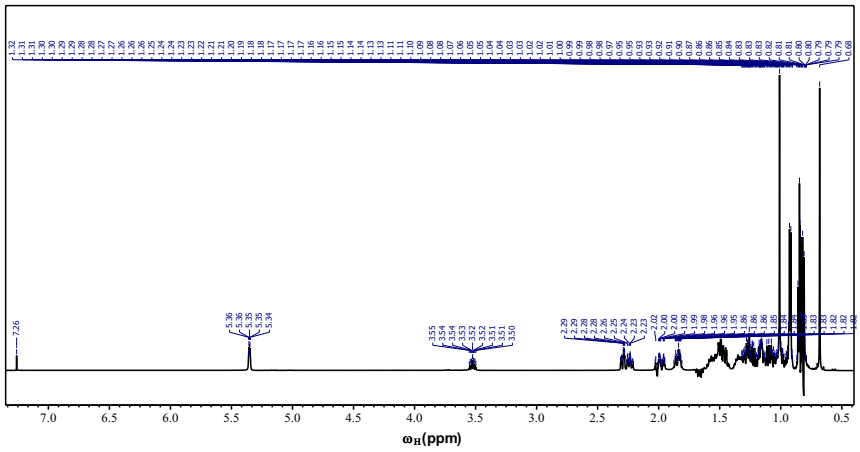

Figure 2.3: 1D ^{1}H NMR spectra of beta-sitosterol recorded at 600 MHz spectrometer.

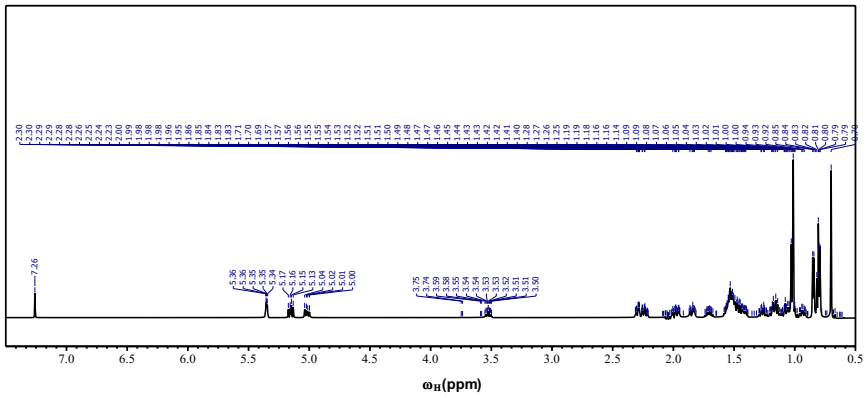

Figure 2.4: 1D ^{1}H NMR spectra of stigmasterol recorded at 600 MHz spectrometer.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

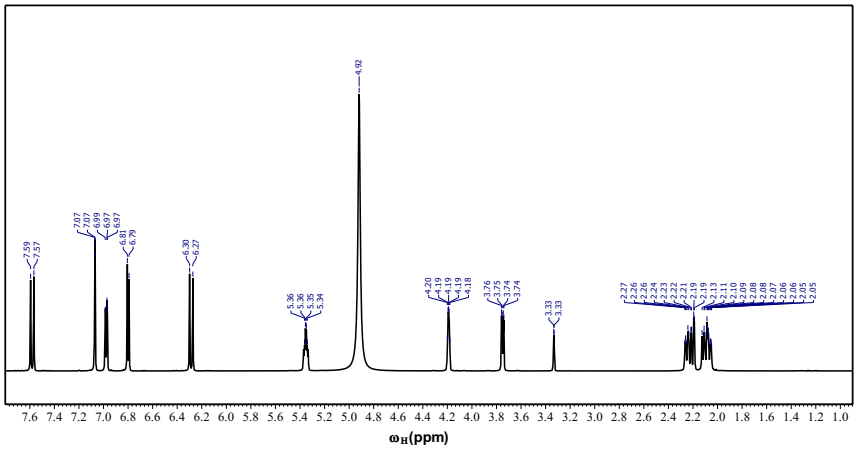

Figure 2.5: 1D ^{1}H NMR spectra of chlorogenic acid recorded at 600 MHz spectrometer.

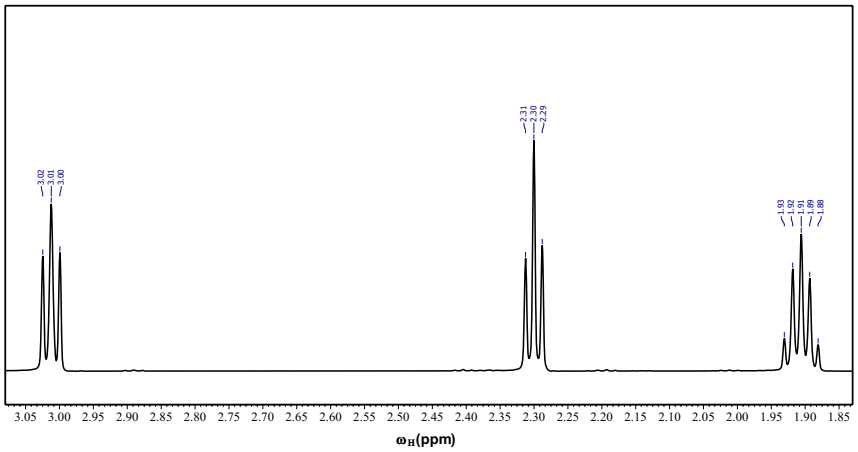

Figure 2.6: 1D ^{1}H NMR spectra of Gamma-Aminobutyric acid recorded at 600 MHz spectrometer.

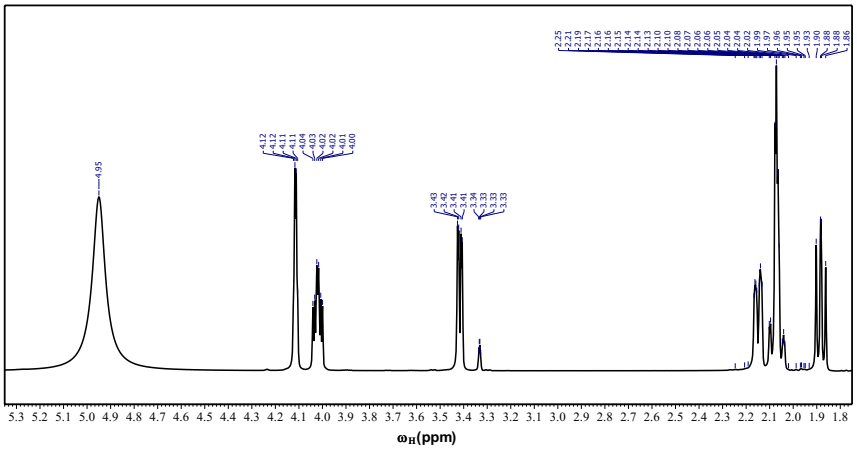

Figure 2.7: 1D ^{1}H NMR spectra of quinic acid recorded at 600 MHz spectrometer.

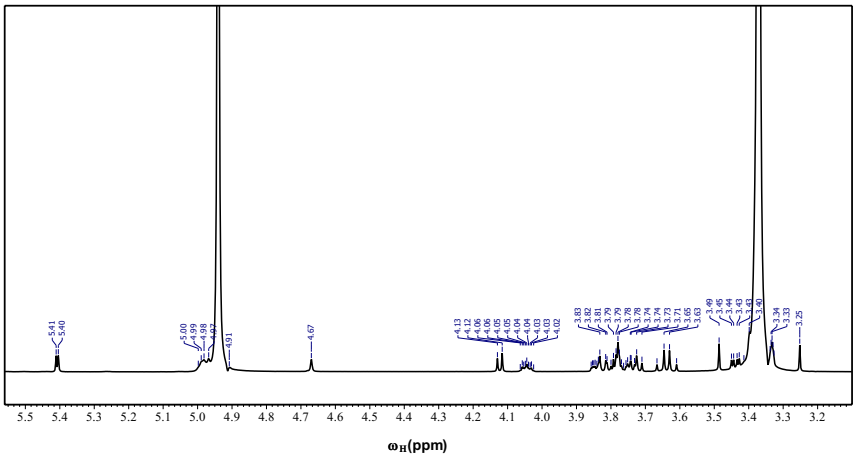

Figure 2.8: 1D ^{1}H NMR spectra of sucrose recorded at 600 MHz spectrometer.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

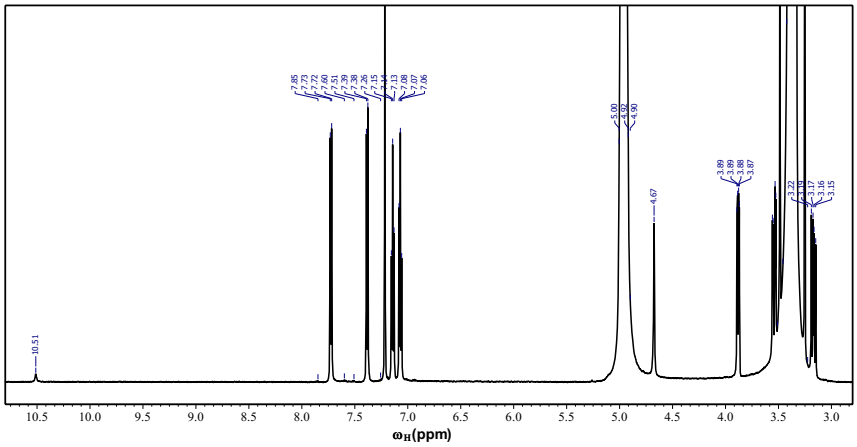

Figure 2.9: 1D ^{1}H NMR spectra of tryptophan recorded at 600 MHz spectrometer.

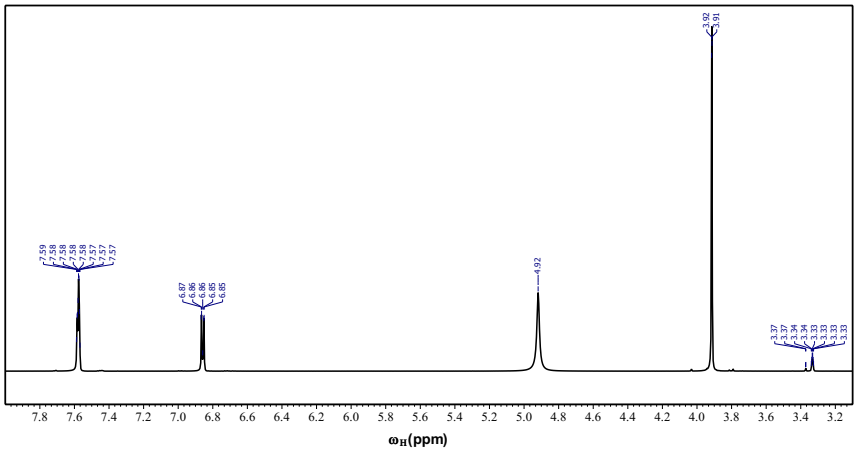

Figure 2.10: 1D ^{1}H NMR spectra of vanillic acid recorded at 600 MHz spectrometer.

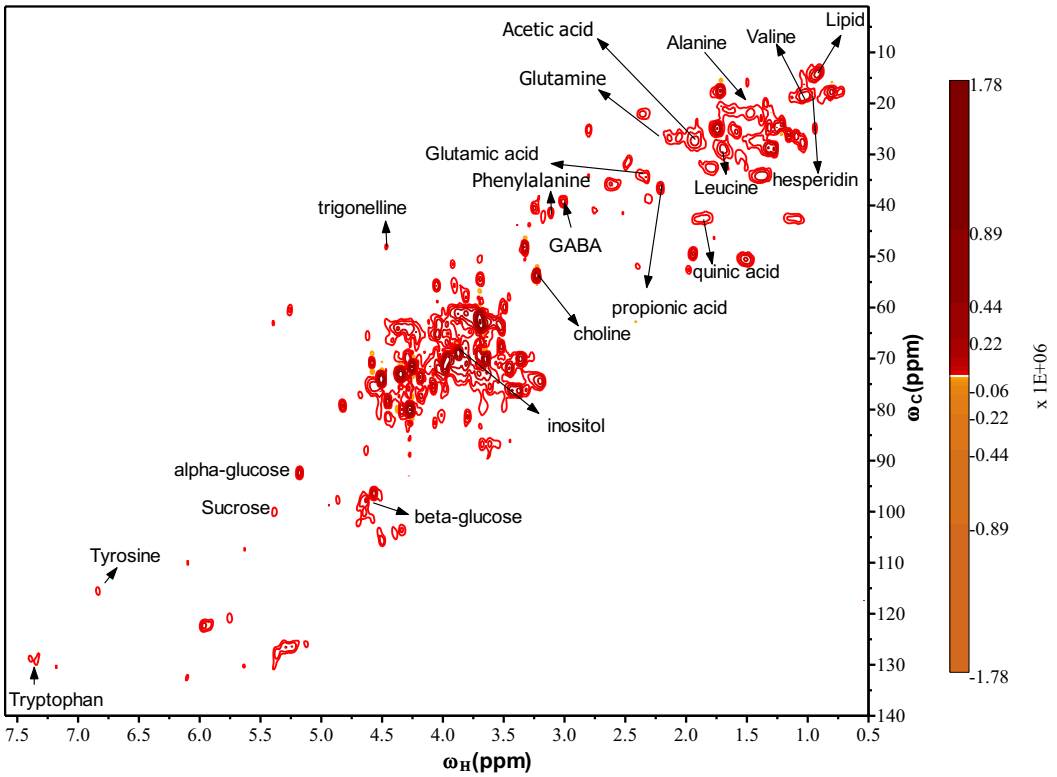

Figure 2.11: 2D ^{1}H-^{13}C HSQC NMR spectrum of Momordica charantia skin, recorded at 600 MHz. Proton chemical shift ω_H ppm is represented along X-axis and carbon chemical shift ω_C ppm) is represented along Y-axis. Spectra recorded with J value = 145.

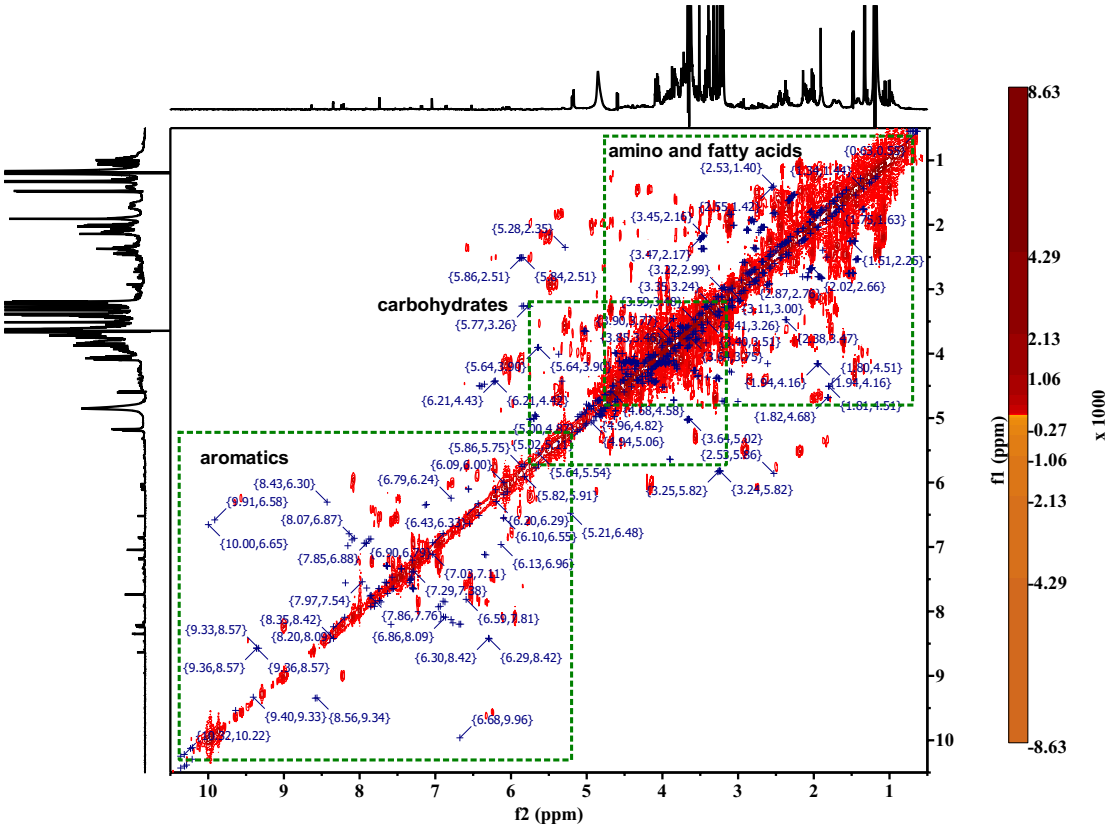

Figure 2.12: ^{1}H-^{1}H COSY NMR spectrum of Momordica charantia skin extract recorded at 600 MHz showing amino acid, lipids, carbohydrates and aromatics peaks. Proton chemical shifts ω_H ppm are represented along both spectral dimensions.

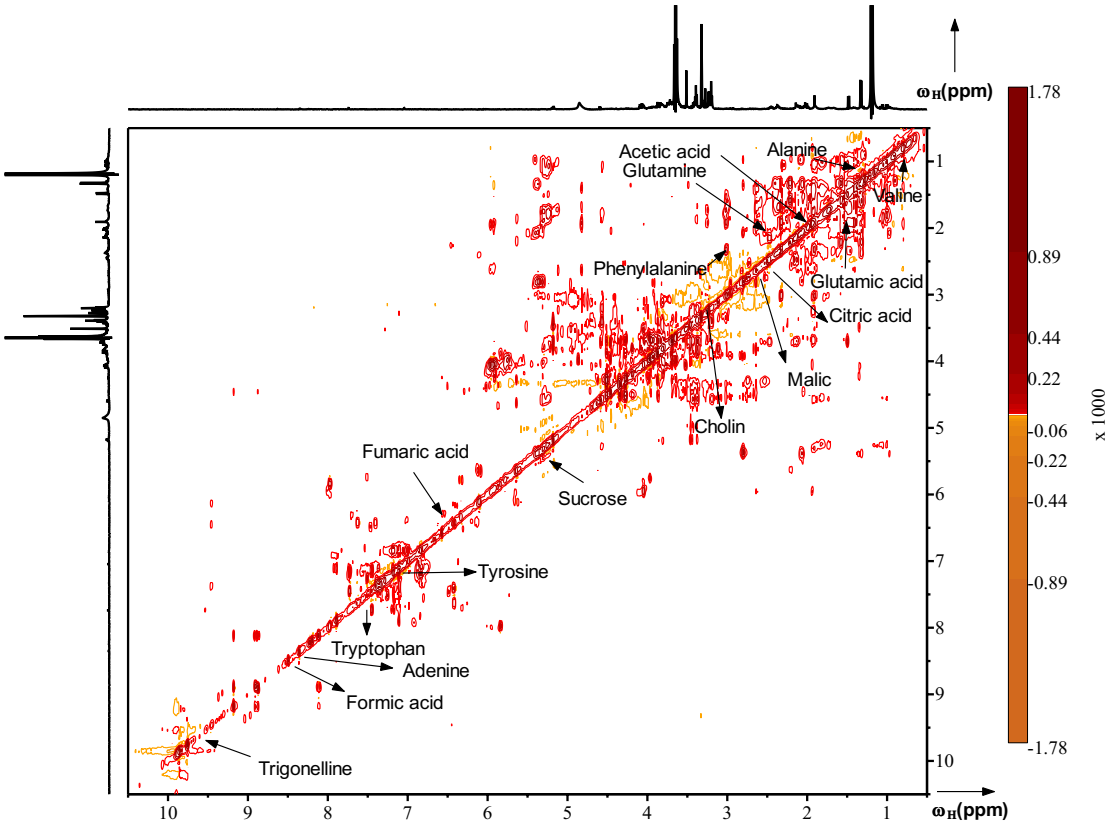

Figure 2.13: The two-dimensional TOCSY spectra of Momordica charantia skin extract in the region of δ 0.5 to 10.5 showing amino acid, lipids, carbohydrate and aromatic peaks.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

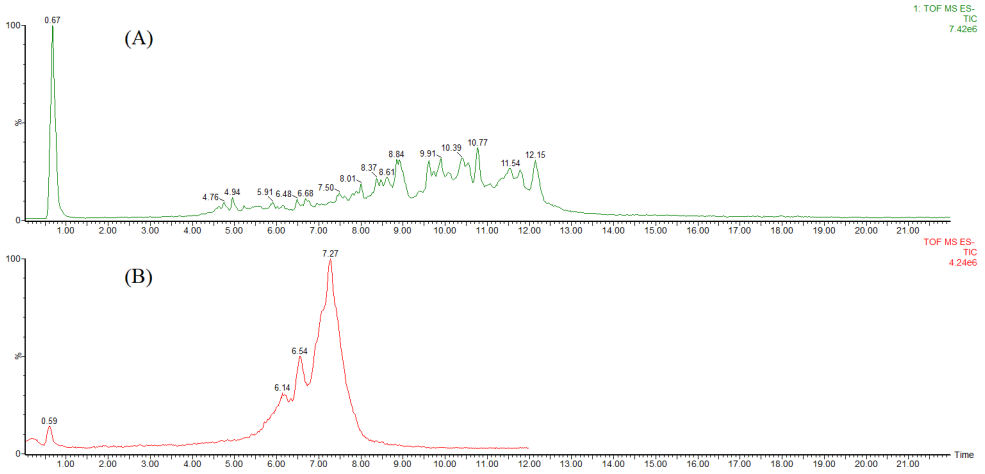

Figure 2.14: UPLC-MS chromatogram (A) shows the of Momordica charantia skin sample, whereas (B) shows the blank sample. These samples run in negative ESI mode.

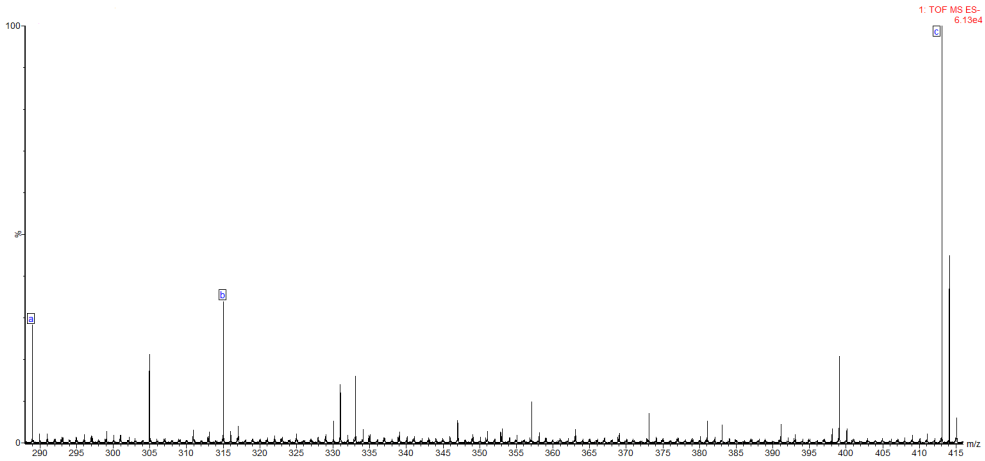

Figure 2.15: Mass spectra showing the detected metabolites masses $[M\text{-}H]^-$ ions: (a); epicatechin, (b); isorhamnetin and (c); beta-sitosterol.

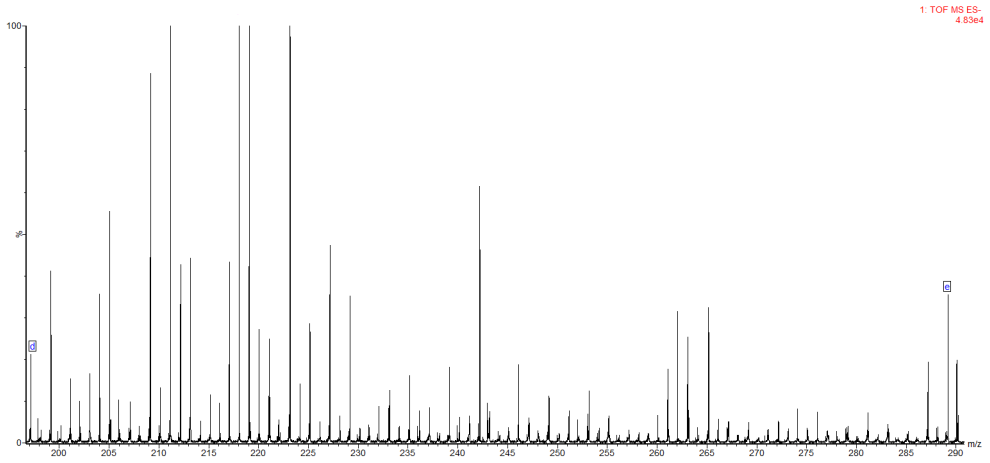

Figure 2.16: Mass spectra showing the detected metabolites masses $[M\text{-}H]^-$ ions: (d); catechin, and (e); syringic acid.

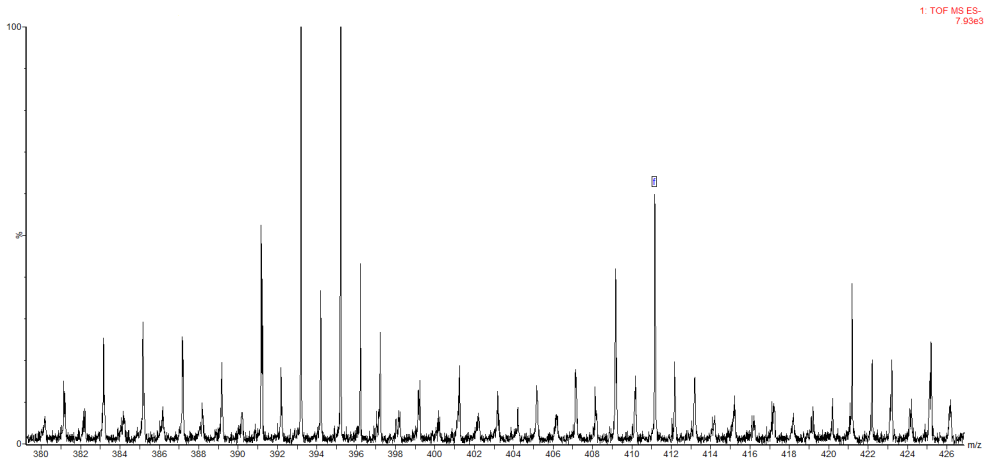

Figure 2.17: Mass spectra showing the detected metabolite mass $[M\text{-}H]^-$ ion: (f); stigmasterol.

Table 2.3: Compounds present in Momordica charantia plant extracts identified using UPLC-ESI-MS in negative ionization mode.

Metabolites	M.W.	$[M\text{-}H]^-$	R_t(min.)
Epicatechin	290	289	0.674
Isorhamnetin	316	315	0.674
Beta-sitosterol	414	413	0.674
Catechin	290	289	5.909
Syringic acid	198	197	5.909
Stigmasterol	412	411	5.909

Multivariate statistical analysis was performed to identify the metabolic differences between the seed, skin and pericarp samples. An initial comparison of all the three samples was performed using the unsupervised method of PCA and the grouping pattern and outliers were detected Fig 2.18, Fig 2.19 also confirmed from the polar dendrogram plotted for all the three groups Fig 2.20.

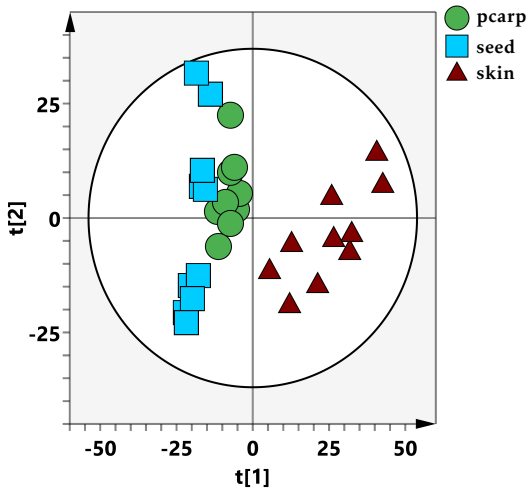

Figure 2.18: Principal component analysis (PCA) score plot of pcarp (pericarp), seed and skin part of Momordica charantia with component 1 explaining 70.1 % of the variation and component 2 explaining 16.6 % of the variation, showing a clear separation between these three groups.

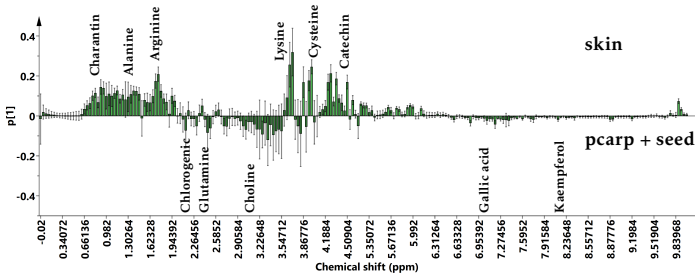

Figure 2.19: PCA column loading plot represent significant differences: The concentration of the following compounds charantin, alanine, arginine, lysine, cysteine and catechin are found to be more in skin part of Momordica charantia While chlorogenic acid, glutamine, choline, gallic acid and kaempferol are showing high concentration in pcarp + seed part of the fruit.

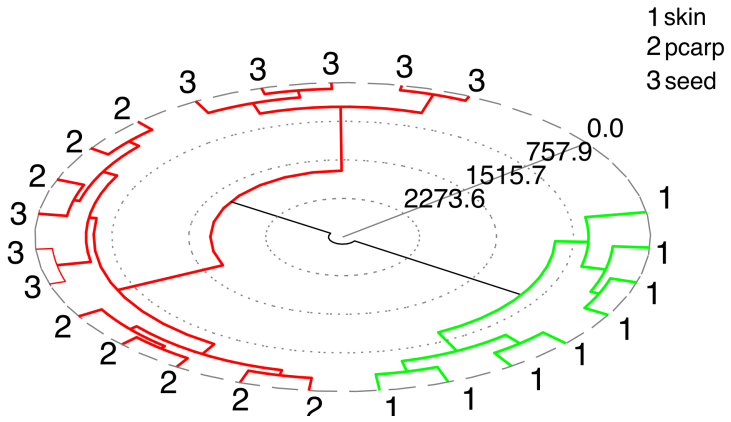

Figure 2.20: The polar dendrogram illustrates hierarchical clustering between all the three parts of Momordica charantia. Labelings are as follows: 1-Skin; 2-pcarp (pericarp); 3-seed.

Two-group comparisons were done using the supervised method of OPLS-DA. Fig 2.21 (a) shows the OPLS-DA score plot for comparison between the skin and pericarp samples, with one predictive and one orthogonal component showing a clear separation between the two groups, with t[1]=40.8% and to[1]=23.6%, respectively. Fig 2.21 (b) shows the corresponding S-plot showing the significant metabolites responsible for group separation. Similar OPLS-DA analysis was performed for comparisons of pericarp with seed and seed with skin samples Fig 2.22, Fig 2.23.

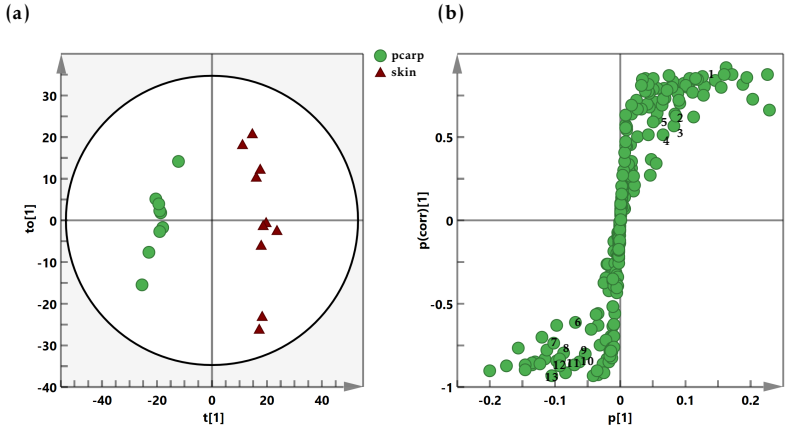

Figure 2.21: (a) OPLS-DA score plot obtained from 1D ^{1}H NMR spectra of skin and pcarp (pericarp). (b) S plot showing the metabolites responsible for the discrimination between the groups. The labels correspond to: 1, lipid methyl group (-CH_2)n; 2, lactic acid; 3, charantin; 4, lipid methyl group; 5, catechin; 6, proline; 7, arginine; 8, glutamic acid; 9, GABA; 10, malic acid; 11, lysine; 12, beta-glucose; 13, alpha-glucose.

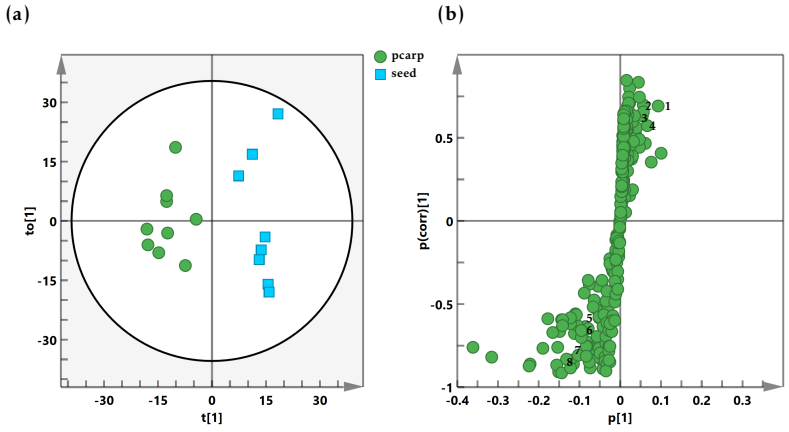

Figure 2.22: (a) OPLS-DA score plot obtained from 1D ^{1}H NMR spectra of seed and pcarp (pericarp). (b) S plot showing the metabolites responsible for the discrimination between the groups. Cutoff values for the covariance of $|p[1]| \geq 0.05$ and for the correlation of $|p(corr)[1]| \geq 0.5$ were used. The labels correspond to: 1, choline; 2, epicatechin; 3, phenylalanine; 4, lysine; 5, lactic acid; 6, methyl group (-CH_2)n; 7, alpha-glucose; 8, beta-glucose.

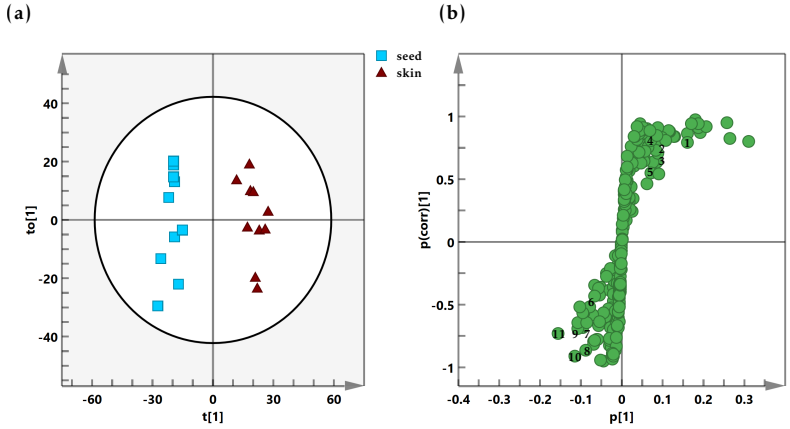

Figure 2.23: (a) OPLS-DA score plot obtained from 1D ^{1}H NMR spectra of seed and skin. (b) S plot showing the metabolites responsible for the discrimination between the groups. Cutoff values for the covariance of $|p[1]| \geq 0.05$ and for the correlation of $|p(corr)[1]| \geq 0.5$ were used. The labels correspond to: 1, threonine; 2, lactic acid; 3, charantin; 4, sucrose; 5, (-CH_2)n; 6, glycine; 7, glutamine; 8, lysine; 9, epicatechin; 10, choline; 11, inositol.

Testing with CV-ANOVA (p-value < 0.05) showed the models to be statistically significant and robust in nature. The significant metabolites were confirmed from the S-plot (abs(p(corr)[1]) > 0.5). Cutoff values for the covariance of $|$p[1]$| \geq 0.05$ and the correlation of $|$p(corr)[1]$| \geq 0.5$ were used. Fig 2.24 shows the relative concentration values of the metabolites that differ significantly in all the three groups of samples, identified from the univariate analysis of t-tests with a threshold of 0.05.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

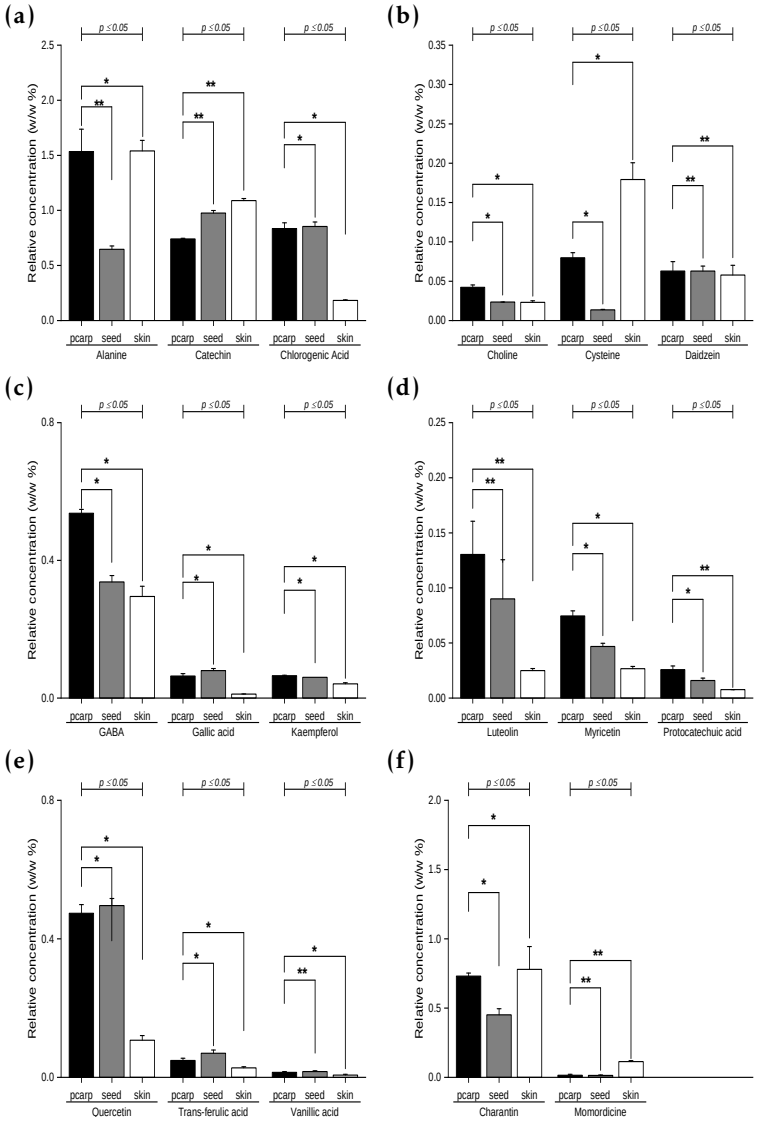

Figure 2.24: Relative concentrations (w/w %) of significantly different metabolites in Momordica charantia extracts (pericarp, skin and seed) as detected by ^{1}H NMR experiments. The label $*$ indicates a significant metabolite ($p \leq 0.05$), while the metabolites marked with $**$ are not significant.

Table 2.4 lists the metabolic pathways associated with statistically significant metabolites.

Table 2.4: List of metabolites with significantly altered concentrations in different parts of Momordica charantia plant extracts and their associated metabolic pathway.

Metabolic pathway	**Key metabolites**
Steroids metabolism	Charantin
	Momordicine
Phenylpropanoids and flavonoids metabolism	Catechin
	Chlorogenic acid
	Gallic acid
	Trans-ferulic acid
	Protocatechuic acid
	Vanillic acid
	Quercetin
	Luteolin
	Daidzein
	Kaempferol
	Myricetin
Amino acids metabolism	Alanine
	Cysteine
Choline metabolism	Choline

The three samples differ significantly in concentrations of medicinally important metabolites, confirming the different metabolic profiles of the three samples. Heat map and hierarchical cluster analysis (HCA) were used to look for natural groupings Fig 2.25. A univariate analysis was also performed to confirm the statistical significance of the metabolites identified to be different in all the three groups using one-way analysis of variance (ANOVA) with $p<0.01$, followed by post- hoc analysis (Fisher's LSD) Table 2.5.

2. NMR metabolomic profiling of phytomedicinal compounds in Momordica charantia

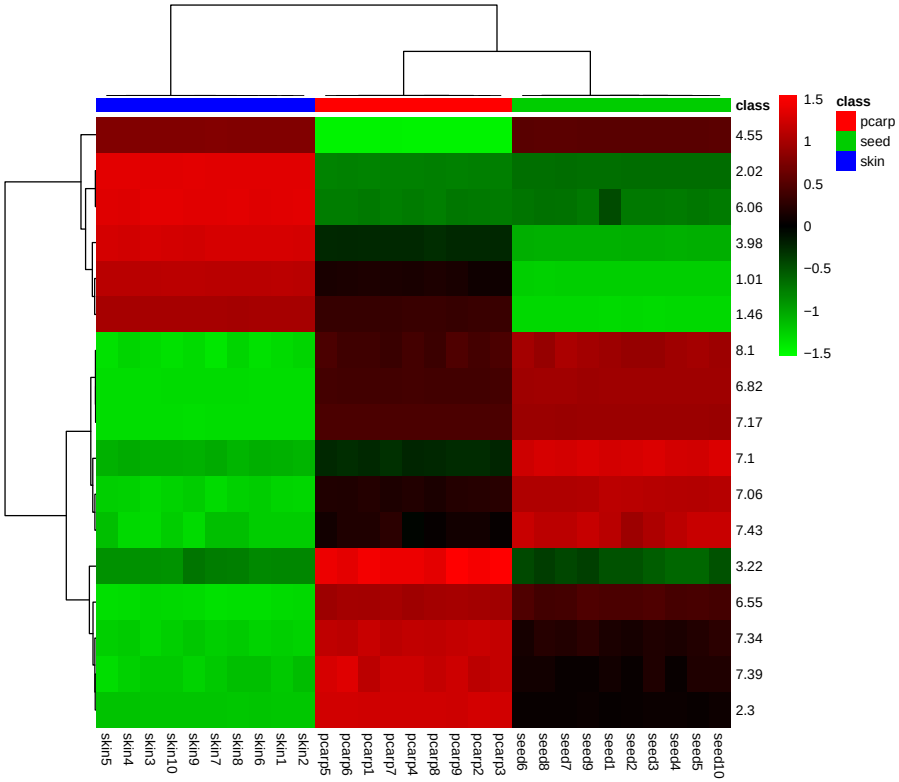

Figure 2.25: Dendrogram and heat map generated by hierarchical cluster analysis for significant metabolites identified from different parts of Momordica charantia (Rows: groups and Columns: metabolites). The two groups pericarp and seed were clustered together and are separated by the third group skin. Positive (red) or negative (green) correlations denote a increase or decrease in the metabolite concentrations, respectively. The hierarchical tree was constructed using Ward's minimum variance and Euclidean distance metric. The dendrogram is shown on the top of the heat map (pericarp-(red), seed-(green), skin-(blue)) and unveils the connection between the different groups based on their metabolite abundance levels.

Table 2.5: Post-hoc analysis showing which groups are different given the p-value threshold of 0.05 for seed, skin and pericarp.

Metabolite	p-value	-log$_{10}$(p-value)	FDR	Fisher's LSD (Post-hoc tests)
Quercetin	9.48E - 11	10.023	3.82E - 09	Seed - Pericarp; Pericarp - Skin; Seed - Skin
Kaempferol	1.27E - 09	8.8947	2.41E - 08	Seed - Pericarp; Pericarp - Skin; Seed - Skin
Alanine	1.64E - 07	6.7863	9.44E - 07	Skin - Pericarp; Skin - Seed
Chlorogenic acid	1.97E - 07	6.7061	1.08E - 06	Seed - Pericarp; Seed - Skin
Catechin	1.60E - 06	5.7947	7.22E - 06	Pericarp - Seed; Pericarp - Skin
Trans-ferulic acid	2.08E - 06	5.6819	9.00E - 06	Seed - Pericarp; Seed - Skin
Luteolin	3.78E - 06	5.4223	1.49E - 05	Pericarp - Skin; Seed - Skin
Protocatechuic acid	4.86E - 06	5.3130	1.86E - 05	Pericarp - Skin; Seed - Skin
Gallic acid	8.12E - 06	5.0903	2.90E - 05	Pericarp - Skin; Seed - Skin
GABA	1.43E - 05	4.8446	4.66E - 05	Seed - Pericarp; Pericarp - Skin; Seed - Skin
Vanillic acid	9.12E - 05	4.0401	2.47E - 04	Pericarp - Skin; Seed - Skin
Daidzein	3.86E - 04	3.4125	9.35E - 04	Pericarp - Skin; Seed - Skin
Myricetin	4.75E - 04	3.3232	1.09E - 03	Pericarp - Skin; Seed - Skin
Cysteine	1.10E - 03	2.9558	2.36E - 03	Pericarp - Skin
Beta-sitosterol	1.11E - 03	2.9535	2.36E - 03	Skin - Pericarp; Skin - Seed

We quantified and compared the total phenolic content of the different parts of the Momordica charantia fruit since several identified metabolites are phenolic compounds [127]. The seeds and pericarp have the highest phenolic content, followed by skin, respectively Table 2.6. The total flavonoid amount present in different parts of the Momordica charantia fruit is shown in Table 2.6, with the pericarp and seeds having the highest quantity, followed by skin. The statistical significance of the results was verified by one-way ANOVA and the results are in accordance with NMR data. The scavenging effects of methanol extracts of the mature fruit as measured by a DPPH assay [128] are in the order: seed $>$ pericarp $>$ skin (Table 2.7 & Fig 2.26 (b). The antioxidant potential of Momordica charantia fruit extracts was also quantified by an ABTS assay (Table 2.8 and Fig 2.26 (a)) demonstrating that the seeds and pericarp have better ABTS scavenging activity as compared to the skin. The α-glucosidase inhibition activity was determined at different concentrations, and it was observed that the α-glucosidase inhibition of pericarp was higher than its skin extracts at the same concentration Fig 2.26(c).

Table 2.6: Total phenolic and flavonoid content of seed, skin and pericarp of Momordica charantia. Data are expressed as mean $\pm$ SD.

	Seed	Skin	Pericarp
Phenolics (mg GAE / g DW)	324.88 $\pm$ 1.16	315.58 $\pm$ 1.16	320.23 $\pm$ 1.16
Flavonoids (mg QE / g DW)	151.85 $\pm$ 0.34	117.47 $\pm$ 3.47	161.46 $\pm$ 3.06

Table 2.7: Mean of EC$_{50}$(μg/mL) of DPPH radical scavenging activities of Momordica charantia fruit extracts (seed, pericarp and skin).

Extract	EC$_{50}$ (μg/mL)
Seed	94.21 $\pm$ 2.82
Pericarp	145.0 $\pm$ 4.35
Skin	245.0 $\pm$ 7.35

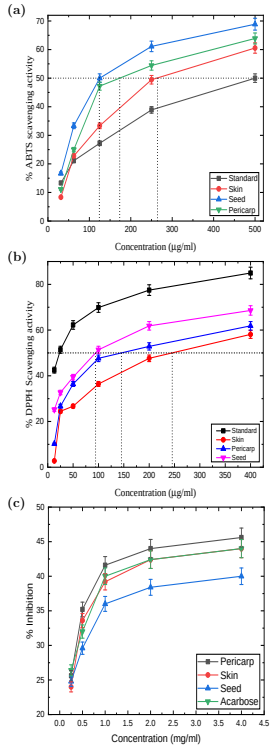

Figure 2.26: (a) Antioxidant potential of different parts of Momordica charantia (seed, pericarp and skin) from ABTS assay and using Trolox as a positive control. (b) Scavenging activity (%) of Momordica charantia extracts (seed, pericarp and skin) on DPPH radicals. (c) The percentage inhibition α-glucosidase activity of Momordica charantia extracts (pericarp, skin, seed) and acarbose (an anti-diabetic drug).

Table 2.8: ABTS assay measurements of antioxidant capacity of Momordica charantia fruit extracts (seed, pericarp and skin). Data are expressed as mean $\pm$ SD.

Extract	**EC_{50} (μg/mL)**
Extract EC_{50} (μg/mL) Seed	124.99 ± 3.74
Pericarp	173.02 ± 5.19
Skin	262.67 ± 7.88

A significant difference in concentrations of important metabolites was found in different parts of the fruit Fig 2.24. The pathway for the synthesis of these medicinally significant metabolites and their associated underlying metabolic pathways are shown in Fig 2.27 and Table 2.4. To determine the possible metabolic pathways, we carried out a metabolic pathway analysis by using the MetPA web-based tool Fig 2.28. Phytosterols are the products of the isoprenoid pathway and are synthesized from the

plant metabolite squalene [129], [130]. Our NMR analysis confirmed the presence of two steroidal glycosides – charantin and momordicine Fig 2.27, which have been previously reported to be useful in the treatment of diabetes due to their hypoglycemic activity and structural similarity to insulin. Our results demonstrate that the skin is the best source for the steroidal glycosides, as compared to the pericarp and seed. Several of the significant secondary metabolites identified in our study are phenolic compounds, which are known to possess antioxidant and anti-inflammatory properties. Vanillic acid and protocatechuic acid have been reported to have neuroprotective, antioxidant and antidiabetic effects [131], [132]. Our study also identified gallic acid, ferulic acid, daidzein and chlorogenic acid, all of which are synthesized from p-coumaric acid through a series of synthase and isomerase enzymatic reactions [133], [134], [135]. Other flavonoid metabolites which are synthesized from naringenin and have been identified in our study include luteolin, catechin, kaempferol, quercetin and myricetin [136], [137], [138], [139]. Our results show that these secondary metabolites are present in higher amounts in the seeds and pericarp, as compared to the skin. Choline is synthesized by the decarboxylation of the amino acid serine and has been reported to be an essential component for plant function [140]. Our study confirmed the presence of choline in Momordica charantia fruits, with the pericarp having the highest concentration of choline as compared to seeds and skin.

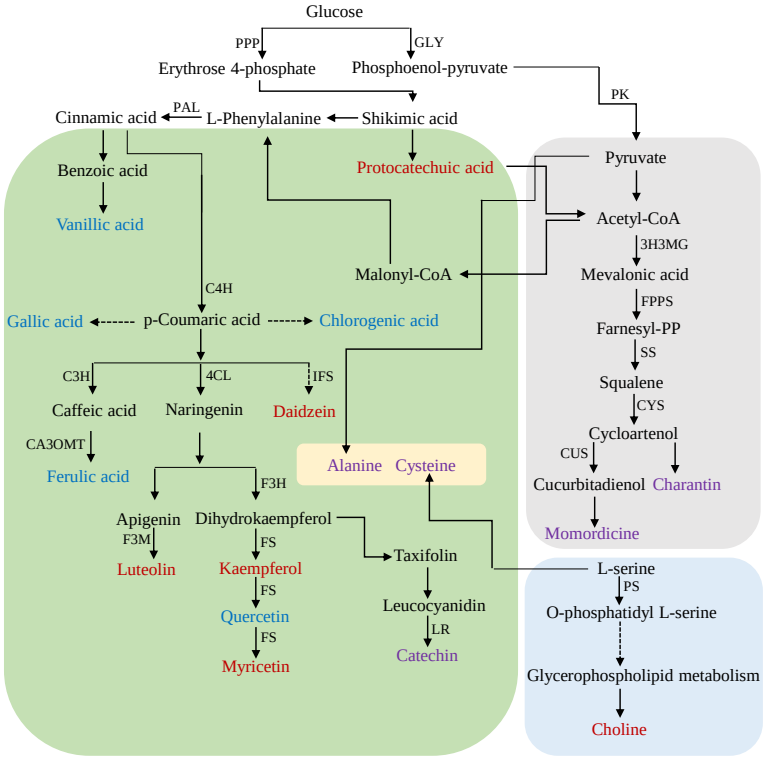

Figure 2.27: Metabolic pathway depicting the synthesis of significant secondary metabolites and linkage between them, identified via 600 MHz ^{1}H NMR experiments. Red, blue and purple colors indicate metabolites that are higher in concentration in pericarp, seed and skin respectively. Reactions considered are: GLY: glycolysis; PK: pyruvate kinase; PPP: pentose phosphate pathway; PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; C3H: coumarate 3-hydroxylase; CA3OMT: caffic acid 3-O-methyltransferase; 4CL: 4- coumaroyl-CoA ligage; IFS: isoflavone synthase; F3H: flavanone 3-hydroxylase; F3M: flavonoid-3'-monooxygenase; FS: flavonol synthase; LR: leucocyanidin reductase; 3H3MG: 3-hydroxy-3-methyl-glutyrl-CoA reductase; FPPS: farnesyl PP synthase; SS: squalene synthase; CYS: cycloartenol synthase; CUS: cucurbitadienol synthase; PS: phosphatidylserine synthase.

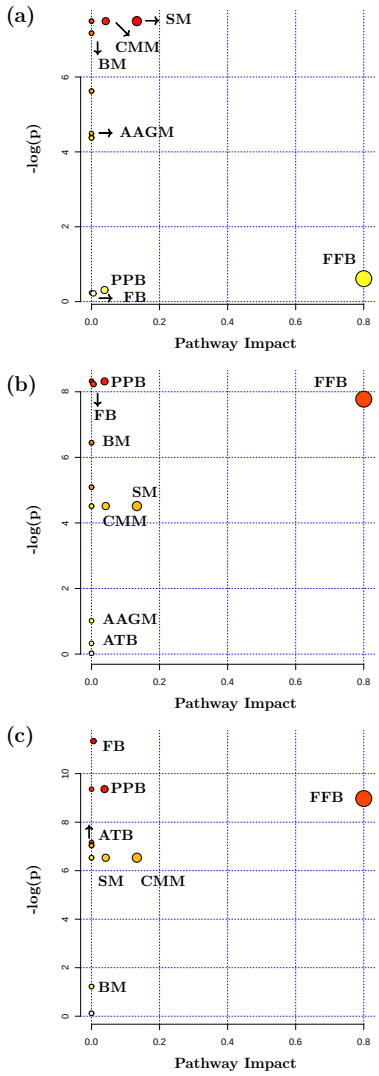

Figure 2.28: Metabolic pathway impact analysis was performed via MetaboAnalyst (MetPa) for (a) Pericarp and seed, (b) Pericarp and skin, (c) Skin and seed, showing the significant changes in all the groups. FB: Flavonoid biosynthesis, PPB: Phenylpropanoid biosynthesis, ATB: Aminoacyl-tRNA biosynthesis, SM: Sulfur metabolism, FFB: Flavone and flavonol biosynthesis, CMM: Cysteine and methionine metabolism, BM: Butanoate metabolism, AAGM: Alanine, aspartate and glutamate metabolism. The compound colors of each circle are based on p-values while the size of the circle gives an idea of the impact value within the metabolome pathway (larger circles imply greater impact). The yellow to red color range denotes different significance levels of the metabolites.

2.4 Conclusions

In summary, we used 1D and 2D NMR spectroscopy-based metabolomics combined with multivariate statistical analysis to investigate the phytochemical constituents of different parts of the Momordica charantia fruit. The pericarp and seeds of the fruit were metabolically characterized as having a higher concentration of antioxidants as compared to the skin. On the other hand, the skin had a higher concentration of steroids such as charantin and momordicine, as compared to the other parts of the fruit, implying that the skin has greater potential for use in plant-based antidiabetic drugs. Due to its reliability, reproducibility and accuracy, NMR-based metabolomics is a useful technique in elucidating the relationship between phytomedicinal components and their role in the remediation of the altered metabolic pathways in diseased organisms and our study is a step forward in this direction. These results have been published in **Natural Product Research, 36(1), 390-395 (2021)** (https://doi.org/10.1080/14786419.2020.1762190).

Chapter 3

NMR metabolomic profiling of phytomedicinal compounds in Phyllanthus emblica and Tinospora cordifolia

3.1 Introduction

Metabolomics is described as the study of all primary and secondary metabolites in cells, biofluids, tissues or organism under various situations [141]. Plants with medicinal properties are a gift from nature to humankind, given with the intention of fostering a disease-free and healthy life. Rasayana is the name given to a series of herbal preparations that are part of the Indian traditional health care system known as Ayurveda. These preparations contain a variety of medicinal plants that are thought to have noteworthy antioxidant effects. Phyllanthus emblica plant is utilized in many traditional medical systems, including Chinese, Tibetan, and Ayurvedic herbal medicine. Phyllanthus emblica (Amla), is also known as Indian gooseberries. It is extensively distributed in subtropical and tropical regions of China, India, Indonesia, and the Malay Peninsula [142]. According to reports, Phyllanthus emblica fruit has antioxidant [143], [144], hypolipidemic [145], [146], hypoglycemic [147], cancer chemopreventive [148], anti-mutation, anti-oxidation, anti-bacterial, anti-virus, anti-inflammatory, and anti-atherosclerosis properties [149], [150]. It is a key ingredient in several commercially available hepatoprotective formulations [151]. Phyllanthus emblica juice is an excellent source of vitamin C. Vitamin C is a water-soluble nutrient that plays an

important role in the body as an antioxidant [152]. Tinospora cordifolia (Giloy), often known as "Guduchi", is widely utilized in traditional Ayurvedic medicine for the treatment of numerous illness. Tinospora cordifolia is an anti-pyretic herb that can be used to treat fever. Dengue fever patients who take this treatment have a higher platelet count and a lower risk of developing complications. Tinospora cordifolia stem is reported to have medicinal properties such as anti-diabetic [153], anti-periodic [154], anti-inflammatory [155], anti-oxidant [156], and anti-malarial activities [157].

In order to describe fruit & stem parts, a metabolomics technique that makes use of ^{1}H NMR spectroscopy was utilized. We studied the metabolic profiles of Phyllanthus emblica fruit (without seeds) & Tinospora cordifolia stem using 1D & 2D NMR spectroscopy. The metabolites in Phyllanthus emblica and Tinospora cordifolia extracts include amino acids, sugars and organic acids. We confirmed the phenolics & flavonoids constituents in Phyllanthus emblica and Tinospora cordifolia extracted with polar and non-polar solvents. A supervised OPLS-DA (Orthogonal partial least squares discriminant analysis) analysis was performed to identify the metabolic differences between the raw fruit, stem and Patanjali products.

3.2 Experimental Methods

3.2.1 Fruit & stem material

Phyllanthus emblica fruits were bought from the market as shown in Fig 3.1. Six fully mature fruits of remarkably similar size were chosen for the study. Fig 3.2 represent the stem samples of Tinospora cordifolia were collected at the IISER Mohali campus. The juices of Phyllanthus emblica (Amla) and Tinospora cordifolia (Giloy) were received from Patanjali Ayurveda, as depicted in the figures Fig 3.3 & Fig 3.4, respectively.

Figure 3.1: Phyllanthus emblica (Amla)

Figure 3.2: Tinospora cordifolia stem

Figure 3.3: Patanjali Ayurvedic product - Phyllanthus emblica juice

Figure 3.4: Patanjali Ayurvedic product - Tinospora cordifolia juice

3.2.2 Chemicals

All of the extraction solvents, reagents, NMR reference standards, and deuterated solvents that were procured from Sigma Aldrich India and possessed an analytical quality with a purity of more than 90 $\%$. Deuterium oxide D_2O, deuterated methanol-D4 CD_3OD and phosphate buffer solution (2.0 mM, pH 7.4) were used as the solvents for the NMR spectroscopy. To serve as an internal standard for all of the ^{1}H NMR studies, trimethylsilyl propionic acid sodium salt (abbreviated as TMSP) was added. To inhibit the growth of the bacteria sodium azide (NaN_3, 99.5$\%$) was utilized. Standard extraction procedures [120] were followed in order to extract metabolites from fruits & stems, and the process was carried out at room temperature utilizing CD_3OD and D_2O as the extraction solvents.

3.2.3 Sample preparation for NMR

Six replicates of each sample type, including fruit and stem, were analyzed. The surface of the fruit and stem were cleaned by distilled water in order to eliminate any debris, and then it was dried with tissue paper. The fruit and stem were sliced into thin pieces. Thin pieces were then ground with a mortar and pestle in liquid nitrogen and stored in different airtight plastic containers. These samples were then frozen overnight at -80 ^{0}C. After that, each of the samples were lyophilized in a freeze dryer for 16 hours at a temperature of -40 degrees Celsius and at a pressure of 30 Pascal. Two different extraction methods were used. Extraction method 1: Polar metabolites extraction (30 mg of dried fruit and stem powder was combined with 500μL methanol-d4+ 300μL D_2O extraction solvents: 400 μL taken as supernatant for freeze drying). After freeze drying, we added 600 μL solution (phosphate buffered saline, pH 7.4 made in D_2O, 100μL of sodium azide to prevent bacterial growth, 2 mM trimethylsilane propionic acid sodium salt (TMSP) as an internal reference) to all eppendorf tubes. These samples were vortexed and centrifuged at 5000 g for 10 minutes and 500 μL of the filtered supernatant was used for the NMR experiments. Extraction method 2: Non-Polar metabolites extraction (60 mg of each sample was combined with 400 μL methanol-d4 + 400 μL D_2O + 800 μL $CDCl_3$ in the ratio of 1:1:2). After the centrifugation process, we got the two layers in the eppendorf tube. The upper and lower layers in the tube contained the polar and non-polar metabolites. The 600 μL upper part of the eppendorf tube was carefully removed using the micro-pipette and 600 μL of the lower part was used for the analysis. Finally, 500 μL supernatant was transferred to 5 mm NMR tube and used for NMR analysis. All sampling were conducted at the same time and with the same controls.

3.2.4 NMR spectroscopy

One-dimensional (1D) ^{1}H NMR spectra of fruit & stem extracts were obtained using a Bruker AVANCE III 600 MHz NMR Spectrometer equipped with a 5mm TXI probe. Prior to the signal acquisition, gradient shimming was carried out. D_2O and $CDCl_3$ were used in the role of an internal lock. For ^{1}H NMR spectra acquisition, a water suppressed Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequence was used to remove the macromolecular signal with the following parameters: 7.55 μs 90 pulse, a 4 s relaxation delay, a 12 ppm spectral width, a 4.54 s acquisition time, and 16 transients of 64k data points. Before performing the Fourier transformation, the data were first padded with zeros by a factor of two, and then the FIDs were multiplied by an exponential weighting function that was equivalent to a line broadening of 0.3 Hz. The NMR spectra were corrected for phase and baseline, and trimethylsilane propionic acid sodium salt (TMSP) was added as an internal NMR reference at 0.00 ppm. 2D NMR experiments were performed to confirm peak assignments, including homonuclear ^{1}H-^{1}H correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and heteronuclear ^{1}H-^{13}C coherence spectroscopy (HSQC). 2D COSY and TOCSY spectra with a spectral width of 12 ppm in both proton F1 and F2 dimensions, 2K data points, 16 scans, and 256 t1 increments were acquired. With 2K data points, 16 scans, and 128 t1 increments, 2D HSQC spectra were obtained with spectral widths of 12 ppm and 180 ppm in the proton and carbon dimensions, respectively.

3.2.5 Metabolite Analysis

The 1D and 2D NMR experiments were used to identify the metabolites, and they were also identified by comparing the chemical shift and coupling constant values with typical NMR metabolite peaks from databases like the Biological Magnetic Resonance Data Bank (BMRB), the Madison Metabolomics Consortium Database (MMCD), and the Human Metabolite Data Base (HMDB).

3.2.6 Multivariate data analysis

The icoshift (interval correlation shifting) [119] MATLAB algorithm was used to align the NMR data after being translated to ASCII format. MestReNova 10.0.2-15465 (Mestrelab Research, Spain) was used to analyze NMR data. In the analysis of the NMR data, the significant chemical shift region, which ranged from 0.7 ppm to 8.5 ppm, was divided into many smaller bins of equal width (0.04 ppm bin size) referenced to a peak of TMSP. A total integral of 100 was used to standardize the data. The spectral region δ4.70-5.10 ppm had residual water peak and was therefore excluded from

the analysis. Orthogonal Partial least-squares discriminant analysis (OPLS-DA), a supervised multivariate analysis of the binned NMR data was conducted using SIMCA version 14.1.0-2047 software (Umetrics, Umea, Sweden). The values of R2Y (variance predicted) and Q2 (variance explained) were used to conduct an analysis that determined the accuracy of the OPLS-DA model. The OPLS-DA scores plot, which has one predictive and one orthogonal component, was used to observe the separation between the groups.

3.3 Results and Discussions

We used NMR data to characterize various primary and secondary metabolites in Phyllanthus emblica (Amla) and Tinospora cordifolia (Giloy) extract. Briefly, metabolite peaks were identified on comparisons with typical compound chemical shifts values and 2D-NMR investigations such as ^{1}H-^{1}H COSY (correlation spectroscopy), ^{1}H-^{1}H TOCSY (Total Correlation Spectroscopy), ^{1}H-^{13}C HSQC (heteronuclear single quantum coherence spectroscopy), and ^{1}H-^{13}C HMBC (heteronuclear multiple bond correlation). The ^{1}H-NMR spectra of raw extract and juice samples exhibited two main regions, an up-field region (δ 0.5–5.5 ppm) for high intensity signals belonging largely to primary metabolites i.e., fatty acids, amino acids and sugars and a down-field region (δ 5.5-9.0 ppm) for low intensity signals belonging mostly to secondary metabolites including some flavonoids and phenolic acids. The presence of oleic fatty acid as a significant unsaturated fatty acid in samples is an important trait. The signals of the terminal methyl group resonating at δ 0.90 ppm and that of the methylene envelope resonating at δ 1.28–1.35 ppm were observed in the proton NMR spectrum. Fig 3.5 shows representative 1D ^{1}H NMR spectra for the Phyllanthus emblica fruit extract, recorded at 600 MHz NMR spectrometer. Table 3.1 lists all the metabolites identified in the sample including fatty acids, amino acids, carbohydrates, organic acids, phenolics and flavonoids.

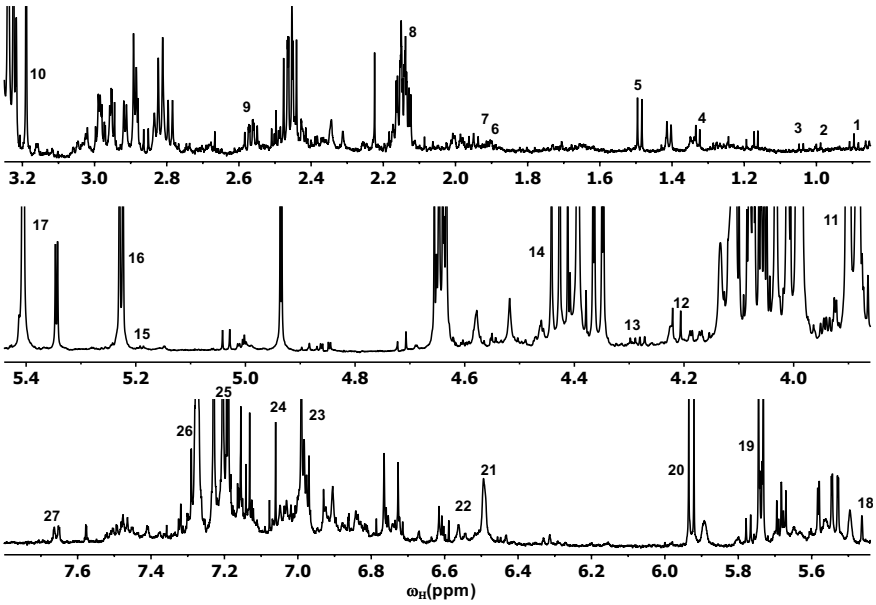

Figure 3.5: Representative ^{1}H NMR spectrum of Phyllanthus emblica (Amla) extracts recorded at 600 MHz, showing identified metabolites. Peaks numbering: 1, Fatty acid methyl ester; 2, Leucine; 3, Isoleucine; 4, Valine; 5, 3-hydroxybutyric acid; 6, Alanine; 7, Arginine; 8, Acetate; 9, Glutamine; 10, GABA; 11, Malate; 12, Phenylalanine; 13, Choline; 14, Proline; 15, Sucrose; 16, alpha-glucose; 17, beta-glucose; 18, sucrose; 19, fumarate; 20, tyrosine; 21, quercetin; 22, chlorogenic acid; 23, ellagic acid; 24, tryptophan; 25, uridine; 26, adenine; 27, formate.

Table 3.1: Metabolites identified from ^{1}H NMR and 2D NMR spectrum of Phyllanthus emblica fruit, with chemical shift given in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)
Term. methyl group	0.89(t,6.95), 0.95(t,6.9)
Isoleucine	0.89(t)
Leucine	0.95(t,6)
Valine	1.04(d,7)
Lactic acid	1.32(d), 4.1(q)
Alanine	1.48(d,7.35)
Arginine	1.90-1.96(m)
GABA	1.94(m)
Glutamine	2.15(m)
Glutamic acid	2.15(m)
Succinic acid	2.58(s)
Choline	3.22(s)
Methyl salicylate	3.90(br,s)
Quinic acid	4.21(d,8.5)
Malic acid	4.28(dd,10,3), 2.68(dd,15.4,3.4)
Trigonelline	4.44(s)
Glucose	5.17(d,3.7)
Alpha-glucose	5.18(d,3.8), 4.63(d,7.98)
Beta-glucose	5.22(d,3.7)
Sucrose	5.4(d,3.8)
Geraniin	5.45(br,s), 6.59(br,s), 6.71(s)
p-coumaroyl glucarate	5.74(dd,2.0,7.3), 4.37(d,2.7)
Catechin	5.92(d)
kaempferol	6.50(d,2)
Fumaric acid	6.56(s)
Taxifolin	6.98(d,6.1)
Gallic acid	7.06(s)
Chlorogenic acid	7.18(d,2.1), 6.98(d,8)
Ferulic acid	7.18(d,2)
Ellagic acid	7.27(s)
Quercetin	7.65(dd,8.4,2.1)

Abbreviations: GABA (gamma-aminobutyric acid), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

In addition, we recorded the proton NMR spectrum for Patanjali juice product, which we then compared to the Phyllanthus emblica raw fruit extract. The results of this comparison are presented in the figure Fig 3.6, Fig 3.7. We were successful in identifying a number of compounds that were present in both samples as shown in the figure Fig 3.6, Fig 3.7. We found some impurities and additive peaks in the proton NMR spectra of Patanjali juice product. Sodium benzoate peak observed at 7.52(s) ppm in the proton spectra of Patanjali juice product confirmed a preservative. Because of the overlap of NMR peaks, it was difficult to identify the compounds using 1D or even 2D NMR spectra. This was notably true for the ^{1}H spectral area between δ 3.2-4.5 ppm, which is dominated by carbohydrates and their derivatives. The chemical shift, scalar coupling, and integration values for some of the compounds in the downfield region could not be matched by us. So we were not able to assigned those peaks. Also the assignments of some compound peaks were not possible due to noise in the downfield region of the proton spectra.

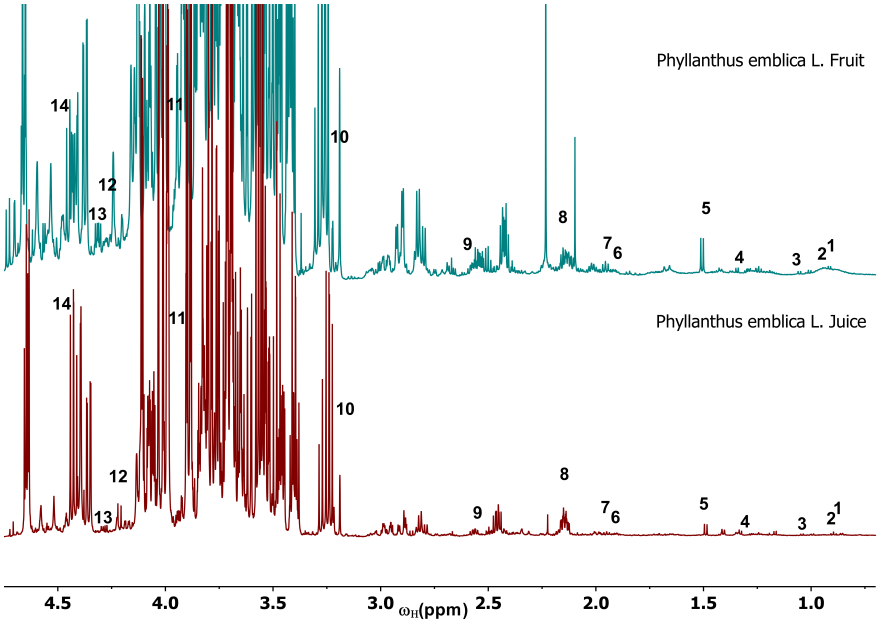

Figure 3.6: Representative ^{1}H NMR spectrum of Phyllanthus emblica (Amla) raw fruit and Juice (Patanjali Product) extracts recorded at 600 MHz, showing identified metabolites. Proton peaks Numbering: 1, Isoleucine; 2, Leucine; 3, Valine; 4, Lactic acid; 5, Alanine; 6, Arginine; 7, GABA; 8, Glutamine; 9, Succinic acid; 10, Choline; 11, Methyl salicylate; 12, Quinic acid; 13, Malic acid; 14, Trigonelline.

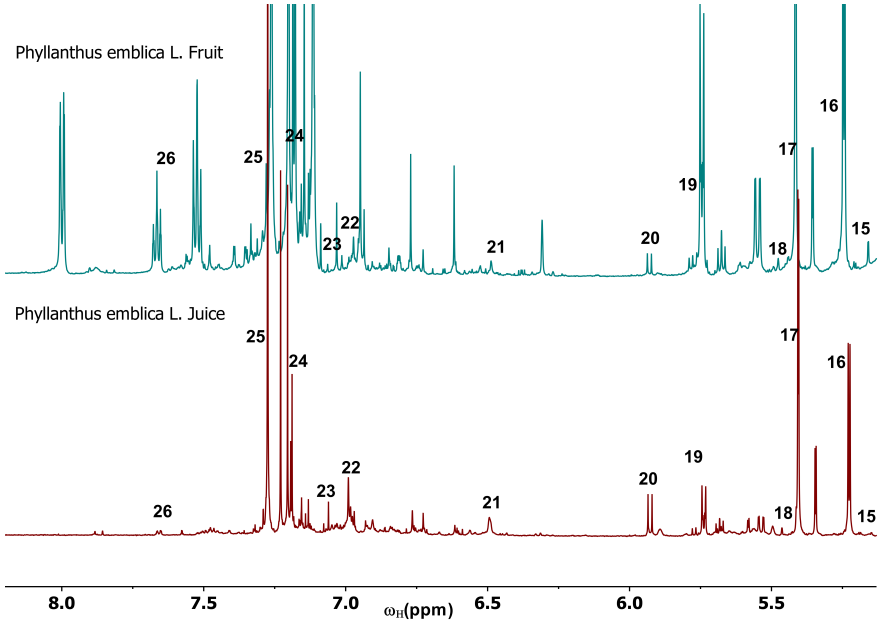

Figure 3.7: Representative ^{1}H NMR spectrum of Phyllanthus emblica (Amla) raw fruit and Juice (Patanjali Product) extracts recorded at 600 MHz, showing identified metabolites. Proton peaks Numbering: 15, Alpha glucose; 16, Beta-glucose; 17, Sucrose; 18, Geraniin; 19, p-Coumaroyl glucarate; 20, Catechin; 21, Kaempferol; 22, Taxifolin; 23, Gallic acid; 24, Chlorogenic acid; 25, Ellagic acid; 26, Quercetin.

Multivariate statistical analysis was performed to identify the metabolic differences between the Phyllanthus emblica fruit extract and Patanjali juice product. Comparisons between the two groups were carried out using the supervised OPLS-DA approach. Fig 3.8 (A) shows the OPLS-DA score plot for comparison between the Phyllanthus emblica fruit extract and Patanjali juice samples, with one predictive and one orthogonal component showing a clear separation between the two groups, with t[1]=70.9% and to[1]=6.5%, respectively. The quality of the OPLS-DA model was assessed from the R2X=0.961, R2Y=0.998 (variance predicted) and Q2=0.996 (variance explained) values. The corresponding S-plot, which illustrates the important metabolites that are responsible for group separation, is shown in figure Fig 3.8 (B). Cut-off values for the covariance of $\mid p[1] \mid \geq 0.05$ and for the correlation of $\mid p(corr)[1] \mid \geq 0.5$ were used. We found the natural active component of Phyllanthus emblica fruit as gallic acid and ellagic acid.

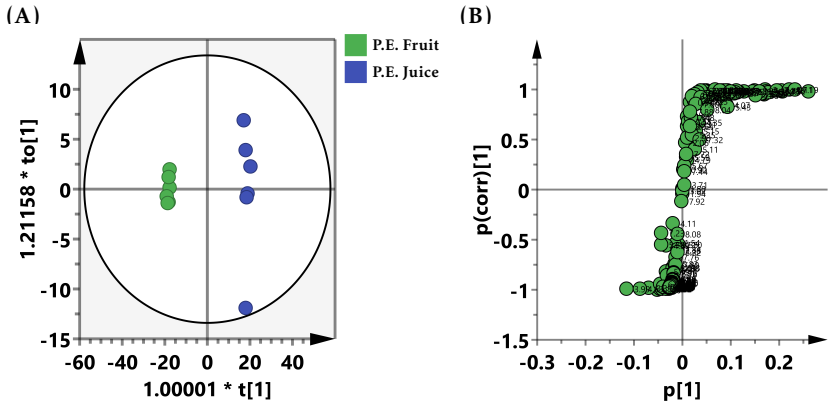

Figure 3.8: (A) OPLS-DA Score plot from ^{1}H NMR, showing clear separation of Phyllanthus emblica Fruit and Juice (Patanjali Product), (B) S-plot giving the information about the significant metabolites which were responsible between these two groups.

A wide range of metabolites were found and confirmed to be present in Tinospora cordifolia stem extract recorded at 600 MHz shown in Fig 3.9. The ^{1}H NMR spectra of giloy stems are marked by the presence of major soluble metabolites. Carbohydrates play the most important role, while organic acids and amino acids are present in smaller amounts. Table 3.2 lists all the metabolites identified in the sample including fatty acids, amino acids, carbohydrates, organic acids, phenolics and flavonoids.

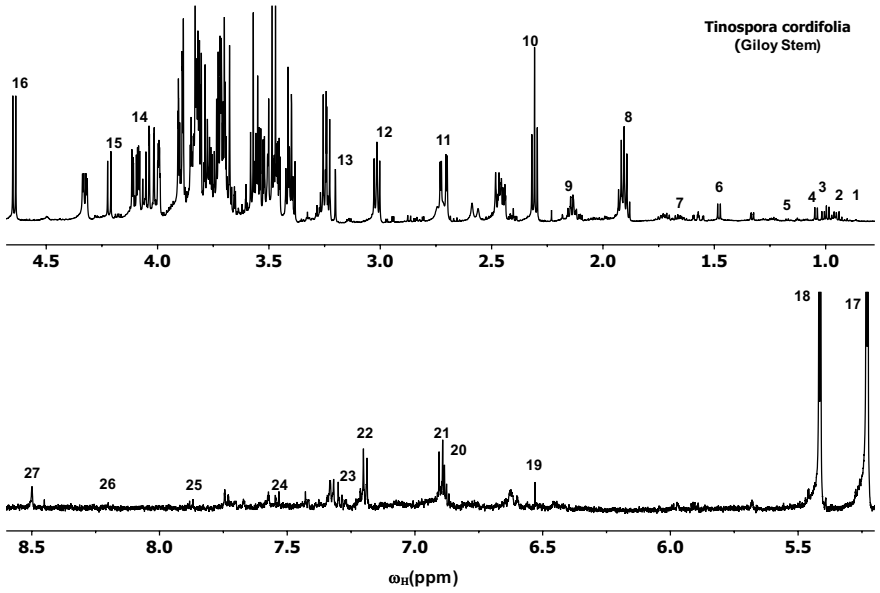

Figure 3.9: Representative ^{1}H NMR spectrum of Tinospora cordifolia stem extract recorded at 600 MHz, showing identified metabolites. Proton peaks Numbering: 1, Fatty acid methyl ester; 2, Leucine; 3, Isoleucine; 4, Valine; 5, 3-hydroxybutyric acid; 6, Alanine; 7, Arginine; 8, Acetate; 9, Glutamine; 10, GABA; 11, Malate; 12, Phenylalanine; 13, Choline; 14, Proline; 15, Sucrose; 16, Alpha-glucose; 17, Beta-glucose, 18, Sucrose; 19, Fumarate; 20, Tyrosine; 21, Quercetin; 22, Chlorogenic acid, 23, Ellagic acid; 24, Tryptophan; 25, Uridine; 26, Adenine; 27, Formate.

Table 3.2: Metabolites identified from ^{1}H NMR and 2D NMR spectrum of Tinospora cordifolia stem, with chemical shift given in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)
Sterols	0.74(s)
Fatty acid methyl esters	0.86(m)
Leucine	0.96(t,6)
Isoleucine	1.01(d,7.01),0.96(t,7.62)
Valine	1.04(d),0.97 (d, 7)
3-Hydroxybutyric acid	1.19(d,6)
Alanine	1.47(d,7.28)
Arginine	1.68(m),3.73(t,6.5)
Acetate	1.91(s)
Glutamine	2.15(m),2.47(m)
Glutamate	2.15(m), 2.0(m)
Gaba	2.30(t,14.8),1.90(m)
Malic acid	2.71(d,15.88,3.5)
Phenylalanine	3.0(m),7.32(d,6.9)
Choline	3.20(s)
Proline	4.07(dd,8.56,6.4)
Sucrose	4.21(d,8.5)
Alpha-glucose	4.64(d,7.9)
Beta-glucose	5.23(d,3.77)
Sucrose	5.41(d,3.85)
Fumaric acid	6.53(s)
Tyrosine	6.86(m),7.19(m)
Quercetin	6.89(d,8.47)
Gallic acid	7.06(s)
Chlorogenic acid	7.17(d,8.47),7.67(d)
Ellagic acid	7.27(s)
Tryptophan	7.54(d,8),7.32(s)
Uridine	7.86(d)
Adenine	8.20(s)
Formic acid	8.50(s)
NAD	8.50(s),8.20(s)
ATP	8.50(s)
AMP	8.50(s),8.20(s)

Abbreviations: GABA (gamma-aminobutyric acid), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

Similarly, OPLS-DA analysis was performed for comparisons of Tinospora cordifolia (Giloy) stem extract and juice (Patanjali product) sample illustrated in Fig 3.10 (A) with one predictive and one orthogonal component showing a clear separation between the two groups, with t[1]=76.2% and to[1]=5%, respectively. The quality of the OPLS-DA model was assessed from the R2X=0.986, R2Y=1 (variance predicted) and Q2=0.999 (variance explained) values. The corresponding S-plot, which illustrates the important metabolites that are responsible for group separation, is shown in figure Fig 3.10 (B). Cut-off values for the covariance of $\mid p[1] \mid \geq 0.05$ and for the correlation of $\mid p(corr)[1] \mid \geq 0.5$ were used. Sodium benzoate (7.52(s) ppm) and potassium sorbate (1.81(m), 5.82 ppm) peaks were observed in the proton spectra of Patanjali juice product confirmed the preservatives.

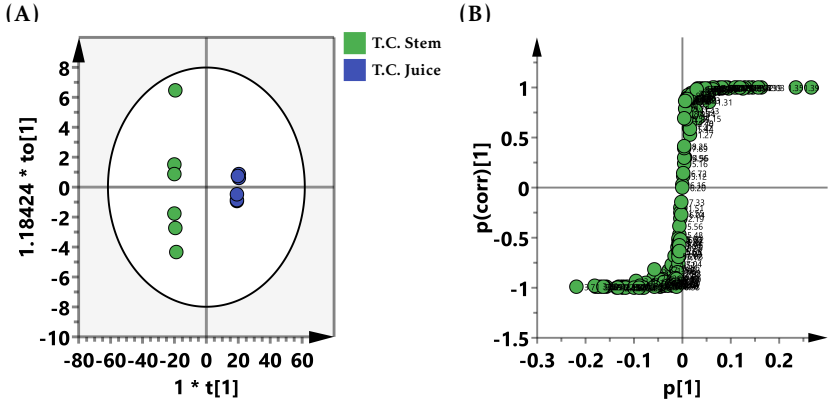

Figure 3.10: (A) OPLS–DA score plot for elucidating the difference among Tinospora cordifolia (Giloy) stem and juice (Patanjali product), (B) S-plot giving the information about the significant metabolites which were responsible for the separation between these two groups.

The Phyllanthus emblica Fruit & Tinospora cordifolia (Giloy) stem have a long history of use as traditional medicines in China, India, Brazil, and Southeast Asian nations. They are effective in treating gastrointestinal disorders, fever, jaundice, gas, and bloating. The secondary metabolites derived from natural sources have proven to be the most fruitful source of possible therapeutic leads [158]. Table 3.3 lists the pharmacological effects [158] of the molecules found in Phyllanthus emblica Fruit & Tinospora cordifolia (Giloy) stem. The molecules gallic acid, quercetin, ellagic acid, and chlorogenic acid were common in Phyllanthus emblica Fruit & Tinospora cordifolia (Giloy) stem and their pharmacological effects were shown in the Table 3.3.

Table 3.3: Pharmacological effects

Molecules	Property
Cinnamic acid	Antioxidant
Kaempferol	Antioxidant
Quercetin	Antioxidant, Anti-inflammatory, Antinociceptive
Rutin	Antioxidant, Anti-inflammatory, Radioprotective
Apigenin	Antioxidant
Ellagic acid	Antioxidant, Antidiabetic, Antitumor
Chlorogenic acid	Antidiabetic, DNA protective, Neuroprotective
Gallic acid	Antiulcer, Antioxidant, Anti-inflammatory

3.4 Conclusions

Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy was applied to identify the metabolites in Phyllanthus emblica (Amla) Fruit & Tinospora cordifolia (Giloy) stem extracted with polar & non-polar solvents. Extraction with polar (water, methanol, and ethanol) and non-polar (chloroform) solvents were used to investigate a wide variety of their biological activities, such as anti-viral, anti-oxidant, anti-diabetic, anti-cancer, anti-inflammatory, hypolipidemic, immunomodulatory, and anti-depressant activities. These activities have been shown to be effective. The useful antioxidant activity of Phyllanthus emblica (Amla) Fruit & Tinospora cordifolia (Giloy) stem led to both these plants being utilized as both food and traditional medicine. It has been proven that phenolic compounds are in responsible of the antioxidant activity of Phyllanthus emblica (Amla) fruit and Tinospora cordifolia (Giloy) stem. The secondary metabolites: quercetin, geraniin, ellagic acid, gallic acid, chlorogenic acid and kaempferol were identified by ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectroscopy. Based on the findings, it appears that the OPLS-DA analysis was able to differentiate between the samples that were purchased from the market. Finally, this study demonstrates the promise of NMR metabolomics for investigating Ayurvedic plant & fruit extracts.

3. NMR metabolomic profiling of phytomedicinal compounds in Phyllanthus emblica and Tinospora cordifolia

Chapter 4

NMR-based investigation of cycling metabolites in Helianthus annuus L. (sunflower) stems during circadian cycle

4.1 Introduction

Metabolomics involves the extensive analysis of metabolites, which are relatively small compounds. In fact, metabolic indicators are increasingly being used in conjunction with or in instead of more conventional molecular markers [159]. All parts of plants, including their roots, leaves, and stems, are able to move and do so on a regular basis. The bending, turning, twisting, and elongation of plant parts are the most common visual cues for plant's motion. As a result, plant movements refer to the ability of plants to alter their position in response to both internal and external stimuli. The plant movements are almost entirely dependent on the accumulation gain and/or loss of pushing forces that are exerted by the plant's cells and tissues. Physiological processes are the key contributors to the repetitive motions in live plants, as discussed in [160]. The Helianthus annuus L. (sunflower) is used as a case study in plant science to illustrate solar tracking [161], [162], [163], heliotropism [164], leaf hydraulic conductance [165], glutamine synthetase circadian oscillations [166], dehydrin transcript fluctuations [167], stem fascinated [168], hormonal influence [169], movements [170]. Circadian rhythms are a subset of biological rhythms with a period of 24 h [171]. Franz Halberg coined the term circadian in 1959 from the Latin words "circa"

(about) and "dies" (day) [172]. Another characteristic of circadian rhythms is that they are produced internally and can maintain themselves; as a result, they are able to function normally even when the surrounding environment remains unchanged, such as when there is no variation in the amount of light or darkness or the temperature [173], [174]. Plant growth hormones, also called phytohormones, are chemicals that are created within plants and occur in extremely low concentrations. Hormones in plants are responsible for controlling every element of a plant's growth and development, from the beginning of embryogenesis to the regulation of organ size, defense against pathogens, tolerance to stress, and finally reproductive development. In the beginning stages of study into plant hormones, five primary groups were recognized. These were abscisic acid, auxin, cytokinins, ethylene, and gibberellins [175]. Auxins play a role in the control of reproductive growth and development in plants [176]. Young Helianthus Annuus L. (sunflower) plants follow the sun from east to west during the day. At night, they turn around to face east to get ready for sunrise. Why do sunflowers follow the sun? Is this a result of the internal circadian cycle being triggered by growth hormone? To achieve this, we use the internal circadian cycle, which exerts control over the metabolic and physiological processes. We report the several metabolites including sugars, amino acids and tri-carboxylic acid (TCA) intermediates of the young sunflower stems using one- and two-dimensional nuclear magnetic resonance (NMR) spectroscopy. These metabolites primarily support cell growth by providing energy and building blocks for the synthesis of essential bio-molecules, as well as serving as signaling molecules. Partial least squares-discriminant analysis (PLS-DA) and Random forest (RF) analysis combined to identify the potential bio-markers that were strongly linked to circadian cycle. According to our findings, relative concentrations differed in carbohydrates, amino acids and organic acids during the circadian cycle. We also confirm the presence of auxin (indole-3-acetic acid, IAA) that regulates various aspects of plant growth, development and to stimulate cell elongation by using mass-spectrometric method. We conclude that the internal circadian clock and growth hormones molecules are the signaling mechanisms in the sunflower stems that maintain the orientation of the plant.

4.2 Materials and Methods

4.2.1 Study site & Plant material

Approximately 200 seeds of sunflower were sown on the same day and monitored till the appearance of flowers at our institute IISER Mohali (residential area), Punjab India. The young, elongating stem samples (picked approximately 15-20 days post-

flowering) of approximately the same height & width (2.5 $\times$ 0.5 cm^2) were chosen for NMR experiments. The close-up of the head of a common sunflower was shown in Fig 4.1. The stem samples were cut by sharp knife and immediately frozen in liquid nitrogen. This is followed by lyophilization of the samples and storage in -80 ^{0}C fridge till extraction. The stems of young sunflower used in this study was collected by Sumit Mishra (latitude 30.663891^0 N & longitude 76.727329^0 E) at IISER Mohali, Punjab in October 2017. The climate of the area during the period of study was summer, with hot and humid weather and average temperature ranges of 35–44 ^{0}C.

Figure 4.1: Sunflower plant at IISER Mohali.

4.2.2 Chemicals & Solvents

HPLC-grade methanol, HPLC-grade acetonitrile, ultrapure water, deuterium oxide (D_2O), methanol-d4 (CD_3OD-D4), 3-(Trimethylsilyl)propionic acid sodium salt (TMSP or TSP), phosphate buffers, sodium azide (NaN_3), indole-3-acetic acid, chlorogenic acid, gamma-amino butyric acid (GABA), sucrose, L-valine, L-alanine, succinic acid, L-phenylalanine, tryptophan, malic acid, and fumaric acid were procured from Sigma-Aldrich India.

4.2.3 Sample preparation

For the untargeted analysis, samples were randomized and prepared with a few modifications in the protocol [177], [178]. Fig 4.2 shows the experimental design, ten replicate stem samples of sunflower were collected at 9 time points during the entire circadian period ($\simeq$ 24-hour cycle), beginning from 9 A.M. on one day till 9 A.M. on the second day. The stem samples were finely ground in liquid nitrogen using a pre-cooled mortar and pestle and kept in separate air-tight plastic containers. The equal amounts of 30 mg dry-weight (DW) per biological replicate were extracted (mix 800 μL of methanol D4 + 200 μL of D_2O). The sodium azide (0.01 %) was used to prevent the bacterial growth in the samples [179]. A total of 600 μL solvent (phosphate buffer solution pH 7.4 in D_2O with 2 mM TMSP) was added in each eppendorf tubes. The samples were then centrifuged at 10,000$\times$ g for 15 minutes and 500 μL of the supernatant was transferred to a 5 mm NMR tube, prior to the NMR experiments.

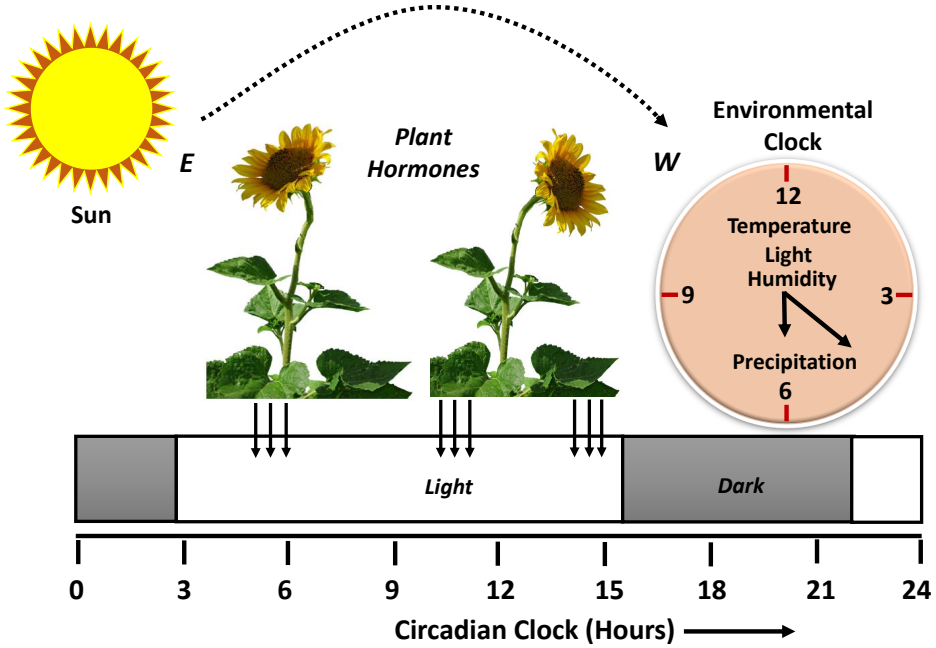

Figure 4.2: Experimental design. Solar tracking of the sunflower: upper leaves of the sunflower plant mediated by growth-hormone. During the day, the apex of the sunflower shoot follows the sun's path and, at night, turns to face east well before sunrise. Time points of harvesting of sunflower stems during circadian cycle. Young sunflower plant stems were collected at 9 time points.

4.2.4 Nuclear magnetic resonance spectroscopy (NMR)

A sets of one-dimensional ^{1}H NMR spectra were performed for each sample at 300 K on Bruker Biospin Avance-III 600 MHz NMR (Fällenden, Switzerland), operating at a proton frequency of 600.21 MHz and equipped with a 5 mm QXI quadrupolar resonance pulse field gradient rf probe. Zg30 pulse sequence spectra were observed to check the water peak and spectral width with the following parameters: 2 scans, 2 dummy scans, 64 K data points, 90^{0} pulse width of 9.95 μs, relaxation delay 4 s and spectral width 20 ppm. The Carr-Purcell-Meiboom-Gill (CPMG) spectra were recorded with a spin-echo delay time t = 300 μs, loop counter n = 400, and a total spin-spin relaxation delay (2nt) time of 240 ms. The following acquisition parameters were used to record the CPMG spectra: 16 scans, 2 dummy scans, 64 K data points, 90^{0} pulse width of 9.95 μs, relaxation delay 4 s and spectral width 12 ppm. ^{1}H ZGPR NMR spectra were acquired with the following parameters: 16 scans, 2 dummy scans, 64 K data points, 90^{0} pulse width of 9.95 μs, relaxation delay 4 s and spectral width 12 ppm. Prior to Fourier transformation (FT), FIDs were zero-filled and an exponential weighing factor corresponding to 0.3 Hz line broadening applied. The acquired NMR spectra were phase, baseline corrected and referenced using TMSP as internal reference with TOPSPIN 3.5pl7 Bruker software. Two-dimensional NMR experiments were carried out to confirm the assignments using Bruker's standard pulse programme library. Two dimensional homo-nuclear ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) spectrum with 80 ms mixing time was acquired with a spectral width of 12 ppm in both the dimensions, relaxation delay of 2 s, 2 K data points, and 8 scans. Two dimensional ^{1}H-^{13}C Hetero-nuclear Multiple-Quantum Correlation (HMQC) was acquired with 8 scans, relaxation delay 1.5 s, 1 k data points, with spectral width of 13, 220 ppm in proton and carbon dimensions respectively. Hetero-nuclear Multiple Bond Correlation (HMBC) was recorded with 8 scans, relaxation delay 1.5 s, 2 k data points, with spectral width of 12, 220 ppm in proton and carbon dimensions respectively. Hetero-nuclear Single Quantum Coherence (HSQC) was performed with 8 scans, relaxation delay 1.5 s, 1 k data points, with spectral width of 20, 220 ppm in proton and carbon dimensions respectively.

4.2.5 Metabolites identification & assignment

For metabolites identification and assignment we used the one and two-dimensional NMR spectra. Further assignments were validated by comparing the chemical shifts (ppm), splitting pattern and coupling constant (J,Hz) to the existing literature values [180], [181], [159]. In addition, the samples were spiked with the commercial compounds. Public or local spectral databases such as the Madison Metabolomics Consortium Database (MMCD), Biological Magnetic Resonance Data Bank (BMRB), the Hu-

man Metabolite Data Base (HMDB), the software packages MetaboHunter [182] and MetaboMiner (http://wishart.biology.ualberta.ca/metabominer) [183] were utilized. The identified metabolites were categorized using the metabolomics standards initiative (MSI) [118].

4.2.6 NMR data processing and statistical analysis

All NMR spectra were preprocessed before statistical analysis using TopSpin 3.5pl7 (Fällenden, Switzerland) and MestReNova software version 10.0.2 (Mestrelab Research, Spain). After phasing and correcting the baseline of the ^{1}H spectra, the chemical shift scale was set to 0.0 ppm for the signal of the internal standard TMSP. Following that, each ^{1}H NMR spectrum was divided into regions with a bin size of 0.001 ppm across the chemical shift range 0.2–10.0 ppm. This allowed for a more accurate reading of the spectrum. The water spectral regions between 4.60 and 5.10 ppm were discarded. Every integral region was first mean-centered, then log transformed, and finally normalized using the total spectral area and scaled according to the pareto scaling. For univariate and multivariate statistical analysis, the final data were imported into SIMCA version 14.0 (developed by Umetrics, Umea, Sweden) and MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/home.xhtml) [184]. The data-set was subjected to unsupervised principal component analysis (PCA) to generate an overview of the data structure, investigate the inherent variation, and identify potential outliers. In addition, OPLS-DA (orthogonal projection latent to structure discriminant analysis) was used to analyze the NMR data as the X matrix with classes and as the Y matrix to identify and differentiate the metabolites. The permutation test and seven-fold cross validation were used to evaluate the model's goodness of fit to the X matrix and its predictive ability, respectively. In order to determine whether or not there are significantly different levels of metabolites, t-test and one-way ANOVA-Tukey tests with a significance level of $p \leq 0.05$ were carried out. Metabolite relative abundance was expressed as mean $\pm$ standard deviation. The differential metabolites were determined by variable importance in projection (VIP) of PLS-DA and Random Forest (RF) models. MetaboAnalyst and the randomForest R package were used to create random forests (RF) models [185]. The RF method of machine learning is an ensemble of identically distributed decision trees that are produced by a tree classification algorithm [186]. This ensemble is known as a random forest. This method, which is based on bootstrap sampling, is widely used in metabolomics analysis [187], [188]. Pathway analysis of remarkably changed significant metabolites in stem samples were analyzed by MetaboAnalyst 5.0 [184]. Among all the perturbed metabolic pathways, those with an impact value > 0.1 and a p - value < 0.05 were considered as significantly perturbed metabolic pathways in stem samples.

4.2.7 UPLC–ESI-MS analysis

To investigate the metabolites, we performed the UPLC-ESI-MS (Ultra Performance Liquid Chromatography-electrospray ionization-mass spectrometry) analysis using with a UPLC (Waters Acquity Class I) and mass spectrometry (Waters Synapt G2-S) system in the same way as those described in [189] with some minor modifications. Reverse phase UPLC chromatographic separation was achieved using a Waters Acquity Column type (ACQUITY UPLCr BEH C18 1.7 μm particle size) maintained at 40 ^{0}C, with a mobile phase flow rate of 0.4 mL/min. The capillary voltages of 3.2 kV and 2.8 kV were used to get positive and negative ions, respectively. The mobile phase elution buffers contained water with 0.1% formic acid and acetonitrile (CH_3CN) with 0.1% formic acid. The chromatographic run parameters were: injection volume 10 μL, source temperature 110 ^{0}C, acquisition mass range 50-900 Da, and experiment reference compound name - Leucine Enkephalin. The following Gradient elution was employed starting at 100 % A & 0 % B to 0 % A to 100 % B. 30 mg of freeze-dried powder were extracted using 1.0 mL of methanol (70% aqueous) during overnight at 4 ^{0}C. The extracts were then absorbed and filtered after being centrifuged at a speed of 10,000 x g for ten minutes. Finally, the filtrate solution was analysed using UPLC. The software MassLynx^{TM} v4.1 was used to process and analyze the spectra.

4.3 Results and Discussions

4.3.1 Biological Clocks

Circadian rhythms are common to many bio-organisms and are endogenous rhythms with a duration of around 24 hours. A simple circadian system model includes input channels that entrain the clock, the central oscillator (clock), and output pathways that create overt cycle [92]. The clock regulates numerous key activities, generates circadian cycles at the molecular [190] and physiological levels [191], and contributes to plant fitness [192], [193]. Diurnal rhythms are biological rhythms synced to the day/night cycle. Representative circadian and diurnal rhythms are depicted in Fig 4.3.

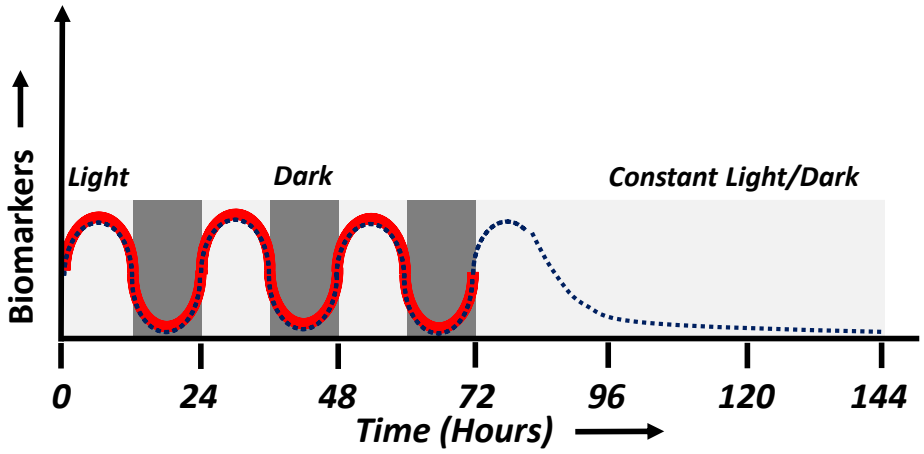

Figure 4.3: Circadian and diurnal rhythms: A conceptual comparison between the circadian and diurnal rhythms. Circadian rhythms last approximately 24 to 25 hours. A diurnal rhythm is one in which the rhythm is synchronized with the day/night cycle. A bold thick red sinusoidal wave depicting the circadian rhythm while blue tick wave form showing the diurnal rhythm which ends due to constant light/dark.

4.3.2 Metabolic profiling

The representative ^{1}H CPMG NMR spectra of stem extract is shown in Fig 4.4. A total of 35 metabolites were detected and quantified. Several amino acids and organic acids including alanine, glycine, isoleucine, leucine, proline, valine, malic acid, citric acid, acetic acid and succinic acid were attributed to the signals in the aliphatic region (δ 1.0 to δ 3.3). Fumaric acid, which is also an organic acid, showed its signal at δ 6.58 (s). Fructose, sucrose, alpha-glucose, beta-glucose, and raffinose were identified major sugars in the carbohydrate region (δ 3.4 to δ 5.5). The aromatic region δ 5.7 - 9.0 reflected some amino acids (phenylalanine & tryptophan) and nucleotides (NAD & AMP). Choline (which is essential component of cellular membranes), polyol (e.g., glycerol) and fatty acids were also confirmed in the spectra. Due to the high complexity and the extensive peak overlap in the spectra, resonance assignments were also confirmed by the two dimensional NMR experiments (TOCSY, HSQC, HMBC and HMQC). A list of identified metabolites is given below in Table 4.1.

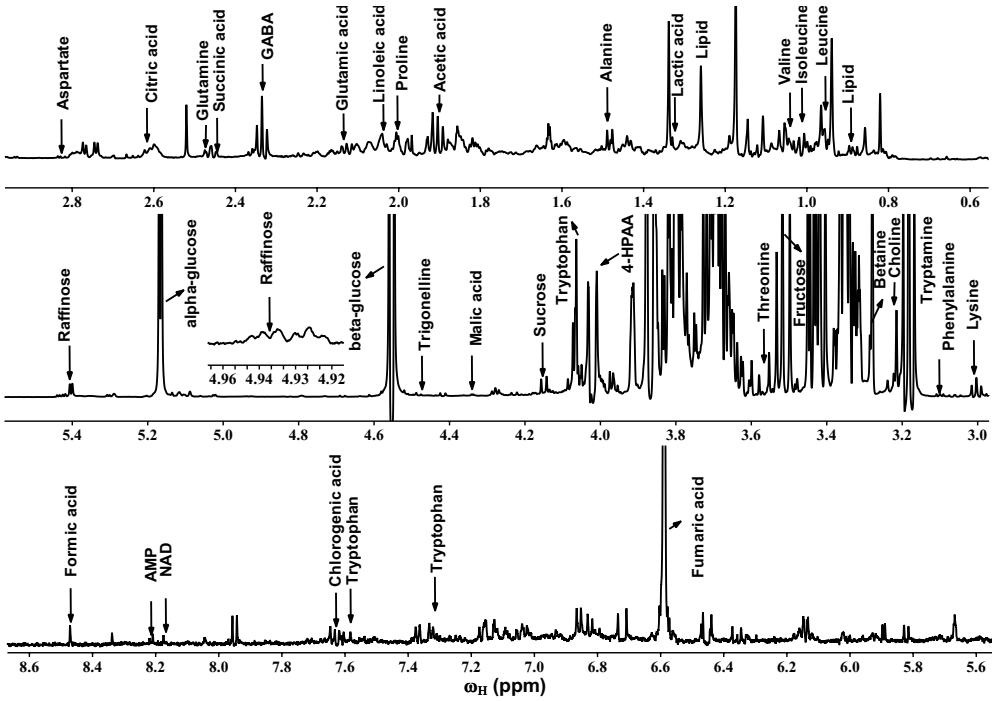

Figure 4.4: Metabolic profile of sunflower stem:The representative 600 MHz 1D-CPMG with pre-saturation proton NMR spectra obtained from the young sunflower stem extracts with metabolites labeled referenced to TMSP (0.0 ppm). The assigned metabolites primarily fall under the categories of amino acids, organic acids, carbohydrates, choline, polyols, and fatty acids. Here the pulse sequence CPMG (Carr-Purcell-Meiboom-Gill) has the form –RD–90^0–(t–180^0–t)n–acquire; where RD indicate the relaxation delay, t is the spin-echo delay and n represents the number of loops.

Table 4.1: Primary metabolites identified from 1D ^{1}H and 2D NMR spectrum of young sunflower stems, with chemical shift given in ppm, the corresponding multiplicity (m), scalar coupling (J) values (in Hz) and their MSI labeling.

Metabolites	Chemical shifts (m, J)	MSI Level
term. methyl group	0.69(s), 0.89(t,6.9)	Level 2
L-Leucine	0.95(t,6), 1.7(m), 3.72(m)	Level 2
L-Isoleucine	1.01(d,7.19), 3.66(d), 1.25(m)	Level 2
L-Valine	1.04(d,6.63), 2.26(m)	Level 1
Fatty acid	1.17(br), 1.64(m)	Level 2
Fatty acid $(-CH_2)$n	1.26(m)	Level 2
L-Lactic acid	1.32(d), 4.10(q)	Level 2
L-Alanine	1.48(d,7.3), 3.76(q,7.2)	Level 1
Acetic acid	1.91(s)	Level 2
Proline	2.00(m), 4.03(dd,8.4,8.4)	Level 2
Linoleic acid	2.05(m), 5.37(m)	Level 2
L-Glutamic acid	2.15(m), 2.0(m)	Level 2
GABA	2.33(t,7.2), 1.90(m)	Level 1
Succinic acid	2.44(s)	Level 1
L-Glutamine	2.46(m), 2.15(m)	Level 2
Citric acid	2.58(dd), 2.74(d,17.6)	Level 2
L-aspartic acid	2.82(m), 3.89(t)	Level 2
L-Lysine	3.00(t,7.59), 3.74(t,6.09)	Level 2
L-Phenylalanine	3.10(t,7), 3.0(m), 3.98(dd,7.8,5.3)	Level 1
Tryptamine	3.18(t,8.4), 3.34(t), 7.20(t)	Level 2
Choline	3.22(s)	Level 2
Glycine betaine	3.27(s)	Level 2
Fructose	3.50(d,11.7), 3.62(m)	Level 2
Glycine	3.54(s)	Level 2
Threonine	3.57(d)	Level 2
Glycerol	3.60(d,4.27)	Level 2
4-HPPA	4.01(s)	Level 2
L-Tryptophan	4.04(t,5.4), 7.32(s), 7.54(d,10)	Level 1
Sucrose	4.14(d,8.5), 5.4(d,3.8)	Level 1
Malic acid	4.32(dd), 2.68(dd,16.6,6.6), 2.36(dd)	Level 1
Trigonelline	4.46(s), 8.89(m), 9.17(s)	Level 2
Beta-glucose	4.55(d,8), 5.22(d,3.6)	Level 2
Raffinose	4.94(d,3.8), 5.40(d,3.97)	Level 2
Alpha-glucose	5.16(d,3.8), 4.63(d,7.98)	Level 2
Fumaric acid	6.58(s)	Level 1
Chlorogenic acid	7.64(d), 7.17(d,1.8), 4.25(d,2.58)	Level 1
NAD	8.17(s), 6.03(d,5.8), 6.08(d,5.7), 6.12 (d,5.8)	Level 2
AMP	8.22(s), 8.59(s)	Level 2
Formic acid	8.47(s)	Level 2

Abbreviations: GABA (gamma-aminobutyric acid), 4-hydroxyphenylpyruvic acid (4-HPPA), NAD (Nicotinamide adenine dinucleotide), AMP (Adenosine monophosphate), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, q=quartet, br=broad.

4.3.3 Mass spectroscopy

Polar extracts were analyzed for targeted metabolites using UPLC–ESI-MS analysis. The presence of amino acids, carbohydrates and organic compounds peak in stem extracts were confirmed in both the positive $[M\text{-}H]^+$ and negative $[M\text{-}H]^-$ ion modes as reported in Table 4.2 and Table 4.3. The malic acid and alpha-glucose peaks were annotated with a retention time around 15.92 min with m/z values 135 & 181 respectively. The valine peak to be identified at retention time of 14.05 min. with m/z 118. The most intense peaks that were detected in the sample extracts, was annotated as indole-3-acetic acid at retention time of 12.32 & 0.079 min. with m/z of 176 in positive ionization mode. Finally, several peaks (isoleucine, indole-3-acetic acid, chlorogenic acid, choline, valine and malic acid) were reported at a retention time of 15.75 min. in negative ionization mode. The mass spectra for pure compound (indole-3-acetic acid) and stem extract are shown in Fig 4.5, Fig 4.6.

Table 4.2: Compounds present in extracts of sunflower stem, identified using UPLC–ESI-MS in positive ionization mode.

Metabolites	**M.W.**	**$[M\text{-}H]^+$**	**R_t (minutes)**
Malic acid	134	135	15.92
Alpha-glucose	180	181	15.92
Valine	117	118	14.05
Indole-3-acetic acid	175	176	12.32
Raffinose	504	505	12.32
Threonine	119	120	0.66
Trigonelline	137	138	0.66
Indole-3-acetic acid	175	176	0.079
Acetic acid	60	61	0.079
Proline	115	116	0.079
Glutamic acid	145	146	0.079
Trigonelline	137	138	0.079

Abbreviations: M.W.- Molecular weight; R_t - Retention time.

Table 4.3: Compounds present in extracts of sunflower stem, identified using UPLC–ESI-MS in negative ionization mode.

Metabolites	**M.W.**	**[M-H]$^-$**	**R$_t$ (minutes)**
Isoleucine	131	130	15.75
Indole-3-acetic acid	175	174	15.75
Chlorogenic acid	354	353	15.75
Choline	104	103	15.75
Valine	117	116	15.75
Malic acid	134	133	15.75

Abbreviations: M.W.- Molecular weight; R$_t$ - Retention time.

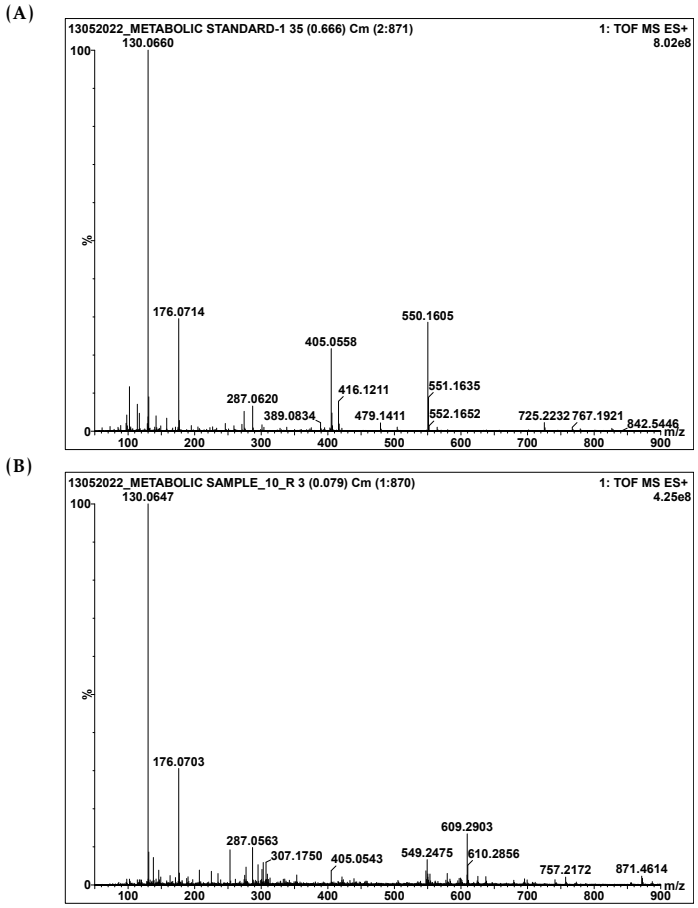

Figure 4.5: Mass spectra showing the detected masses in Positive ESI mode $[M\text{-}H]^{+}$: (A) Pure compound, and (B) Sunflower stem extract.

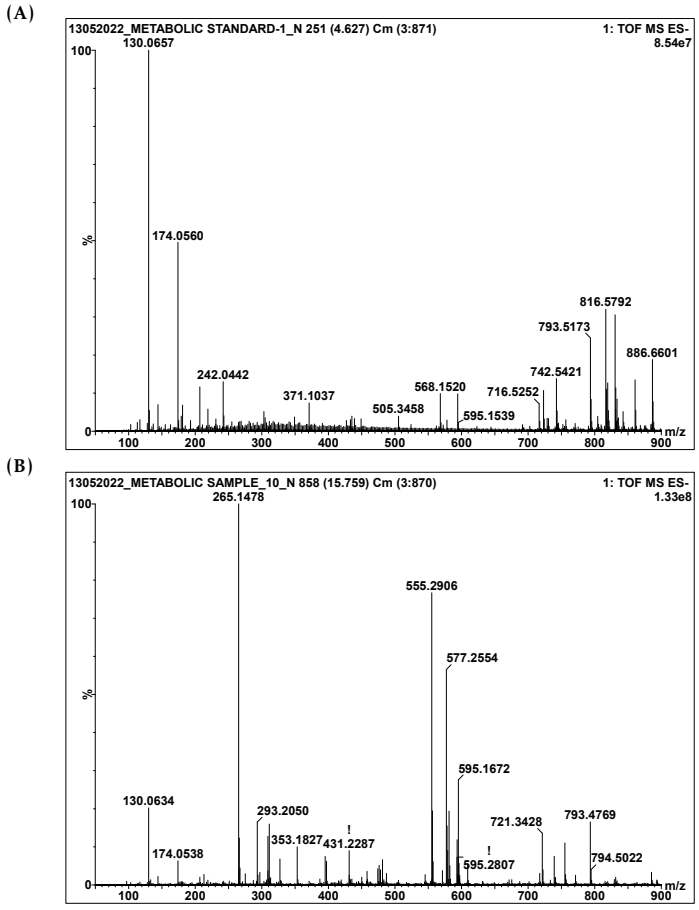

Figure 4.6: Mass spectra showing the detected masses in Negative ESI mode $[M\text{-}H]^-$: (A) Pure compound, and (B) Sunflower stem extract.

4.3.4 Statistical analysis

In this study, untargeted NMR-based metabolomics was used to detect changes in the stem metabolite profile of sunflower due to internal circadian cycle. Multivariate analysis (MVA) was conducted to explain variance within a dataset and to identify the biologically relevant spectral features or to mark the outliers [67]. The unsupervised principal component analysis (PCA) and supervised orthogonal partial least squares discriminant analysis (OPLS-DA) are popular methods to produce the score matrix as well as a loading matrix. The score matrix depicts the relationship between observations while the loading matrix explain the individual peak contribution in the NMR spectra [194], [195], [196], [197]. In order to determine which metabolites undergo cyclical changes in concentration, samples of sunflower stems were taken from the

IISER Mohali campus at nine different time points, with each time point separated by three hours. At first, the samples from all time points were compared using the unsupervised method of principal component analysis (PCA). The PCA model (with two principal components, t[1] = 29.5% & t[2] = 13.5 %, R2X(cum) = 0.827, Q2(cum) = 0.589) showed good separation between all time points. Biological outliers were found in two samples, when compared to the larger group of samples in the centre Fig 6.6. A supervised OPLS-DA model was performed to get a better visualization of clustering between the groups. This model for classifying the sampling showed complete separation along the first principal component t[1] Fig 6.6 with R2X = 0.827 and Q2 = 0.589 (represent the cumulative values for this model). In order to get a more clearer picture of how the time points are grouped together, a polar dendrogram also obtained by hierarchical clustering shown in Fig 4.8. Here the leaf nodes are spread around the circumference of a circle. There are two main nodes resulting in grouping of the time points depicting by two colors red and blue. In addition, the greatest difference was found between the hours of 6 am and 6 pm, which correspond to sunrise and sunset, respectively. It might be a sign that the environmental cue of light has induced the internal clock to synchronize. We performed the OPLS-DA analysis to further investigate the metabolic perturbations between these two time points 06:00 AM and 06:00 PM which showed the greatest difference. Fig 6.7 shows the OPLS-DA score plot of the stem samples for comparison between 06:00 AM and 06:00 PM, with one predictive t[1] = 28.3 % and one orthogonal component to[1] = 20.8 % showing a clear separation between these two time points. The R2X, R2Y and Q2 values for the model were 0.766, 0.995 and 0.99 respectively. Additionally, the CV-ANOVA permutation test revealed that the model was statistically significant (p-value $\leq$ 0.05), demonstrating its validity and robustness. S-plots was generated from OPLS-DA model to identify the metabolites responsible for the differentiation in the score plot between two time points. Stem samples at 6:00 A.M. and 6:00 P.M. were characterized by higher levels of tryptamine, glutamine, citric acid, oleic acid, fumaric acid, raffinose and lower levels of fructose, AMP in the samples. The statistical significance of the metabolites were also confirmed by a univariate t-test analysis, where a p-value $\leq$ 0.05 was considered statistically significant.

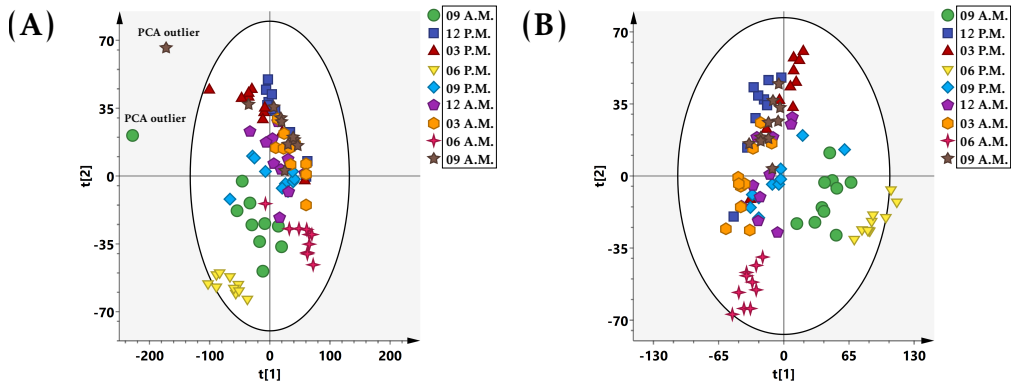

Figure 4.7: PCA & OPLS-DA of samples collected at different times during the circadian cycle from sunflower stem. A: Metabolomic profiles analyzed using ^{1}H-NMR Spectroscopy: Principal component analysis (PCA) score plot of sunflower stem extracts between the first two components t[1] = 29.5 % and t[2] = 13.5 % at different sampling time points. Principal Component Analyses revealed a good clusterization of samples were collected at nine time points and measured in decuplet. The plot showed clear separation between the samples from the initial time point 09:00 A.M. to final time point 09:00 A.M.. Two samplings were marked as the outliers. A = 8+1+0, N = 90, R2X(cum) = 0.827, R2Y(cum) = 0.714, Q2(cum) = 0.589. B: Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) score plot was derived from the ^{1}H NMR spectra of sunflower stem extracts between the first two components at different sampling time points. The ellipse represents the 95 % confidence region for Hotelling's T2. OPLS-DA analysis define a perfect metabolic signature for all the groups. Samples were collected at nine time points and all samples were measured in decuplet. The plot showed clear separation between the samples from the initial time point 09:00 A.M. to final time point 09:00 A.M.

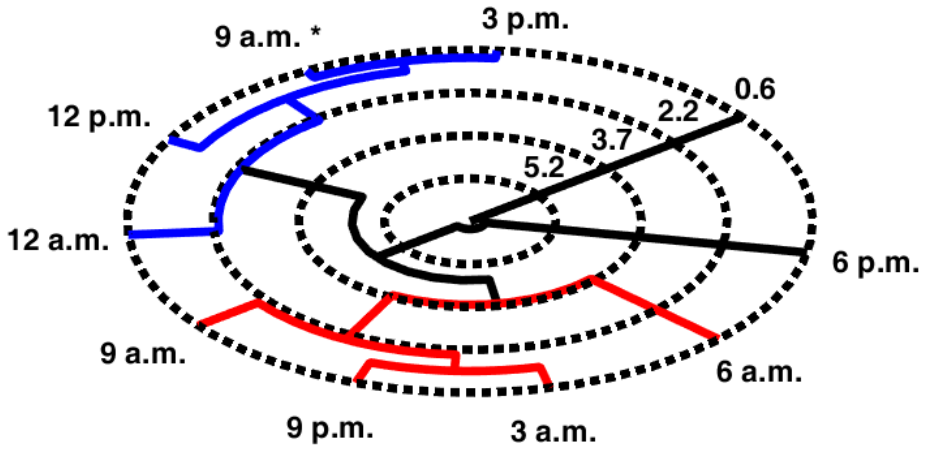

Figure 4.8: Polar dendrogram.

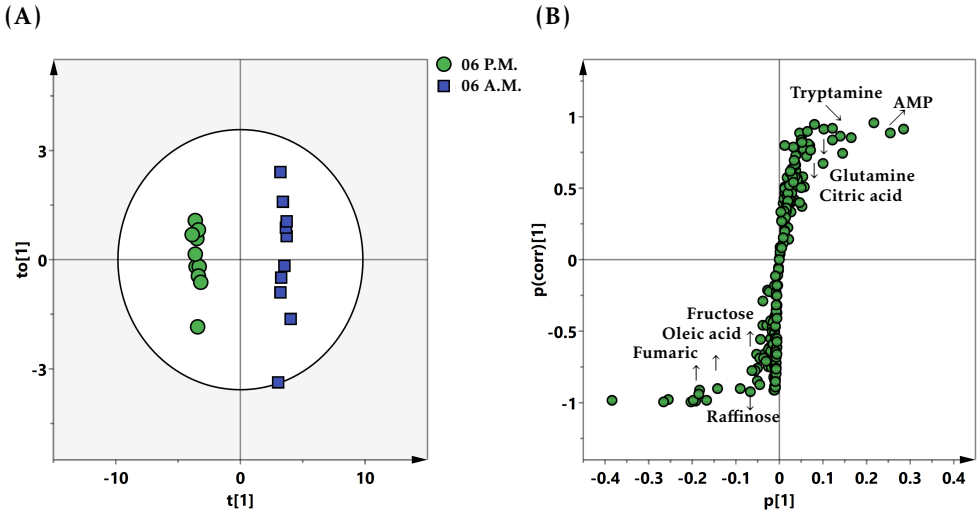

Figure 4.9: OPLS-DA & S-plot. (A): OPLS-DA score plot applied to the ^{1}H NMR spectra of sunflower stem extracts at two sampling time points 06:00 P.M. and 06:00 A.M. with explained variance t[1] = 28.3 % and to[1] = 20.8 %. The goodness-of-fit parameters R2X=0.766, R2Y=0.995, Q2=0.99 and R2=0.995 for the OPLS-DA models. A (1+1+0), N = 20. (B) S-plot indicated that the two sampling time points 06:00 P.M. and 06:00 A.M. depicted in OPLS-DA score plot could be separated based on these significant metabolites.

4.3.5 Circadian rhythm of the significant metabolites

It has been found that sunflower solar tracking movements are driven by antiphasic patterns of stem elongation on the east and west sides, and interactions between environmental response pathways and the internal circadian oscillator coordinate physiological processes with predictable changes in the environment to influence growth and reproduction [164]. The circadian cycle strongly influences many plant metabolic and physiological processes [192], [198], [199], [200]. It should be noted that through this analysis, we were able to pinpoint the metabolites that cycle either as a result of circadian rhythm or as a result of the environment's day and night cycle. Additional studies carried out under constant conditions, such as constant light or constant dark, will confirm that the identified metabolite cycling is in fact, caused by the circadian rhythmicity and not due to cyclic day/night. NMR-based metabolomics approach is used to see the effect of circadian rhythms on primary metabolism of young sunflower grown in natural environmental conditions. We identified and investigated the significant metabolites that fluctuate in response to the circadian rhythm and are directly or indirectly involved in the heliotropic movement of the sunflower plant. The relative concentration of the cycling metabolites with peaks and trough were plotted over the course of the nine time points that were recorded every three hours interval. Peaks and troughs varies for individual metabolites at different time points of the circadian cycle Fig 4.10.

Amino acids work as intermediates of final metabolites in some metabolic pathways and regulate several metabolic, physiological, and biochemical pathways, affecting many plant physiological activities [201], [202]. In the Fig 4.10 (A), (J), (K), and (L), it can be seen that the amino acids alanine, glutamine, leucine, in addition to proline, all exhibit a rhythmic pattern during the light dark cycle.

Glucose and fructose participate in various pathways in plant metabolism and act as a signaling molecules controlling plant growth and development [203]. The levels of sugar derivatives glucose, and fructose are reportedly affected by the light/dark cycle [204], [205]. Fig 4.10 (B), (G) show rhythmic changes in sugar levels. The concentration of glucose increased during the day, reached a maximum, and then decreased at the end of the light period. The concentration level was roughly the same during the dark period. The amount of fructose that was contained in the stems experienced an initial steep decrease during the light period, which was then followed by an increase to a level that was the highest possible at the end of the light period. After that, it's concentration dropped during the time when there was darkness. The plant consumes these sugar and reserves for night. Low sugar level at the start of the light period is the expression of clock genes [206].

4-Hydroxyphenylpyruvic acid, also known as 4-HPPA, is a compound that belongs to the keto acid family and is an essential intermediate in the tyrosine and phenylala-

nine metabolic pathway. [207], [208]. There was a rising trend in 4-HPPA concentration during the day and a falling trend during the night Fig 4.10 (C).

Choline is an important neurotransmitter and a precursor for the biosynthesis of phospholipids & acetylcholine, essential components of all membranes [209]. Choline concentration showed a decreasing pattern during the light period while it is increased during the dark period as shown in Fig 4.10 (E). Day-length affects the stems lipid membrane composition [210].

Choline can be converted to the organic osmolyte betaine through oxidation [211], [212]. In a manner analogous to that of choline, the concentration of betaine was found to fall during the period of light and to increase during the period of darkness Fig 4.10 (D).

Fumaric acid is a significant intercellular transport molecule and found highly concentrated in stems [213]. It is a major source of fixed carbon even in plant species that carry out C_3 photosynthesis and found increase in concentration with plant age and light intensity [214]. According to earlier research, fumaric acid can be metabolized to produce energy and carbon skeletons for the synthesis of other compounds, just like starch and soluble sugars can do. Fig 4.10 (H) demonstrates that the concentration of fumaric acid rises during the day, reaches a maximum value, then falls and begins to end during the night. During the night the concentration level become nearly equal may be because the rates at which fumaric acid is formed and degraded are equal. A second possibility is that fumaric acid is exuded from the leaves during development in constant darkness.

Citric acid can be used to synthesize amino acids or GABA by producing glutamate [215], [216]. Citric acid levels were also measured during the light-dark cycle. During the periods of light, the concentration of citric acid was lower, while it was higher during the periods of darkness Fig 4.10 (F).

γ-Aminobutyric acid (GABA) most likely has a dual role as both a signaling molecule and a metabolite, could be similar to those performed by other metabolites, such as glutamate or carbohydrates [217]. It is widely documented that GABA occupies a central place at the interface between the carbon and nitrogen metabolism of plants [218]. It was shown in the Fig 4.10 (I) that the concentration of GABA in the stems of young sunflower increased throughout the light period and reduced during the dark period.

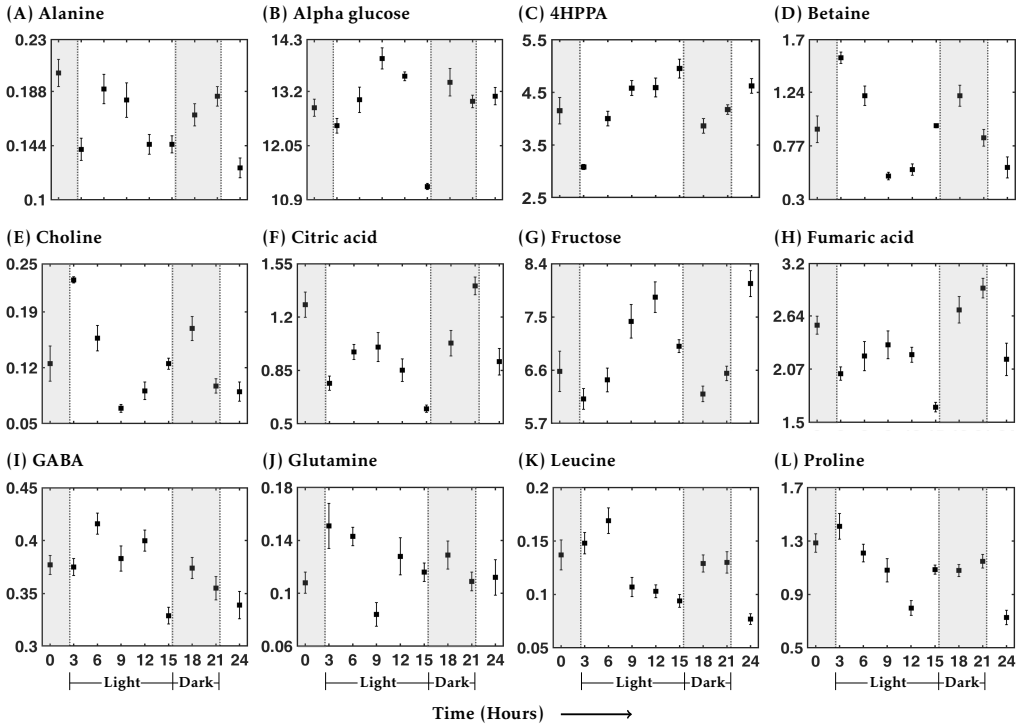

Figure 4.10: Plots of relative concentrations of oscillating metabolites sampled at 3 hour intervals across 24 hours. Each data point is an average of ten replicates. Selected metabolites with relative concentration from ^{1}H NMR-based metabolomics profiles during the circadian cycle. The X-axis shows time (hours) and the Y-axis shows relative concentration levels to TMSP. All data are expressed as the mean $\pm$ SEM (standard error of mean). In all analyses, statistical significance is shown as: $p \leq 0.05$.

4.3.6 Circadian rhythmicity of the metabolites with solar elevation

The circadian rhythmicity of tryptamine and tryptophan is maintained throughout the 24-hour day and night cycle, as shown in Fig 4.11. The levels of tryptamine and tryptophan were found to be higher during the day, correlating with solar activity Fig 4.11 (A), (B), (C). The data for solar elevation angle is taken from the following website Keisan Online Calculator: https://keisan.casio.com/ (High accuracy calculation for life or science).

Tryptamine maybe a precursor of the indole acetic acid that stimulates plant growth in some plants [219]. In our study, we found that tryptamine levels tended to rise during the day and fall at night Fig 4.11 (B).

Tryptophan is an aromatic amino acid [220] that makes up the backbone of hun-

dreds of thousands of secondary metabolites in plants, such as the indoleamines [221], auxin (indole-3-acetic acid; IAA) [222], [223], alkaloids [224], benzoxazinoids [225], and glucosinolates [226]. Fig 4.11 (C) shows that the amount of tryptophan rises during the day and reaches its highest during the light period. After that, it starts to go down and reaches its lowest point at the end of the dark period.

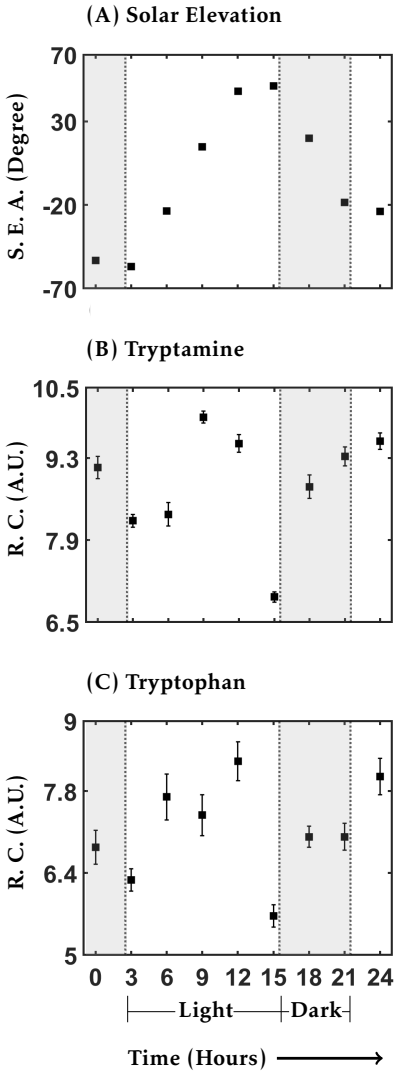

Figure 4.11: Comparison of significant metabolites with the solar position throughout the day and night. During the study, significant metabolites were compared to solar elevation angle (degree).

4.3.7 Metabolite significance analysis

Random Forest is a supervised learning algorithm that works well with high-dimensional data. RF analysis is performed using the RandomForest package [185]. PLS-DA VIP's and RF important features were compared to identify which bins are significant. The discriminating metabolites were identified from the chemical shifts in the PLS-DA VIP plot and from the VIP (variable importance) values in the Random Forests for each cluster of stem samples. VIP levels greater than 1 were deemed significant. The number of trees in the random forest in this study was set at 500. RF method developed a list of significant variables (variables of importance) by assessing the out-of-bag (OOB) error increase associated with variables during permutations (mean decrease accuracy; MDA). Fig 4.12 shows the important features ranked by random forest. Summary of significant bins for differentiation between two time points 06:00 A.M. Vs 06:00 P.M. is listed for RF mean decrease accuracy (MDA) and PLS-DA variable importance in projection (VIP). Bonferroni-corrected p-values and false discovery rates (FDR) by Kruskal-Wallis testing are listed below in Table 4.4.

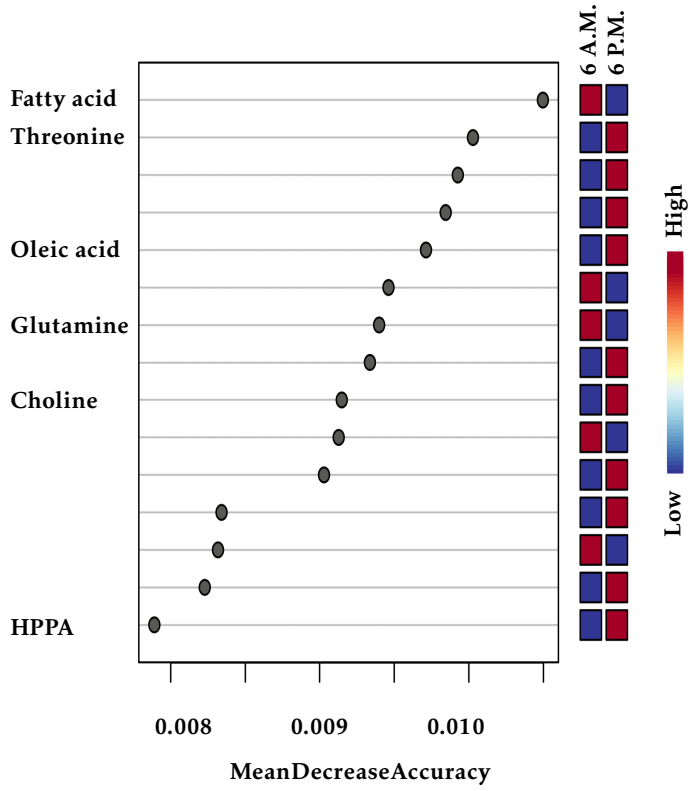

Figure 4.12: Random Forest (RF) method: Significant features found using the Random Forest method. The features are ranked by the mean decrease in classification accuracy when they are permuted

Table 4.4: Random Forest (RF) classification.

Compound	VIP score	RF MDA	p-value	FDR
Fatty acid	1.48	1.04E-2	2.84E-08	4.65E-07
Threonine	3.56	1.00E-2	1.96E-13	5.96E-12
Oleic acid	2.01	0.97E-2	7.66E-08	9.60E-07
Glutamine	1.18	0.93E-2	1.02E-09	2.42E-08
Choline	3.82	0.91E-2	5.39E-08	8.19E-07
4-HPPA	2.32	0.78E-2	1.30E-13	4.63E-12

Abbreviations - MDA : Mean decrease in accuracy; 4-HPPA : 4-Hydroxyphenylpyruvic acid.

4.3.8 Pathway analysis

The metabolites can be linked to their distinctive metabolic pathways to pinpoint the underlying circadian-regulated metabolic pathways. Pathway analysis shows that there were significant changes to four metabolic pathways namely Alanine, aspartate and glutamate metabolism, Glyoxylate and dicarboxylate metabolism, Glycine, serine and threonine metabolism and Tryptophan metabolism as shown in Fig 4.13. IAA, also known as indole-3-acetic acid, is the most common type of active auxin and can be synthesised either from tryptophan or from tryptophan precursors [227], [228]. Four tryptophan (Trp)-dependent and one Trp-independent pathways have thus far been proposed for the biosynthesis of auxin in plants [229]. In tryptophan-dependent IAA biosynthesis, four pathways exist: (i) the tryptamine (TAM) pathway, (ii) the indole-3-acetamide (IAM) pathway, (iii) the indole-3-pyruvic acid (IPA), (iv) the indole-3-acetaldoxime (IAOx) pathway as shown in Fig 4.14.

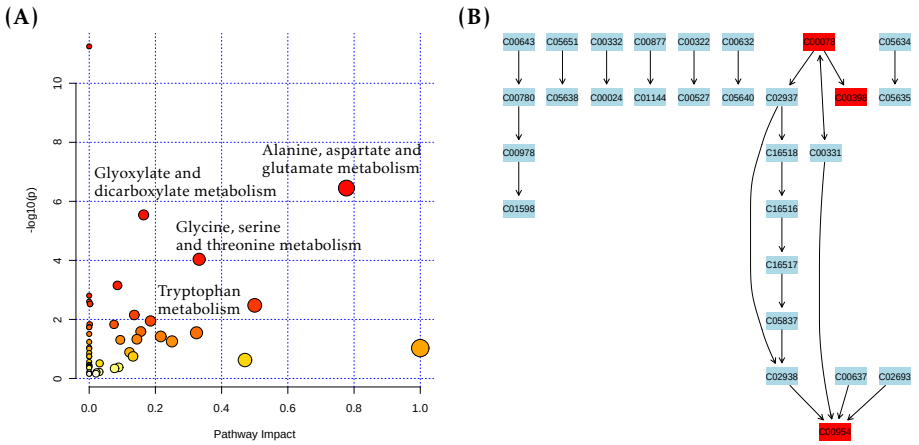

Figure 4.13: Pathway analysis of significantly changed metabolites in stem samples: (A) The metabolome view on the left shows all matched pathways according to the p values from the pathway enrichment analysis and pathway impact values from the pathway topology analysis. Solid color circle (varying from yellow to red) shows the metabolites are in the data with different levels of significance. (B) Pathway view of tryptophan metabolism: ID's: C00078 L-Tryptophan, C00398 Tryptamine, C00954 Indole-3-acetate. About compound colors within the pathway - light blue means those metabolites are not in our data and are used as background for enrichment analysis

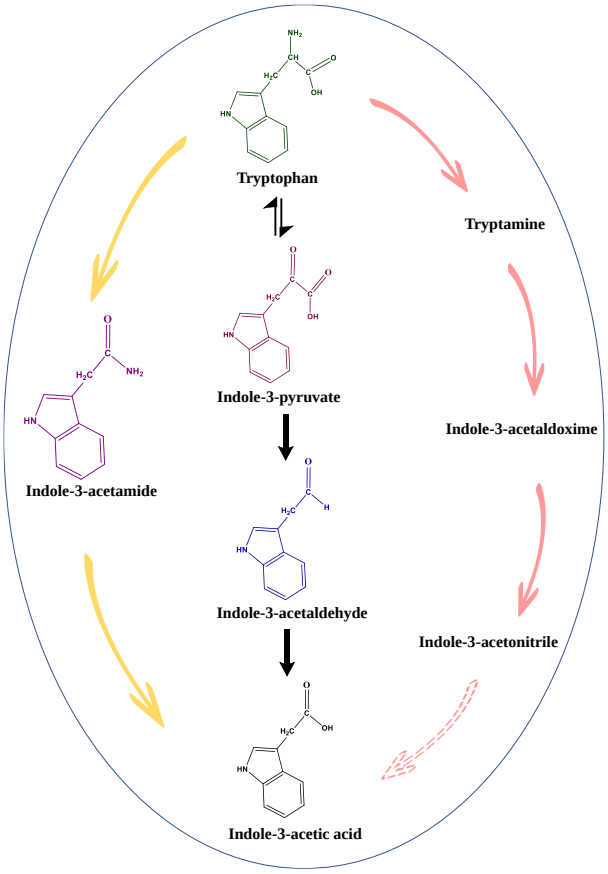

Figure 4.14: Tryptophan dependent pathway for IAA (indole-3-acetic acid) production: Biosynthetic routes leading to Indole-3-acetic acid (IAA) production that are dependent on tryptophan.

4.4 Conclusions

The application of NMR spectroscopy to metabolomics has significant potential to make important contributions to the development of this area of plant biology. The circadian rhythm plays an important role in the progression of plant growth and development. A plant's internal circadian clock runs on a 24-hour cycle, allowing different genes to be activated at different times of day. In response to both abiotic and biotic factors, as well as the right temperature, the circadian rhythm forecasts a 24-hour cycle consisting of 16 hours of light and 8 hours of darkness. Natural day/night cycles synchronise this internal clock with actual daytime. Changing the duration of daytime or nighttime can reset the clock. Continuous light completely throws off the clock in sun-

flowers. In the present work we report the biomarker signatures for young Helianthus annuus L. (sunflower) stems, which offers biochemical explanations for the metabolic change. The metabolic profiling of young Helianthus annuus L. (sunflower) stems revealed 35 metabolites including lipids, amino acids, carbohydrates and organic acids. Sugar metabolism is a key circadian output that contributes to plant physiology optimization. These metabolites and the metabolic changes that they undergo, in turn, provide feedback that regulates the diurnal and circadian oscillations of gene expression. Indeed, several metabolites have the potential to recognize the phenomenon of circadian rhythm in the physiological processes of plants. This is accomplished by recognizing the rhythmicity in the metabolites that contribute to those physiological processes. The concentration levels of the biomarkers showed the alteration during the circadian cycle. Indole-3-acetic acid (IAA), which is involved in plant growth & developmental processes such as cell division and elongation, tissue differentiation, apical dominance, and responses to light, gravity, and pathogens was confirmed by the UPLC-ESI-MS analysis.

Chapter 5

NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to air pollution stress

5.1 Introduction

Air pollution from hazardous substances and contaminants such as particulate matter, ground-level ozone, poisonous gases, and volatile organic compounds harms human health and global ecosystems [230]. Changes in plant physiology, morphology, and metabolism can be used to assess air pollution over time [231], [232]. Plants may help reduce air pollution and improve air quality [233]. The metabolome of an organism determines its phenotype and is regulated temporally, spatially, environmentally and genetically [234]. Plants' metabolism is disrupted by abiotic stressors, which slows plant growth & development [52]. However, plants can "reprogram" their metabolic networks in response to abiotic challenges, allowing them to both maintain metabolic equilibrium and create secondary metabolites to combat the stresses [235]. Therefore, metabolomics studies can shed light on the complex biochemical and genetic interactions that define organismal development [236], [237]. The environmental metabolomics is a relatively new area of research that assesses the changes that occur in the metabolic profile of an organism as a result of exposure to various types of environmental stress [238]. Plants produce a wide variety of specialized secondary metabolites, such as flavonoids, phenolics, and phytosterols, as an adaptive response to the many different kinds of biotic and abiotic environmental stresses that they are subjected to [84]. Plant metabolism variations in response to stress can be utilized as sensitive markers of environmental changes [85], [239]. NMR spectroscopy is frequently

employed for metabolomic investigations in many organisms due to its non-invasive, quantitative, robust, and reproducible nature [240]. NMR metabolomics investigations help scientists understand how environmental stressors including toxicity, contamination, climate change, or growth conditions affect biochemical response pathways in the metabolome [241]. It has been suggested that if plants can synchronize their internal circadian clocks with the light–dark cycle of their environment, they will have an advantage in terms of their ability to survive, grow, and compete [192]. The circadian clocks of plants have changed over time to take on a distinct periodicity, which can be utilized to characterize a variety of physiological, behavioral, and metabolic processes [242]. As a result, circadian rhythms are a reliable indicator of the way in which plants react to environmental challenges, which in turn can have a feedback effect on circadian rhythms [243]. High-Resolution-Magic-Angle Spinning (HR-MAS) NMR has been utilized to measure the in vivo metabolic profile of Arabidopsis thaliana leaves during the circadian cycle [244]. It has been found in the past that certain plant species have the capacity to withstand high levels of the stress caused by air pollution, which does not adversely affect their rate of growth, the senescence of their leaves, or their photosynthetic activity [245], [246], [247]. Bougainvillea spectabilis is a plant that is commonly grown along highway traffic dividers in India. Several studies conducted in the context of India have reported that this plant is one of several ornamental plant species that have shown evidence of tolerance to air pollutants. Depending on the severity of the stress and how long it lasts, air pollution can have a significant impact on the entire metabolic process of Bougainvillea spectabilis plant as it progresses through the growth phase. This includes the physiological, biochemical, and molecular levels. Air pollution can cause plants to exhibit mottled foliage, "burning" at the leaf tips or edges, stunted growth, early leaf drop, delayed maturation, and decreased yield or quality [247]. Pollutants in the air cause damage to the leaf cuticles and influence the conductance of the stomata [247]. Air pollutants like sulphur dioxide, ozone, and oxides of nitrogen can influence the physiological processes of plants, consequently influencing patterns of growth [247]. However, a detailed mechanistic picture account has yet to emerge [248], [249], [250], [251]. Plants react to the diverse abiotic stress situations by activating the stress signals, which then leads to a variety of changes in their physiology, genetics, and metabolism. We investigated the variations that occurred in the metabolome of Bougainvillea spectabilis in response to extended air pollution stress by using NMR spectroscopy in conjunction with multivariate statistical approaches. Previous research have concentrated on identifying metabolites whose periodicity matches the circadian clock of the organism because these rhythms are robust to random changes [244]. We first examined the cycling patterns of selected metabolites found in Bougainvillea spectabilis leaves and compared them to those found in the model plant A. thaliana. The presence of important metabo-

lites discovered using 1D and 2D NMR investigations was confirmed using ultra performance liquid chromatography-electrospray ionization-mass spectrometry (UPLC-ESI-MS). After that, we compared the metabolic profiles of polluted leaves to those of control leaves (from plants grown on highway dividers and exposed for prolonged durations to high levels of vehicular emissions). Using two distinct free radical scavenging assays, we were able to determine how much of a difference there was between the control samples and the polluted samples in terms of their antioxidant capabilities. We used a combined analysis of the identified metabolites that differed significantly between the control and polluted samples to gain insight into the metabolic pathways that are "re-programmed" in Bougainvillea spectabilis to promote plant regrowth and to maintain regular metabolic processes that were disrupted by air pollution stress.

5.2 Experimental Methods

5.2.1 Study sites

Two sites were chosen for the study: IISER Mohali campus (30.663611^0N, 76.7275^0E), is a residential educational institute with low levels of vehicular pollution and was chosen as the control site for the study. Zirakpur town (30.6425^0N, 76.8173^0E) is a crowded industrial town with a national highway passing through its central area with heavy vehicular traffic and combustion from industry, transportation emissions, chlorofluorocarbons (CFCs) in aerosols and stubble burning are major sources of air pollution in this region. It was hence chosen as the heavy pollution site for the study. Both the control and the polluted sites are located in the state of Punjab in North India, and the distance between the two study sites is 12 km. The climate of the area during the period of study was summer, with hot and humid weather and average temperature ranges of 35–44 ^{0}C.

5.2.2 Plant cultivation

Leaves from the Bougainvillea spectabilis plant were collected from the two study sites and classified (according to the pollution level of that area) into two categories: control (C, from plants not exposed to air pollution), and polluted samples (P, from plants exposed to heavy air pollution). The control samples of Bougainvillea spectabilis plants (bought from a nursery at eight weeks old) were grown in small pots at IISER Mohali under natural conditions for nine weeks, beginning in May 2017: 22–44 ^{0}C, 12:12 h light–dark cycle and 61% average humidity. A secluded area behind the NMR lab on campus with no vehicular traffic was chosen to grow the plants and to ensure that

there is no exposure to vehicular emissions. Plants were watered regularly and grown without additional fertilization Fig 5.1 (Plants grown in pots at control site). After nine weeks of growth, several fresh equal sized leaves (the total area of each leaf was similar) were collected from the plants growing in the control site. For samples from the polluted site, we obtained leaves from Bougainvillea spectabilis plants grown and maintained along the highway road dividers by the Highways Authorities of India. Sample collection was performed in November 2017, and plants growing in the polluted site show the following features: bigger leaves (2.0–2.5 cm length $\times$ 1.4–1.8 cm breadth), darker green in color and a thicker stem as compared to the plants at the control site. These observations indicate that the plants at the polluted site are much older than the control site samples and have faced prolonged exposure to air pollution and might even have faced different types of abiotic stresses, including drought stress, during their growth period and beyond.

Figure 5.1: Photo of a green and growing Bougainvillea spectabilis plants in IISER Mohali.

5.2.3 Solvents and chemicals

All reagents, NMR reference standards, extraction chemicals and deuterated solvents such as deuterium oxide (D_2O, 99.90%), deuterated methanol-D_4 (CD_3OD-D_4, 99.80%), trimethylsilane propionic acid sodium salt (TMSP) and di-sodium hydrogen phosphate (Na_2HPO_4), which were used for the experiments, were procured from Sigma-Aldrich (India).

5.2.4 Sample preparation

The sample preparation included: freeze-drying, weighing and extraction [121]. The collected leaves were first cleaned by distilled water to remove surface contaminants and then frozen using liquid nitrogen. The frozen leaves were ground using a pre-cooled pestle and mortar under liquid nitrogen and then stored at -80 ^{0}C prior to lyophilization. These leaves were freeze-dried in a lyophilizer for 12 h. NMR samples were prepared with 25 mg of dried leaf powder per sample. The extraction was performed at room temperature to avoid degradation of heat-sensitive plant metabolites. The extraction solvent used was prepared by mixing methanol-D_4 (560 μL), D_2O, (240 μL) and Na_2HPO_4 (0.95 mg) with TMSP powder (1.5 mg/mL) used as an internal NMR reference. A total of 800 μL of the extraction solvent was added to the leaf samples and left to stand for 30 m for metabolite extraction. The samples were then centrifuged at 5000 rcf for 5–10 min and 500 μL of the supernatant was transferred to a 5 mm NMR tube, prior to the NMR experiments.

5.2.5 Sample collection

For the identification of cycling metabolites and in order to profile the changes in their levels across the circadian period, plant leaves were sampled at 2 h intervals over 26 h. A total of seventy eight control leaf samples were prepared for the thirteen equally spaced time points analysis (each time point has six replicates). After monitoring the flux and volume of vehicular traffic at the polluted site, we chose two time points for sample collection: 8 AM and 8 PM, when traffic and hence vehicular emission levels were at their peak. Twelve samples (six replicates at each of these two time points) were collected from the plants exposed to heavy air pollution.

5.2.6 1D and 2D NMR spectroscopy

All NMR experiments were performed at 298 K on a Bruker Biospin Avance-III 600 spectrometer (Fällenden, Switzerland), operating at a proton frequency of 600.13 MHz and equipped with a 5 mm QXI quadrupolar resonance pulse field gradient rf probe. NMR spectra were recorded without sample spinning and with the addition of 16 dummy scans prior to acquisition. 1D ^{1}H NMR spectra were acquired with a Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequence with incorporated water suppression and optimized with a spin-echo delay time t = 300 μs and loop counter n = 400, and a total spin-spin relaxation delay (2nt) time of 240 ms. This sequence is often used for metabolomics experiments as it achieves attenuation of fast-relaxing broad signals arising from larger molecules. Proton spectra were recorded with a 90^0 pulse

width of 9.15 μs, a relaxation delay of 2 s, 16 scans, free induction decay (FID) size of 64 K data points, and a spectral width of 7211.54 Hz. Data were zero-filled by a factor of 2 and the FIDs were multiplied by an exponential weighting function equivalent to a line broadening of 1 Hz prior to Fourier transformation. The spectra were then phase and baseline corrected and referenced to the TMSP signal at 0 ppm. Multiplicities in the observed NMR resonances were labeled according to the convention: s = singlet; d = doublet; dd = doublet of doublets; t = triplet; and m = multiplet. To corroborate metabolite identification using 1D spectra, 2D NMR experiments recorded including homonuclear ^{1}H-^{1}H correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and J-resolved (JRES) experiments, and heteronuclear ^{1}H–^{13}C coherence spectroscopy (HSQC) experiments. The COSY and TOCSY spectra were recorded with a spectral width of 12 ppm in both the dimensions, relaxation delay of 2 s, 2 K data points, and 16 scans. The 2D JRES spectra were recorded with a spectral width of 2400.768 Hz in the F1 and 40 Hz in the F2 dimensions respectively, relaxation delay of 4 s and 16 scans. The HSQC experiments were acquired with a relaxation delay of 2 s, 2 K data points, 32 scans, 128 t_1 increments, and a spectral width of 12 ppm and 200 ppm in proton and carbon dimensions, respectively.

5.2.7 Metabolite identification and quantification

Metabolites were identified based on the 1D and 2D NMR recorded spectra, together with comparing the obtained chemical shift values with those listed in databases such as Biological Magnetic Resonance Data Bank (BMRB), the Madison Metabolomics Consortium Database (MMCD) and the Human Metabolite Data Base (HMDB) and the software packages MetaboHunter [182] and MetaboMiner [183]. Metabolites were quantified using the ratio method since NMR signal intensity is proportional to the molar concentration of metabolites:

$$\text{m(A)} = \frac{m(S) \times I(A) \times MW(A) \times N(S)}{I(S) \times MW(S) \times N(A)}$$

where, m(A), m(S) are metabolite mass and reference standard mass respectively, I(A), I(S) are metabolite and reference standard peak integrals respectively, MW(A), MW(S) are metabolite and reference standard molecular weights respectively, and N(A), N(S) are the number of metabolite protons or reference standard protons contributing to the corresponding peak integral, respectively. The calculated metabolite mass is represented as %w/w.

5.2.8 Multivariate statistical analysis

The NMR data analysis was performed using MestReNova software version 10.0.2 (Mestrelab Research, Spain). The spectra were binned into regions of equal spectral width (0.01 ppm). The spectral region δ 4.65–5.11 ppm contained residual water peaks and was not considered for the analysis. Data were normalized to total integral of 100 to remove differences due to different signal-to-noise ratios between the spectra. Principal component analysis (PCA), Orthogonal projections to latent structures discriminant analysis (OPLS-DA), one-way analysis of variance (ANOVA), statistical t-tests, and hierarchical clustering heat map analysis were performed using SIMCA version 14.1.0-2047 software (Umetrics, Umea, Sweden) and Metaboanalyst software http://www.metaboanalyst.ca/). To look for clustering, groupings, outliers and other patterns within the data, it was first treated with the unsupervised method of PCA (with Pareto-scaling). The data points located outside the 95% confidence region of the Hotelling's ellipse in the PCA score plot were removed prior to the subsequent analyses using the supervised method of OPLS-DA. To maximize group differentiation and identify clustering patterns, the data was then analyzed using OPLS-DA, which is used to maximize covariance between the measured data (NMR peak intensities) and the response variable (predictive classifications) and the OPLS-DA coefficient loading plot can be used to identify significant metabolites contributing to the separation between the two classes. The separation between the groups was observed using the OPLS-DA scores plot, having one predictive and one orthogonal component. Significant metabolites, responsible for group separation were then identified from the S-plot, which visualizes the covariance (p(ctr)) and correlation coefficient (abs(p(corr)) between the variables and the classification score in the model. Metabolites with (abs(p(corr)[1]) > 0.6) in the S-plot were considered to be statistically significant. R2Y (measuring goodness of fit) and Q2 (measuring predictability) values were used to assess the quality of the OPLS-DA model obtained, which provide the variance explained and the variance predicted by the model, respectively. The performance of the OPLSA-DA model was evaluated using CV-ANOVA (cross-validated ANOVA [252], where a p-value < 0.05 was considered to be statistically significant to validate the OPLSA-DA model. This was followed by the permutation analysis on the best model, with a threshold p-value of < 0.05 considered to be statistically significant to indicate that none of the results are better than the original one. Heat map and hierarchical cluster analysis (HCA) were also used to look for natural clustering in the time points and to gain an understanding of the trend in metabolite cycling across the 24 h circadian cycle. The hierarchical tree was constructed using Ward's minimum variance and the Euclidean distance. For better visualization of grouping and pattern, a polar dendrogram was generated using a MATLAB script [121], [253], [254], [255]. The significant metabolites identified using multivariate analysis were further tested for their statistical significance using the

univariate method of t-test, where a p-value < 0.05 was considered to be significant. A p-value correction method of Benjamini and Hochberg was applied to the data [256]. Metabolic pathways involved were identified and visualized using the Metabolomic Pathway Analysis (MetPA) software package (http://metpa.metabolomics.ca).

5.2.9 UPLC–ESI-MS experiments

Mass spectrometric experiments were performed to support the metabolite identification obtained from the NMR data analysis. Fresh samples of Bougainvillea spectabilis leaves were prepared, using the same procedure as described above for the NMR analysis. UPLC–ESI-MS (Ultra performance liquid chromatography–electrospray ionization-mass spectrometry) analysis was performed with a UPLC (Waters Acquity Class I) and mass spectrometry (Waters Synapt G2) system. Reverse phase chromatographic separation was performed using a ACQUITY UPLC (Column type: BEH Amide 2.1 $\times$ 150 mm, C18 1.7 μm) system and LC/MS grade, 0.1-μm membrane filtered solvents were used for mobile phases. Elution buffers were as follows: (A) water with 0.1% formic acid and (B) acetonitrile (CH_3CN) with 0.1% formic acid. Linear gradients from 95% A/5% B to 5% A/95% B over 30 min were used after the fixed initial conditions. The instrumental parameters were selected as follows: injection volume 10μL with flow rate 0.2mL/min, capillary voltage 2.00kV, source temperature 150^0C, sampling cone voltage 40.0 eV, desolvation temperature 350 ^{0}C, cone gas flow 61.0 L/h, desolvation gas flow 625.0 L/h. ESI-MS was conducted in the negative ion mode with lock spray configuration for recording the spectra between m/z 50 and 1500. The lock spray reference compound used was leucine enkephalin (554.2615 Da in negative mode) for calibration of all the masses. Samples were extracted using 1 mL of MeOH with 10 mg dried leaves of Bougainvillea spectabilis. The mixture was micro-centrifuged at 5000 rcf for 10 min. Afterwards, the 1 mL of the supernatant was transferred to a 1.5 mL Eppendorf tube. The spectra were processed and analyzed using the MassLynx^{TM} v4.1 software package.

5.2.10 Free radical scavenging assay

Free radical scavenging activity was measured from a 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay using the method described by [123], with some modifications. A stock solution at 400 μg/mL concentration of the sample was prepared and then serial dilution was performed from 12.5 to 400μg/mL. As a positive control, 2mg of quercetin was dissolved in 2.5 mL distilled water. 1 mL methanol leaf extract/standard of different concentrations was mixed in a vial with 3 mL solution of DPPH in methanol. The vials were incubated at room temperature for 30 min in the dark to complete the reac-

tion. Absorbance was of the solution was measured at 517 nm against a blank solution using a UV–Vis spectrophotometer. The result was expressed as mg quercetin equivalent antioxidant content (mg QEAC) per g dry weight (g DW). The DPPH radical scavenging activity (SC%) was calculated as follows:

$$\text{SC } \% = \frac{A_0 - A_s}{A_0} \times 100$$

where, A_0 is the absorbance of the reagent blank and A_s is the absorbance of the sample. The results were expressed as EC_{50} (the half maximal effective concentration), which denotes the concentration of the sample required to scavenge 50% of the DPPH free radicals.

5.2.11 ABTS radical cation scavenging assay

The antioxidant activity of Bougainvillea spectabilis leaf extracts was measured via an 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) diammonium salt (ABTS+) cation radical decolorization assay using the method reported by [124], with some modifications. A stock solution of ABTS+ radical was prepared by mixing ABTS (7 mM) with potassium persulfate (2.45 mM) in distilled water. The solution was kept in the dark at room temperature for 12–16 h before use, and then diluted with methanol to obtain an absorbance of 0.70 ± 0.02 units at 734 nm on the UV–Vis spectrophotometer. Methanol leaf extracts (10 μL of each sample) were mixed thoroughly with 2.99 mL of ABTS radical working solution. Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) was used an antioxidant standard and 2 mg of trolox was dissolved in 2.5 mL distilled water and then serial dilutions were performed of different concentrations 31.25–500μg/mL. ABTS scavenging ability was expressed as half-maximal effective concentration (EC_{50} in μg/mL).

5.3 Results and Discussions

5.3.1 Metabolite identification via 1D and 2D NMR

1D and 2D NMR spectra were recorded to identify the presence of a wide variety of metabolites in Bougainvillea spectabilis leaves. Fig 5.2 shows the representative 1D ^{1}H NMR spectrum of Bougainvillea spectabilis leaves, recorded at 600 MHz. Fig 5.3 represent 1D ^{1}H NMR spectrum of control & polluted Bougainvillea spectabilis leaves, recorded at 600 MHz. The upfield region from δH 0.8 to 3.5 ppm, showed peaks mainly for amino acids, organic acids and lipids. The amino acids alanine, valine,

leucine, glutamine, glutamic acid, phenylalanine, cysteine, proline and γ-aminobutyrate (GABA) were identified and quantified in all the replicates. Organic acids identified in this region include propionic acid, acetic acid, malic acid, citric acid, chlorogenic acid, and quinic acid. Phenolics identified in this region include hesperidin and rutin. The mid-field region from δH 3.5 to 5.5 ppm contained peaks mainly for carbohydrates which were mostly glucose (α and β-configurations), fructose, sucrose, raffinose and an sugar alcohol myo-inositol. The downfield region of δH beyond 6.0 ppm contained peaks mainly for aromatic compounds such as aromatic amino acids and phenolics. These signals were relatively weak and were from aromatic moieties from amino acids such as tyrosine and tryptophan. Organic acids such as fumaric acid and formic acid were also identified in this region as well as phenolics such as coumaric acid, gallic acid and flavonoids such as catechin and kaempferol. Other metabolites identified include choline (an essential component of cell membrane biosynthesis), adenine (a nitrogenous base), adenosine monophosphate (AMP), nicotinamide adenine dinucleotide (NAD), methyl salicylate, putrescine (a polyamine), and trigonelline (which is derived from the methylation of the nitrogen atom of niacin which is a naturally occurring vitamin B complex). Table 5.1 and Table 5.2 summarize the assignments of the leaf NMR spectrum, with a wide range of primary and secondary plant metabolites including amino acids, carbohydrates and phenolic compounds. The presence of all the metabolites suggested by the 1D ^{1}H NMR spectra were further confirmed by their peaks in 2D experiments such as COSY, TOCSY, JRES and HSQC.

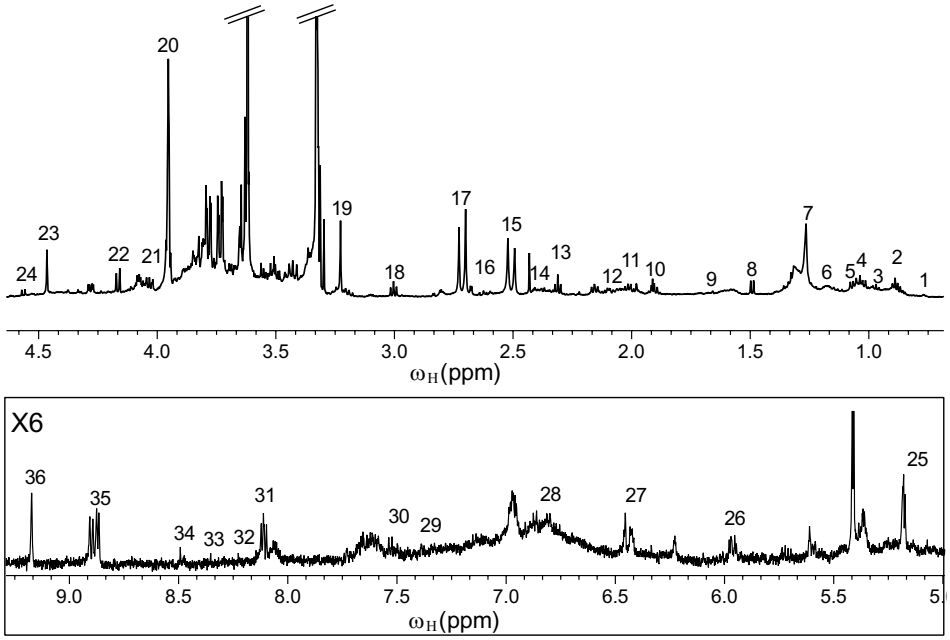

Figure 5.2: 1D ^{1}H NMR spectrum of Bougainvillea spectabilis leaves, recorded at 600 MHz showing specific resonances of metabolites identified. Peaks numbering: 1 – phytosterol, 2 – lipid terminal methyl, 3 – leucine, 4 – valine, 5 – hesperidin, 6 – rutin, 7 – lipid (CH_2)n, 8 – alanine, 9 – lipid (CH_2CH_2COOH), 10 – quinic acid, 11 – chlorogenic acid, 12 – glutamine/glutamic acid, 13 – GABA, 14 – glutamic acid, 15 – citric acid, 16 – malic acid, 17 – citric acid, 18 – putrescine, 19 – choline, 20 – methyl salicylate, 21 – inositol, 22 – sucrose, 23 – trigonelline, 24 – beta-glucose, 25 – alpha-glucose, 26 – NAD, 27 – fumaric acid, 28 – tyrosine, 29 – tryptophan, 30 – tryptophan, 31 – kaempferol, 32 – adenine, 33, 34 – AMP, 35, 36 – trigonelline.

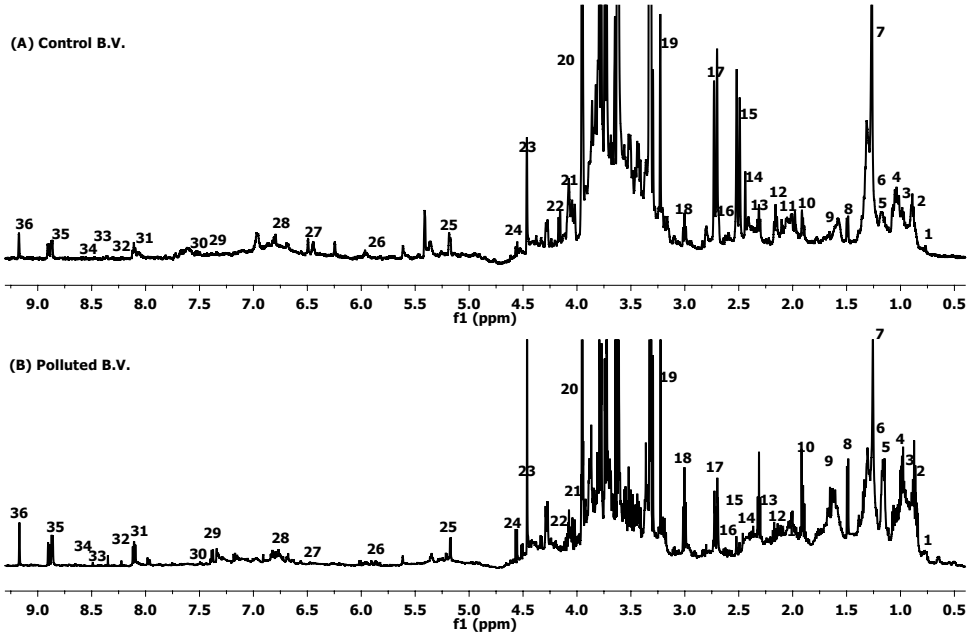

Figure 5.3: (A) Control, (B) Polluted: 1D ^{1}H NMR spectrum of Bougainvillea spectabilis leaves, recorded at 600 MHz showing specific resonances of metabolites identified. Peaks numbering: 1 – phytosterol, 2 – lipid terminal methyl, 3 – leucine, 4 – valine, 5 – hesperidin, 6 – rutin, 7 – lipid (CH_2)n, 8 – alanine, 9 – lipid (CH_2CH_2COOH), 10 – quinic acid, 11 – chlorogenic acid, 12 – glutamine/glutamic acid, 13 – GABA, 14 – glutamic acid, 15 – citric acid, 16 – malic acid, 17 – citric acid, 18 – putrescine, 19 – choline, 20 – methyl salicylate, 21 – inositol, 22 – sucrose, 23 – trigonelline, 24 – beta-glucose, 25 – alpha-glucose, 26 – NAD, 27 – fumaric acid, 28 – tyrosine, 29 – tryptophan, 30 – tryptophan, 31 – kaempferol, 32 – adenine, 33, 34 – AMP, 35, 36 – trigonelline.

Table 5.1: Primary Metabolites identified from ^{1}H NMR and 2D NMR spectrum of Bougainvillea spectabilis leaves, with chemical shift given in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	**Chemical shifts (m, J)**
Term. methyl group	0.86 (t, J = 6.95), 0.95 (t, J = 6.9)
(CH_2)n	1.27 (m)
$CH_2CH_2COOH(C_3)$	1.64 (m)
CH_2COOH (C_2)	2.37 (t, J = 7.5)
Fatty acid	1.64 (m)
Fatty acid	5.3 (m)
Leucine	0.95 (t, J = 6), 1.9 (m)
Valine	1.06 (d, J = 7), 0.97 (d, J = 7)
Alanine	1.49 (d, J = 7.3), 3.76 (q, J = 7.2)
Glutamine	2.15 (m), 2.47 (m)
Glutamic acid	2.15 (m), 2.0 (m)
Phenylalanine	3.0 (m), 3.98 (dd, J = 7.8, 5.3)
Tyrosine	6.86 (m), 7.1 (m)
Tryptophan	7.32 (s), 7.54 (d, J = 7.54)
GABA	1.90 (m), 2.30 (t, J = 7.2)
Cysteine	3.06 (m), 3.97 (dd)
Proline	4.03 (dd, J = 8.4, 8.4), 1.99 (m)
Beta-glucose	4.58 (d, J = 7.8), 5.22 (d, J = 3.6)
Alpha-glucose	5.18 (d, J = 3.8), 4.63 (d, J = 7.98)
Sucrose	4.17 (d, J = 8.5), 5.4 (d, J = 3.8)
Fructose	3.5 (d, J = 11.77), 3.62 (m)
Raffinose	5.4 (d, J = 3.8), 4.98 (d, J = 3.8)
Propionic acid	1.04 (t, J = 7.6), 2.17 (q, J = 7.6)
Acetic acid	1.96 (s)
Malic acid	2.68 (dd, J = 16.6, 6.6), 2.36 (dd)
Citric acid	2.56 (d, J = 17.6), 2.74 (d, J = 17.6)
Fumaric acid	6.56 (s)
Formic acid	8.46 (s)
Quinic acid	1.84 (dd, J = 13.4, 10.8), 2.05 (m)

Abbreviations: GABA (gamma-aminobutyric acid), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

Table 5.2: Secondary and other metabolites identified from ^{1}H NMR and 2D NMR spectrum of Bougainvillea spectabilis leaves, with chemical shift given in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)
Phytosterol	0.69 (s), 2.02 (m), 5.3 (m)
Gallic acid	7.06 (s)
Chlorogenic acid	7.17 (d, J = 1.8), 4.25 (d, J = 2.58), 3.89 (dd), 2.13 (m), 2.02 (m)
Hesperidin	1.08 (dd), 3.11 (m)
Rutin	1.12 (d, J = 6.22), 3.31 (m)
m-Coumaric acid	6.48 (d, J = 28), 7.16 (d, J = 7.69)
Catechin	4.55 (d, J = 7.90), 2.5 (d, J = 16, 8)
Kaempferol	6.18 (d, J = 2.09), 8.08 (m)
Adenine	8.11 (s), 8.21 (s)
Inositol	4.00 (t, J = 2.8)
AMP	8.26 (s), 8.59 (s)
NAD	6.03 (d, J = 5.8), 6.08 (d, J = 5.7) 6.12 (d, J = 5.8), 8.16 (s)
Trigonelline	4.46 (s), 8.89 (m), 9.17 (s)
Methyl salicylate	3.92 (s), 6.87 (t, J = 7.6)
Putrescine	1.75 (m), 3.0 (t, J = 7.29)
Choline	3.22 (s)

Abbreviations: AMP (Adenosine monophosphate), NAD (Nicotinamide adenine dinucleotide), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

5.3.2 UPLC–ESI-MS analysis

We performed a targeted mass-spectrometric analysis to identify the presence of peaks corresponding to the metabolites of our interest. The presence of the metabolites in Bougainvillea spectabilis leaves on the detection of $[M\text{-}H]^-$ ions with retention time is reported in Table 5.3. We identified amino acids, amino acid derivative, organic acids, sugars, nucleobases, vitamins and phenolic compounds at different retention times. The first four compounds to be identified at R_t = 0.691 min with m/z = 89, 111, 191, 341 were lactic acid, uracil, quinic acid and sucrose, respectively. This was followed by the detection of citric acid and lactose at R_t = 1.156 min with m/z 191 and 341, respectively. The next compounds to be identified were riboflavin, rutin and glutathione oxidized at R_t = 3.016 min with m/z 375, 609, 611, respectively. Naringenin, chlorogenic acid and hesperidin peaks were identified with m/z 271, 353, 609 respectively at R_t = 3.876 min. A few intense peaks, which did not emerge from the sample and were present in the mass spectra of the blank run as well, were attributed to the background signals from solvents and other unidentified contaminants.

Table 5.3: Compounds present in Bougainvillea spectabilis leaves extracts identified using UPLC-ESI-MS in negative ionization mode.

Metabolites	M.W.	$[M-H]^-$	R_t(min.)
Lactic acid	90	89	0.691
Uracil	112	111	0.691
Quinic acid	192	191	0.691
Sucrose	342	341	0.691
Citric acid	192	191	1.156
Lactose	342	341	1.156
Riboflavin	376	375	3.016
Rutin	610	609	3.016
Glutathione oxidized	612	611	3.016
Naringenin	272	271	3.876
Chlorogenic acid	354	353	3.876
Hesperidin	610	609	3.876

5.3.3 Quantifying free radical scavenging activity

A DPPH assay was performed to determine the antioxidant capacity of the Bougainvillea spectabilis control and polluted leaves, to act as hydrogen donors or free radical scavengers. The % inhibition of DPPH radical increased with increase in concentration of the extract for all the leaf samples Table 5.4. The free radical scavenging activity of the polluted leaf samples was higher as compared to the control leaf samples, indicating a higher concentration of antioxidants in the polluted leaves Fig 5.4.

Table 5.4: Mean of EC_{50} (μg/ml) of DPPH free radical scavenging activities of Bougainvillea spectabilis leaves extracts (control and polluted region).

Leaf extract	EC_{50} (μg/ml)
Control leaves	147.30 ± 4.41
Polluted leaves	93.80 ± 2.81

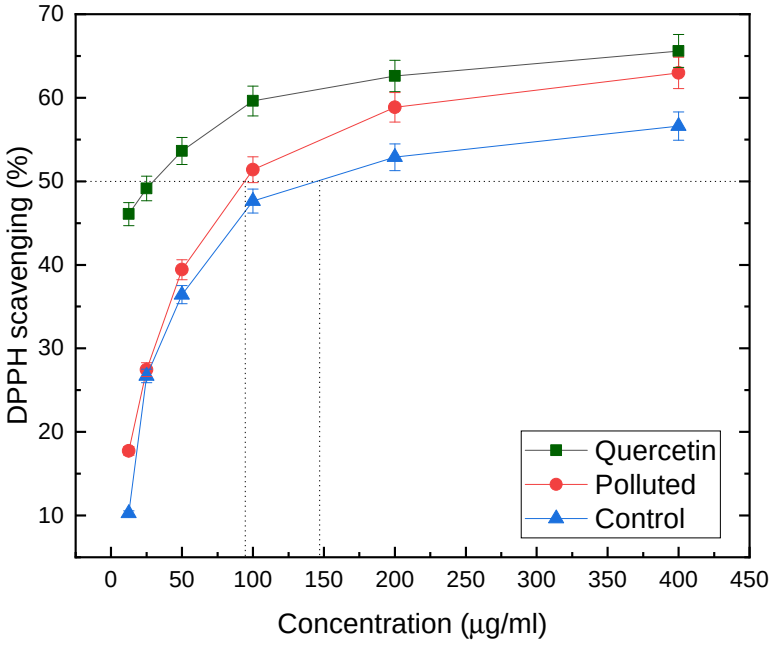

Figure 5.4: DPPH scavenging activities (%) of the methanol extracts of the Bougainvillea spectabilis leaves (control and polluted samples). The quercetin was used as standard. Values are expressed as mean ± SD for triplicates.

The antioxidant potential of Bougainvillea spectabilis leaves was also quantified using an ABTS assay with the scavenging ability expressed as EC50 (μg/ mL). The extract of both control and polluted Bougainvillea spectabilis leaf samples were found to be effective in scavenging free radicals and the scavenging activity increased with increasing concentration Table 5.5. The ABTS radical scavenging activity of the polluted leaf samples was higher as compared to the control leaf samples Fig 5.5, corroborating the results from the NMR analysis, as well as the DPPH assay.

Table 5.5: Mean of EC_{50} (μg/ml) of ABTS free radical scavenging activities of Bougainvillea spectabilis leaves extracts (control and polluted region).

Leaf extract	**EC_{50} (μg/ml)**
Control leaves	223.59 ± 6.70
Polluted leaves	113.33 ± 3.39

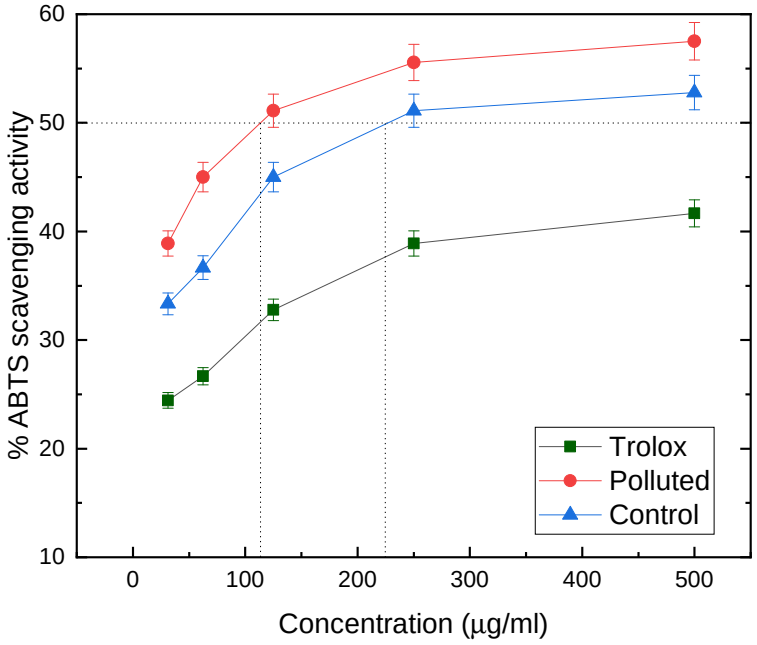

Figure 5.5: Antioxidant potential of Bougainvillea spectabilis leaves (polluted and control region leaves). Using ABTS assay, Trolox served as a positive control. Values are expressed as mean ± SD for triplicates.

5.3.4 Multivariate data analysis

To identify metabolites that show cycling in concentrations, Bougainvillea spectabilis leaf samples were collected at thirteen time points, at a gap of 2 h each, from the IISER Mohali campus (which is the designated non-polluted site) and labeled as controls. The samples from all the thirteen time points initially compared using the unsupervised method of PCA Fig 5.6 showed no grouping or clustering patterns and had two replicates as outliers, which were removed from subsequent analysis.

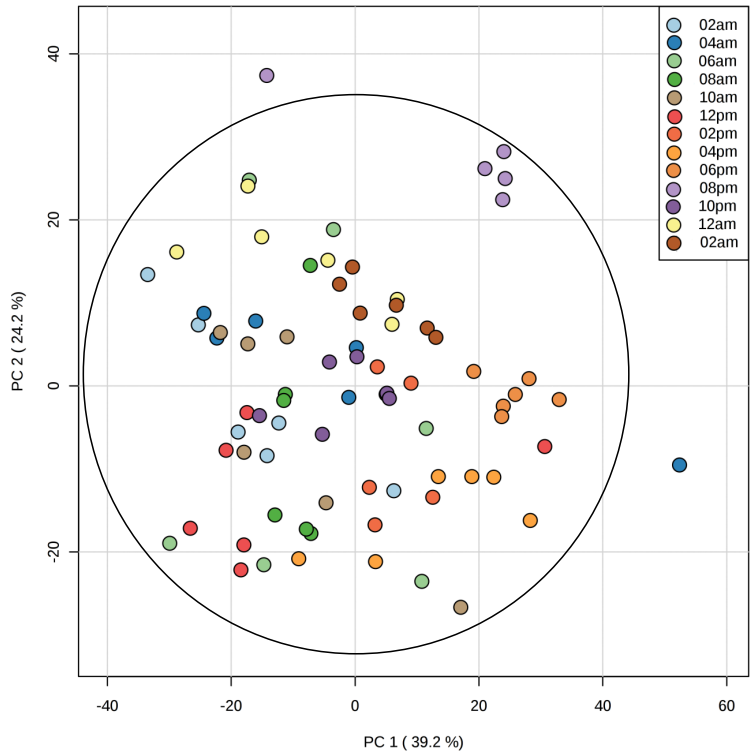

Figure 5.6: Principal component analysis (PCA) score plot derived from ^{1}H NMR spectra of Bougainvillea spectabilis leaves of all 13 time points with PC1 and PC2 showing variation 39.2 % and 24.2 % respectively.

To better visualize the clustering of the time points, a polar dendrogram was plotted to radially map the leaf nodes onto the circumference of a circle. Three main nodes resulting in three time groupings can be identified in the polar dendrogram shown in Fig 5.7. The first group had times corresponding to early morning (times 12 AM, 2 AM, 4 AM, 6 AM) while the second group had times corresponding to morning–afternoon (at times 8AM, 10AM, 12PM, 2PM). A third grouping is of four time points (4 PM, 6 PM, 8 PM, 10 PM) which are different from the other two groups and fall in-between the two time zones. This indicates that these four time points are different yet intermediate to the other two groups and could be the region in which the Bougainvillea spectabilis metabolism changes from the day to the night phases. Since the two major clusters have time points that are 10–12 h apart, indicating a possible 10–12 h cycling in metabolite concentrations, we focused on comparing leaf samples at 8 AM and 8 PM which are 12 h apart. These are also the time points at which road

traffic is the heaviest, leading to plant exposure to high levels of air pollution.

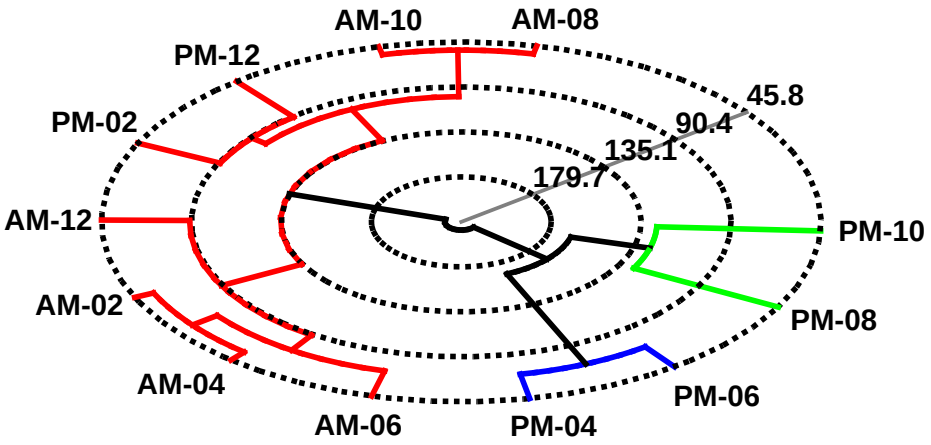

Figure 5.7: Clustering of control samples of Bougainvillea spectabilis at 12 different time points is represented on a polar dendrogram. All time points have six replicates each. The significant clusters are indicated using Euclidean distance between pairs of observations. There are three significant clusters: early morning (12 AM–6 AM), morning to early afternoon (8 AM–2 PM), and late afternoon to night (4 PM–10 PM).

Heat map and hierarchical cluster analysis (HCA) were also used to look for natural clustering in the time points as shown in Fig 5.8. The dendrogram is shown on the top of the heat map and reveals the connection between the different clusters of time points, based on their metabolite concentration levels. The tree dendrogram obtained from HCA encapsulates the variation present in the dataset. To obtain a clear picture of the metabolite rhythms, we considered the mean of all replicates per time point instead of analyzing individual replicates which reduced to 12 data points instead of 72 (12 $\times$ 6 =72) corresponding to 6 replicates per time point sampled. The time points are indicated along the x-axis, with the labels: a: 12 AM, b: 02 AM, c: 04 AM, d: 06 AM, e: 08 AM, f: 10 AM, g: 12 PM, h: 02 PM, i: 04 PM, j: 06 PM, k: 08 PM, and l: 10 PM. The statistically significant variables (representative of metabolites) are presented along the y-axis and the major classes are color-coded based on the measured concentrations. The intensity of each bin scaled between 2 (red) to -2 (green) denoting metabolite concentrations that are either significantly higher or lower in the samples. The HCA analysis indicated that the metabolite intensity profiles showed a variation over all the time points.

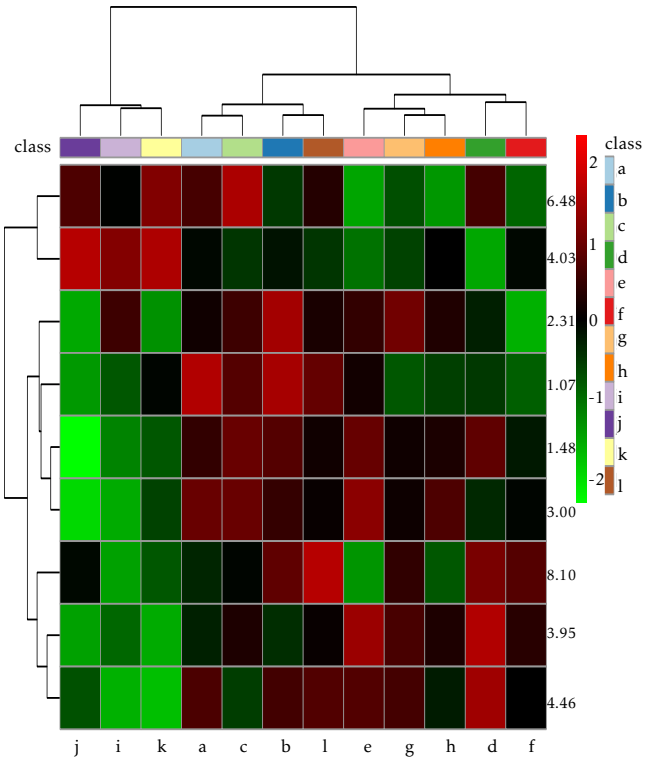

Figure 5.8: Heat map and dendrogram constructed using hierarchical cluster analysis (HCA) for statistically significant variables from Bougainvillea spectabilis leaves (control samples) at different time points during the circadian cycle. Columns correspond to different time points (a: 12 AM, b: 02 AM, c: 04 AM, d: 06 AM, e: 08 AM, f: 10 AM, g: 12 PM, h: 02 PM, i: 04 PM, j: 06 PM, k: 08 PM, and l: 10 PM). The rows correspond to statistically significant variables (representative of metabolites). Positive (red) and negative (green) correlations denote an increase or decrease in the metabolite concentrations, respectively. The major contributing signals are from the bins δ 8.10, 3.95, 4.46 and 4.03 ppm.

Fig 5.9 shows the 3D score plot of the OPLS-DA analysis of 1D ^{1}H NMR spectra showing the metabolites responsible for the separation between leaf samples collected at time points: 4 AM, 10 AM, 4 PM, and 10 PM, during the circadian cycle. The OPLS-DA 3D plot shows a clear differentiation between the different time points in the circadian cycle, with clustering and separation of four time windows: early morning, late morning, afternoon and night. As seen from the figure, the total explained variation for the first three latent components was LC1 21.0%, LC2 18.7% and LC3 15.0%. The explained variation of each component is given by: R2X(cum) = 0.997, R2(cum)

= 0.632 and Q2(cum) = 0.174. Compared to the PCA model, the OPLS-DA models showed an improvement in predictability. Based on the polar dendrogram clustering and the predictive scores obtained from the OPLS-DA scores plot for each comparison, we selected two time points, namely 8 AM and 8 PM, which showed maximum difference in group clustering, for further analysis. Fig 5.10 (a) shows the OPLS-DA scores plot of the control samples for comparison between 8AM and 8PM, with one predictive and one orthogonal component showing a clear separation between the two time points. The R2X(cum), R2Y(cum) and Q2(cum) values for the model were 0.992, 0.861 and 0.697, respectively. The model was also found to be statistically significant after testing with CV-ANOVA permutation test (p-value < 0.05) which proved the credibility and robust nature of this model. Fig 5.10 (b) shows the corresponding S-plot with annotations denoting the significant metabolites. The metabolite peaks for leucine, valine, alanine, malic acid, sucrose and tyrosine were found to be present in higher concentrations and the metabolite peaks for glutamic acid, GABA, choline and trigonelline were found to be present in lower concentrations at 8AM, as compared to 8PM. The statistical significance of the metabolites identified from the OPLS-DA analysis was also confirmed by performing the univariate analysis of t-test, where a p-value < 0.05 was considered to be statistically significant.

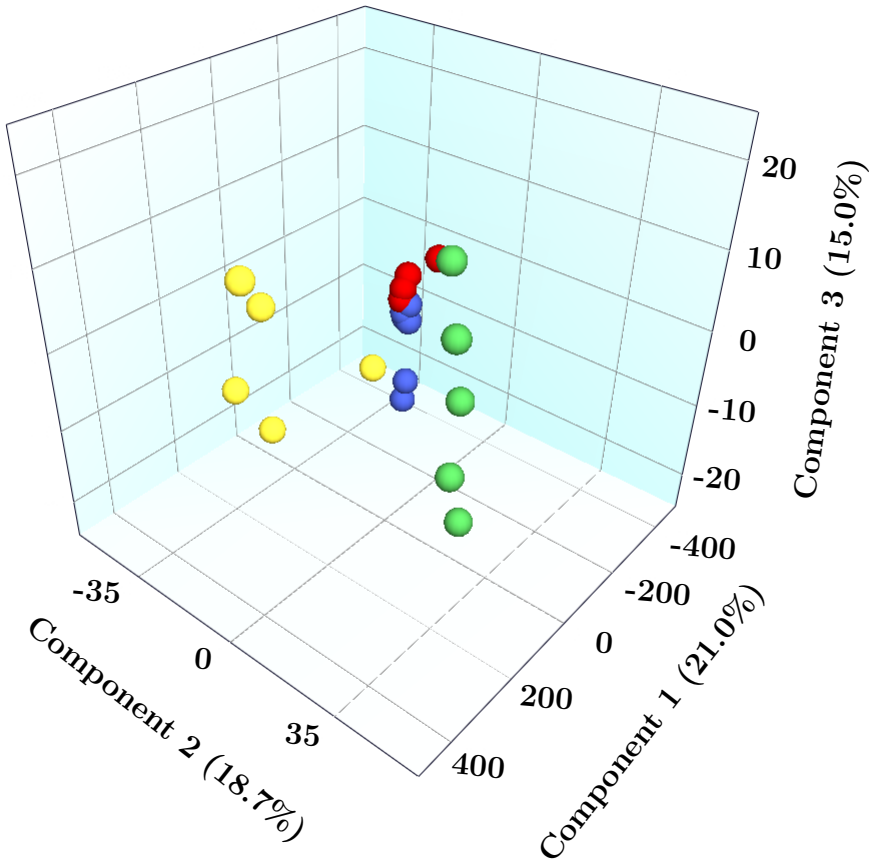

Figure 5.9: 3D OPLS-DA scores plot showing one predictive component (x-axis) and two orthogonal (y-axis, z-axis) components of Bougainvillea spectabilis leaves (control samples), showing clear metabolic differences between different time points in the circadian period. The various symbols represent samples collected at different time points during the circadian cycle: 10 AM (yellow circles), 4 AM (red circles), 10 PM (blue circles) and 4 PM (green circles).

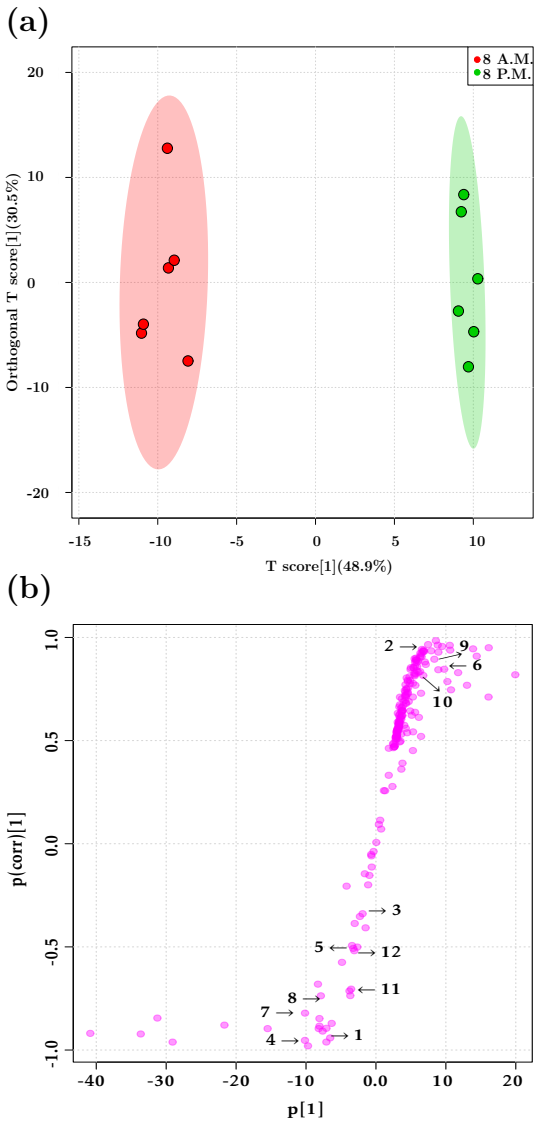

Figure 5.10: (a) OPLS-DA score plot obtained from ^{1}H NMR spectra of Bougainvillea spectabilis leaves (control samples) at two time points (8 AM and 8 PM). (b) S-plot showing concentration differences of the significant metabolites at 8 AM and 8 PM. The labels correspond to: 1, leucine; 2, valine; 3, alanine; 4, glutamic acid; 5, GABA; 6, malic acid; 7, choline; 8, sucrose; 10, tyrosine; 9, 11, 12, trigonelline.

5.3.5 Metabolite rhythms in control samples during circadian cycling

To confirm that the significant metabolites identified indeed show cycling with a period close to 24 h, intensities of these metabolites were quantified in the control samples of Bougainvillea spectabilis leaves. Changes in metabolite cycling rhythms over the 24 h circadian cycle were analyzed by identifying the spectral bins that changed significantly between the time points compared and quantifying the intensities of individual metabolites in these significant bins. Metabolite intensities were obtained from NMR spectra after normalizing and scaling of spectra to make all the time points comparable and are depicted in Fig 5.11. Tangible metabolite rhythms with characteristic peaks and troughs in their cycling patterns can be identified from the intensity plots. Specific metabolites that show cycling in tandem with the circadian clock are discussed below. Fumaric acid is a major form of fixed carbon and can hence be metabolized to generate energy, similar to starch and sugars [214]. In our study, fumaric acid concentration increases in the day, reaches a maximum at the end of light period, decreases and reaches a minimum at the end of the dark period. Furthermore, the concentration of fumaric acid remained at a steady minimum during the dark period, which could either be due to its production and degradation rates during night being equal or due to it being transported out of the leaves during the continuous dark period. Sugar levels in plants are influenced by light/dark periods [257]. The multivariate analysis revealed that sucrose levels remained at a minimum during the light period, then increased to a maximum amount toward the end of the light period. Subsequently, the sucrose level decreased during the dark period and remained at a minimum. These sugar stores are used by the plant during the night. The low concentration of sugars in the beginning of the diurnal cycle is related to the regulation of genes associated with the circadian clock. Glutamic acid levels depend on available nitrate, which is converted to ammonium in the leaves and reduced via the glutamine synthetase/glutamate synthase pathway to glutamic acid or glutamine [258]. Nitrate uptake is stimulated by light, which could be responsible for elevated glutamic acid levels during the day. Glutamic acid rhythmicity can also influence cycling of other amino acids. The concentration plots in Fig 5.11 show that amino acids such as proline, alanine and valine show a pronounced cycling in their concentrations. The non-protein amino acid γ-aminobutyric acid (GABA) is an important component in plant carbon metabolism when growth conditions are carbon-limited [259]. As is expected, GABA levels rose during the day, reached a maximum toward the end of the light period and subsequently decreased and remained minimum during the dark period. Choline is an important plant metabolite which is required in the synthesis of membrane phospholipids. Choline concentration was decreased during the light period and increased slightly and subsequently remained constant during the

dark period, which might be related to changes in the composition of membrane lipids during the day and the night in the plant [210]. Polyamines (PAs) have been known to help in retarding plant senescence and in stabilizing membrane and cell walls [260]. Putrescine is a polyamine which is a phytohormone like aliphatic amine which plays an important role in several plant physiological processes [261]. Recent research has suggested that PAs are activated when plants are exposed to multiple types of environmental stresses such as oxidative damage, adverse soil conditions and metal toxicity [262]. In Bougainvillea spectabilis leaves, putrescine levels showed a double-cycling pattern i.e. they increased during the day, reached a maximum and decreased toward the end of the day. This cycling pattern was repeated during the night. Trigonelline (N-methyl nicotinamide) is a metabolite which is metabolized from nicotinamide via amidohydrolase and methyltransferase pathways and plays an important role in several physiological processes in plants [263]. Trigonelline is considered an active hypoglycemic agent [264], acts as an osmolyte [265], and is involved in leaf closure mechanisms and hence is closely associated with the plant circadian rhythm [266]. In the control samples of Bougainvillea spectabilis leaves, trigonelline levels in general decreased during the day, then increased and reached a maximum at the end of the light period and then decreased again during the dark period.

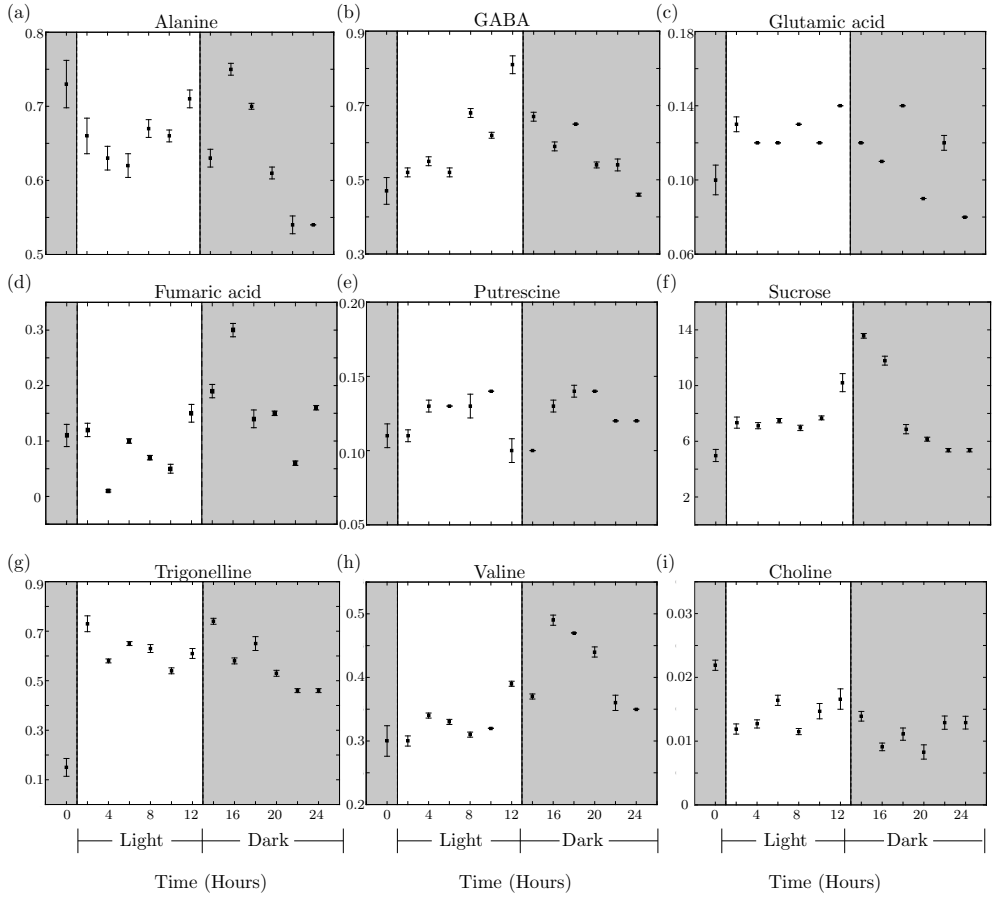

Figure 5.11: Plots of relative concentration of cycling metabolites sampled in Bougainvillea spectabilis leaves (control samples) at 2 h intervals across 24 h. Each data point is the average of 6 replicates and error bars are standard errors of the mean (SEM). The gray and white color boxes represent the night and day time points, respectively.

5.3.6 Altered leaf metabolism in polluted samples after prolonged exposure to air pollution

After identifying the metabolites which cycle in the control leaf samples, we proceeded toward analyzing whether the concentrations of these metabolites vary significantly in the plants exposed to high levels of air pollution. We collected Bougainvillea spectabilis leaf samples at two time points: 8 AM and 8 PM from the polluted site and compared their metabolic profiles with those of control samples obtained at the same time points. Fig 5.12 shows the OPLS-DA scores plot for comparison between control

(C) and polluted (P) leaf samples, showing a clear separation between the two sample sets. The S-plot (data not shown) was used to identify the significant metabolites responsible for differentiation between control (C) and polluted (P) samples. Table 5.6 and Table 5.7 contain the list of metabolites and their concentrations which are significantly different in the control and the polluted samples at 8 AM and 8 PM, respectively. The statistical significance of the metabolites identified was again confirmed from the t-test ($p<0.05$) as before. At 8 AM, the levels of valine, alanine, leucine, glutamic acid, GABA, putrescine, choline, trigonelline, glucose, fructose, rutin, quinic acid and malic acid were significantly increased in the polluted (P) samples as compared to the control samples. Concomitantly, the levels of sucrose, fumaric acid, citric acid, chlorogenic acid, raffinose, methyl salicylate and kaempferol were decreased in the polluted samples, as compared to the control samples. At 8 PM, after exposure to high levels of vehicular traffic for several hours during the day, the picture remained nearly the same, except for a slight decrease in the levels of valine, malic acid and fructose in the polluted samples as compared to the control samples, and a slight increase in the level of kaempferol in the polluted samples.

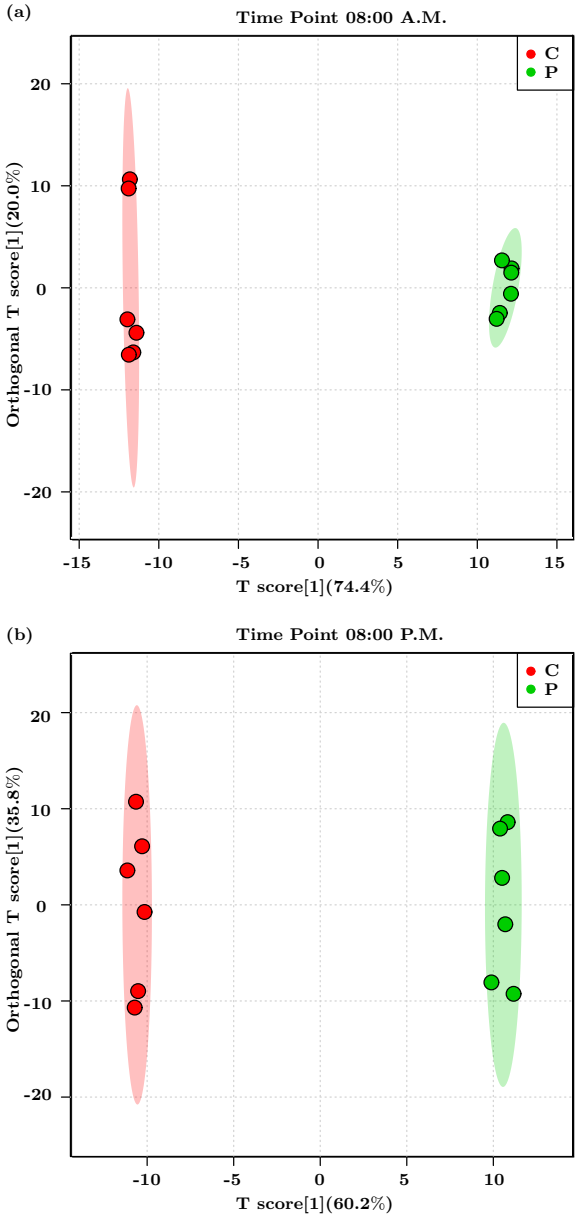

Figure 5.12: OPLS-DA score plot obtained from ^{1}H NMR spectra of control (C) and polluted (P) Bougainvillea spectabilis leaf samples, at two time points (a) 8 AM and (b) 8 PM.

Table 5.6: Relative concentrations (% w/w) of the significant metabolites in control (C) and polluted (P) Bougainvillea spectabilis leaves at 8:00 AM. The asterisk ($*$) denotes a cycling metabolite and $\uparrow$, $\downarrow$ denote increase or decrease in metabolite concentration in P samples, respectively. The superscripts: a; denotes the statistical significance p ≤ 0.05 and b; give the cutoff values for the correlation of $\mid p(corr) \mid \geqslant 0.5$.

Metabolites	**C**	**P**
Valine*	$0.37 \pm 0.00^{a,b}$	$0.78 \pm 0.04^{a,b}$ $\uparrow$
Alanine*	$0.69 \pm 0.02^{a,b}$	$2.20 \pm 0.14^{a,b}$ $\uparrow$
Leucine	$0.05 \pm 0.00^{a,b}$	$0.13 \pm 0.00^{a,b}$ $\uparrow$
Glutamic acid*	$0.12 \pm 0.00^{a,b}$	$0.22 \pm 0.01^{a,b}$ $\uparrow$
GABA*	$0.60 \pm 0.01^{a,b}$	$1.46 \pm 0.08^{a,b}$ $\uparrow$
Putrescine*	$0.15 \pm 0.00^{a,b}$	0.46 ± 0.02^{a} $\uparrow$
Choline*	$0.03 \pm 0.00^{a,b}$	$0.04 \pm 0.00^{a,b}$ $\uparrow$
Methyl salicylate	5.20 ± 0.20^{a}	4.84 ± 0.30^{a} $\downarrow$
Sucrose*	$7.62 \pm 0.73^{a,b}$	$6.48 \pm 0.44^{a,b}$ $\downarrow$
Trigonelline*	$0.70 \pm 0.01^{a,b}$	$1.26 \pm 0.08^{a,b}$ $\uparrow$
Fumaric acid*	0.12 ± 0.01^{a}	0.05 ± 0.00^{a} $\downarrow$
Kaempferol	0.78 ± 0.04^{a}	0.65 ± 0.14^{a} $\downarrow$
Rutin	4.12 ± 0.42^{a}	5.50 ± 0.41^{a} $\uparrow$
Quinic acid	1.49 ± 0.07^{a}	2.89 ± 0.17^{a} $\uparrow$
Glucose	$0.08 \pm 0.01^{a,b}$	$0.12 \pm 0.01^{a,b}$ $\uparrow$
Malic acid	$4.98 \pm 0.10^{a,b}$	$5.40 \pm 0.32^{a,b}$ $\uparrow$
Fructose	$6.00 \pm 0.30^{a,b}$	$6.30 \pm 0.40^{a,b}$ $\uparrow$
Chlorogenic acid	4.12 ± 0.28^{a}	2.80 ± 0.23^{a} $\downarrow$
Raffinose	3.92 ± 0.83^{a}	0.76 ± 0.11^{a} $\downarrow$
Citric acid	3.41 ± 0.08^{a}	1.65 ± 0.10^{a} $\downarrow$

Table 5.7: Relative concentrations (% w/w) of the significant metabolites in control (C) and polluted (P) Bougainvillea spectabilis leaves at 8:00 PM. The asterisk ($*$) denotes a cycling metabolite and $\uparrow$, $\downarrow$ denote increase or decrease in metabolite concentration in P samples, respectively. The superscripts: a; denotes the statistical significance p ≤ 0.05 and b; give the cutoff values for the correlation of $\mid p(corr) \mid \geqslant 0.5$.

Metabolites	C	P
Valine*	$0.49 \pm 0.02^{a,b}$	$0.48 \pm 0.03^{a,b}$ $\downarrow$
Alanine*	$0.75 \pm 0.02^{a,b}$	$1.28 \pm 0.10^{a,b}$ $\uparrow$
Leucine	$0.05 \pm 0.00^{a,b}$	$0.08 \pm 0.00^{a,b}$ $\uparrow$
Glutamic acid*	$0.11 \pm 0.00^{a,b}$	$0.16 \pm 0.01^{a,b}$ $\uparrow$
GABA*	$0.62 \pm 0.02^{a,b}$	$0.94 \pm 0.07^{a,b}$ $\uparrow$
Putrescine*	0.16 ± 0.01^{a}	0.26 ± 0.01^{a} $\uparrow$
Choline*	$0.02 \pm 0.00^{a,b}$	$0.03 \pm 0.00^{a,b}$ $\uparrow$
Methyl salicylate	4.32 ± 0.15^{a}	4.51 ± 0.42^{a} $\uparrow$
Sucrose*	$10.67 \pm 0.41^{a,b}$	$7.55 \pm 0.86^{a,b}$ $\downarrow$
Trigonelline*	$0.68 \pm 0.03^{a,b}$	$1.23 \pm 0.12^{a,b}$ $\uparrow$
Fumaric acid*	$0.24 \pm 0.01^{a,b}$	0.11 ± 0.02^{a} $\downarrow$
Kaempferol	0.64 ± 0.06^{a}	0.68 ± 0.16^{a} $\uparrow$
Rutin	$5.12 \pm 0.31^{a,b}$	$6.63 \pm 0.81^{a,b}$ $\uparrow$
Quinic acid	$1.38 \pm 0.05^{a,b}$	$2.25 \pm 0.16^{a,b}$ $\uparrow$
Glucose	$0.12 \pm 0.01^{a,b}$	$0.18 \pm 0.02^{a,b}$ $\uparrow$
Malic acid	7.09 ± 0.27^{a}	4.54 ± 0.34^{a} $\downarrow$
Fructose	$8.34 \pm 0.34^{a,b}$	$6.32 \pm 0.47^{a,b}$ $\downarrow$
Chlorogenic acid	$6.30 \pm 0.35^{a,b}$	$4.12 \pm 0.57^{a,b}$ $\downarrow$
Raffinose	8.01 ± 0.30^{a}	$1.32 \pm 0.25^{a,b}$ $\downarrow$
Citric acid	4.35 ± 0.17^{a}	1.47 ± 0.15^{a} $\downarrow$

Pathway analysis was performed on the polluted leaf samples of Bougainvillea spectabilis, to identify patterns in the data and the metabolic pathways that are associated with plant exposure to air pollution stress and show dysregulation. Metabolic pathways were analyzed using Metabolomic Pathway Analysis (MetPA) which is based on data from HMDB and from Kyoto Encyclopedia of Genes and Genomes (KEGG). Differential metabolite enrichment networks were constructed based on significant metabolites. Nodes represent the correlated metabolic pathways and the edges represent the biological relationship between two pathway associated metabolites. Results were represented graphically with metabolite concentrations as X variables and other parameters as Y variables and are shown in Fig 5.13 (a), where log(p) denotes the transformed p-value calculated from the enrichment analysis and pathway impact value is proportional to the bubble size and bubble color. Panel (b) in Fig 5.13 shows the correlation network with metabolites as nodes and correlations as edges. The nodes in the network were shaded in color intensity according to their p-value. As is evidenced from Fig 5.13, the major metabolic pathways that are disturbed due to prolonged exposure to air pollution are the branched amino acid biosynthesis, the citrate cycle, arginine and proline metabolism and alanine, aspartate and glutamate metabolism. We summarize below the effect of metabolic pathway disruption on the enhancement or decrease in production of specific metabolites and the implications for our study.

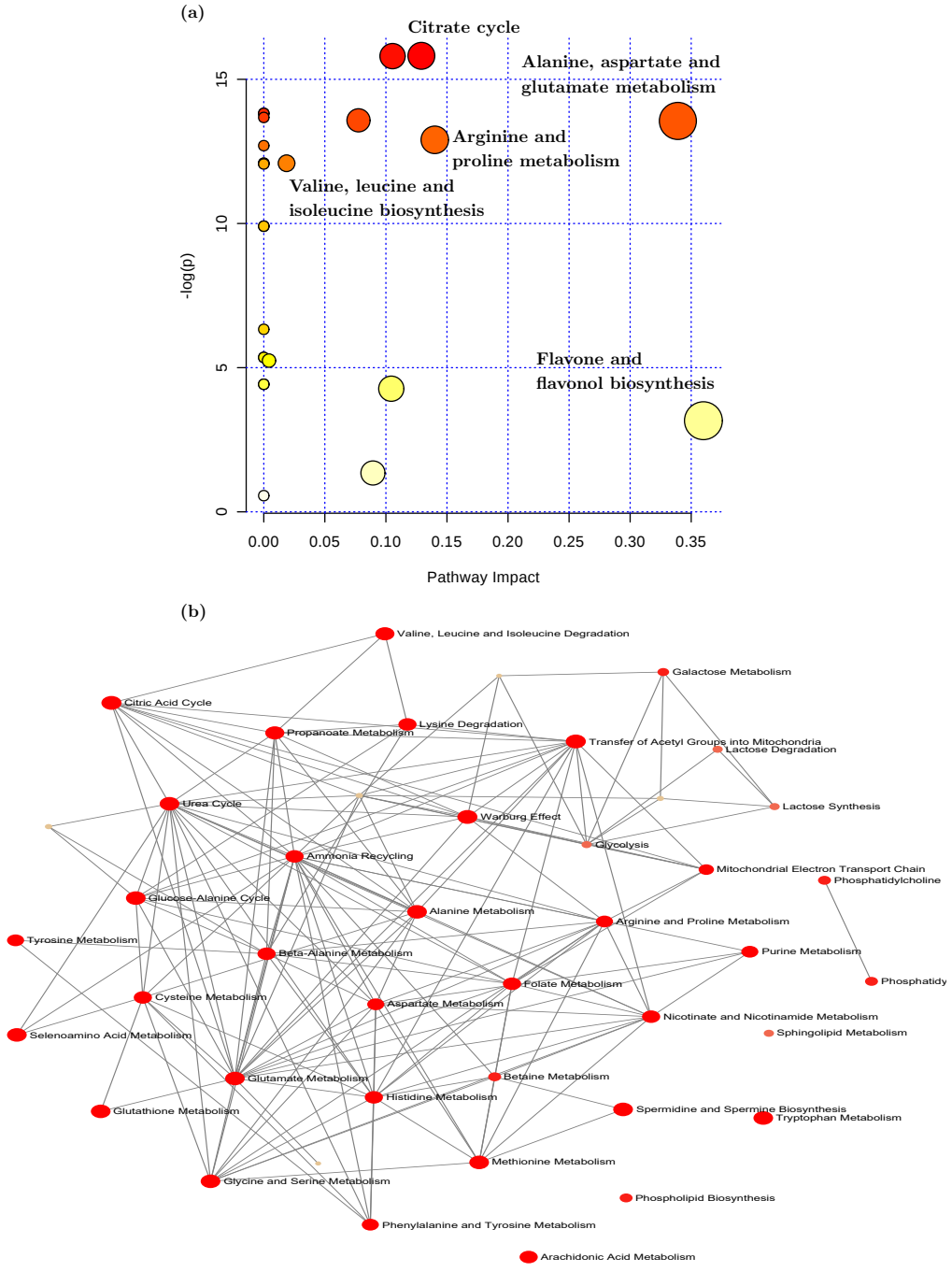

Figure 5.13: (a) Pathway topology analysis associated with exposure to air pollution stress, carried out using MetPA. (b) Correlation network with metabolites as nodes and correlations as edges. Each node indicates a metabolite set with its color intensity based on its p value. Two metabolite sets are linked by an edge which indicates the number of shared metabolites.

5. NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to air pollution stress

Oxidative stress leads to production of reactive oxygen species (ROS) in plant cells, which are associated with age-related diseases and environmental pollution. Plant defense against free radicals includes the increased production of nonenzymatic antioxidants such as glutathione, thioredoxin, ascorbate, beta-carotene and α-tocopherol as well as amino acids such as tryptophan, tyrosine, histidine, alanine, and proline [267]. We observed a significant elevation in the levels of alanine in the polluted leaf samples, indicating an increase in free radical scavenging activity. Exposure to ultraviolet (UV) light can cause DNA, RNA and protein damage and increase the production of free radicals [268]. The major response of A. thaliana to UV stress is characterized by alteration of the phenylpropanoid metabolic pathway, and in our study we also found evidence of alteration in levels of associated phenolic compounds such as chlorogenic acid, malic acid, citric acid, quinic acid, and fumaric acid, indicating that vehicular emission can produce metabolic response similar to UV exposure. Trigonelline levels get elevated as a plant response to biotic and abiotic stress [269]. We also observed a significant increase in trigonelline levels in the polluted leaf samples as compared to the control samples. Polyamines such as spermidine, spermine and putrescine are known to play the role of stress messengers in activating plant response to different kinds of environmental stress and pathogen attack [270]. Putrescine has been applied exogenously to help trigger the plant defense mechanism against a variety of environmental stresses [271]. We found polluted leaves to have enhanced levels of putrescine, suggesting that the plant maintains high levels of this polyamine in order to cope with the continuous environmental stress due to vehicular pollution. Tropospheric ozone is considered a major stress condition for plants, which arises due to heavy primary air pollution due to vehicle emissions, burning of fossil fuels and industrial emissions [272]. A major metabolic pathway which gets immediately modified in response to ozone stress is the phenylpropanoid biosynthesis pathway and a variation in levels of coumaric acid, ferulic acid, gallic acid, catechin, kaempferol and chlorogenic acid [270]. We also observed significant differences in levels of phenolic compounds in the polluted leaves as compared to the control leaves. It has been noted that in response to drought stress several metabolites get accumulated including GABA, raffinose and other oligosaccharides, putrescine and TCA cycle metabolites [273]. The citric acid (TCA) cycle is a major energy generator for an organism and dysregulation in this cycle will produce an energy deficit. TCA intermediates are malate, succinate and fumaric acid and if TCA cycle is disturbed, other means of ATP production will have to get enhanced. The levels of GABA, putrescine and TCA cycle intermediates such as fumaric acid were significantly elevated in the polluted samples in our study, indicating the plant response to lack of moisture and hence dehydrating air conditions due to vehicular pollution. Further, plant response to high and low temperatures has also been previously studied and it was found that a global reprogramming of plant metabolism occurred, with metabolic

pathways involving proline, trehalose, ascorbate, putrescine, citrulline and TCA cycle intermediates being severely affected [274]. The central carbohydrate metabolism is also strongly affected as evidenced by fluctuations in levels of sugars glucose, fructose and sucrose. In our study, we also observed an increase in putrescine, fumaric acid, glucose and fructose and a decrease in sucrose levels, indicating that vehicular pollution triggers a response similar to that of heat shock stress. Since exposure to air pollution involves a combination of stress factors, we analyzed the available published literature for metabolic profiling of A. thaliana exposed to a variety of stress factors, in order to gain a complete overall picture. In general, previous studies on A. thaliana showed that most metabolites are accumulated in excess as a first response to abiotic stress in order to support growth recovery, which is also the case for our results on Bougainvillea spectabilis. The metabolite GABA is also rapidly triggered in response to biotic and abiotic stress conditions and accumulation of GABA has been suggested as protecting the plant against oxidative stress [275]. Similarly branched amino acids such as valine, leucine and isoleucine also tend to get accumulated under abiotic stress and could function as alternative electron donors for the mitochondrial electron transport chain [276]. In our study the polluted leaf samples showed a significant increase in the levels of GABA, valine and leucine. Rigorous statistical analysis of our data showed that Bougainvillea spectabilis shows a varied response to environmental stress induced by exposure to heavy and prolonged vehicular traffic emissions. In summary, our study highlights the role of altered plant metabolism, expressed as changes in the concentrations of metabolites belonging to different yet interconnecting metabolic pathways in coping with abiotic stress. Fig 5.14 shows metabolites and their associated metabolic pathways which get altered in response to abiotic stress due to air pollution. Our study strengthens the fact that plants modulate the activities of metabolic pathways so as to maintain optimal fitness for their activities such as defense, reproduction etc even under stressful conditions. The carbohydrate and energy yielding pathways such as TCA, identified to undergo alterations in our study, have previously been reported to indicate the physiological state of the plants, as plants subjected to air pollution are known to have increased respiration and decreased CO_2 fixation due to chlorophyll deterioration [277]. A decrease in most of the metabolites belonging to such pathways in our study implicates the lower physiological health of plants in polluted areas. Further, metabolites such as polyamines (putrescine), certain amino acids, GABA and choline are known to have osmoprotectant functions which are known to restore plant cellular redox metabolism, preserve cellular turgor by maintaining osmotic balance and stabilize proteins and cellular structures during environmental stresses [278]. A change in such metabolic pathways synthesizing osmolytes highlight plants ability to maintain osmotic balance in response to air pollution stress. Finally, plants are known to produce secondary metabolites such as phenolics in response to various stresses [278].

Gaseous pollutants like SO_2, NO_2 and other toxic particulate matter can affect plant cells by blocking stomatal pores and altering gaseous exchanges resulting in disturbed plant physiology [279]. Changes in metabolic pathways resulting in the increased production of secondary metabolites such as phenylpropanoids can thus be considered as adaptive measures by the plant to overcome such damaging effects as phenolics are known to help plants by cycling nutrients and gases and being stored at strategically significant places in plants where they play a role in signaling and defense [279]. Our study thus confirms that the altered metabolic pathways in polluted leaves indeed help the plants to survive severe abiotic stress conditions.

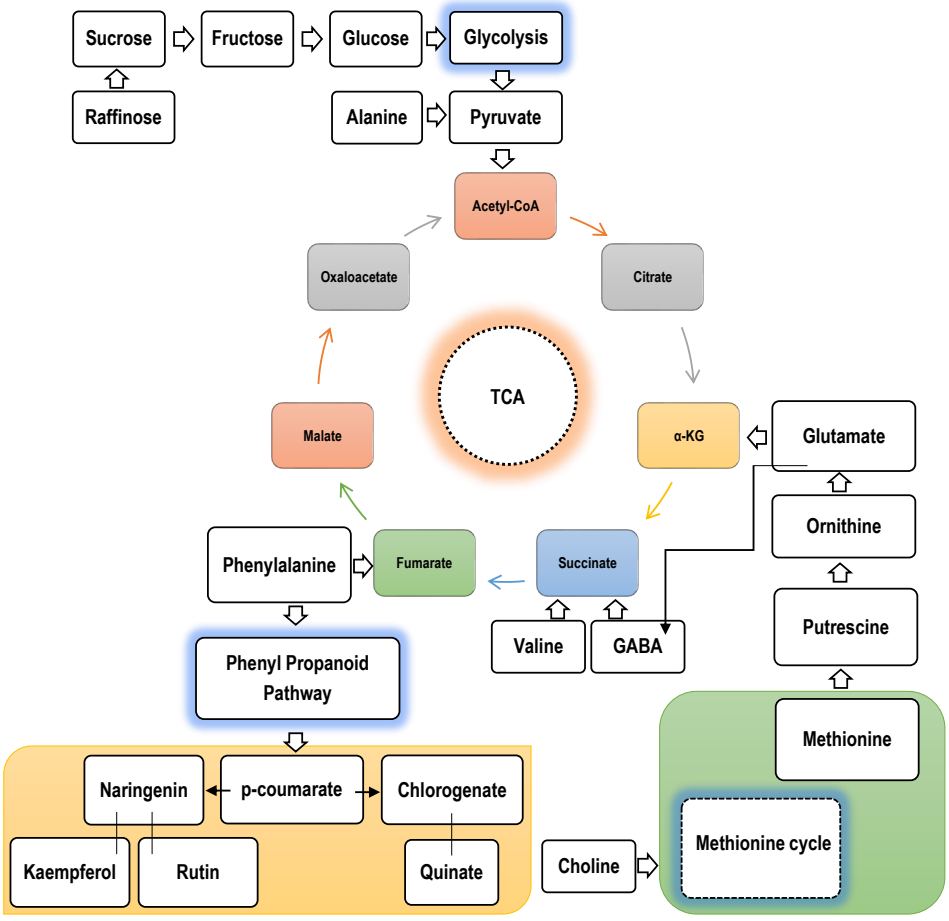

Figure 5.14: Diagram of the significant metabolites and their associated metabolic pathways which get altered due to abiotic stress of air pollution in our study.

5.4 Conclusions

Plants respond adaptively to environmental stresses by biochemically adjusting the disrupted metabolic pathways: in the short-term by an effort to ameliorate the stress and in the long-term by establishing a new steady state to maintain and regulate their metabolism. Metabolomics studies are useful in understanding plant stress response and in evolving strategies for bioengineering of plants to enable them to withstand different types of stress. NMR metabolomics is emerging as a powerful tool to gain a comprehensive picture of how metabolic networks in plants are reprogrammed in response to various types of abiotic stress conditions or even a combination of these stresses. Our study focused on establishing the identity and quantity of metabolites as the function of a rhythmic pattern during the circadian cycle, and then using this information to elucidate the difference in metabolite patterns associated with plants subjected to heavy vehicular emissions. Our work is a step forward in the direction of understanding the molecular mechanisms underlying plant tolerance to abiotic stress and has implications for studies on understanding plant response to climate change. More detailed studies vis-a-vis metabolic cycling testing in plants from polluted regions are required in order to investigate the relationship between extreme air pollution and alteration of metabolic cycling rhythms in plants. This work has been published in **Environmental and Experimental Botany, 164, 58-70, (2019)** (https://doi.org/10.1016/j.envexpbot.2019.04.019.)

5. NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to air pollution stress

Chapter 6

NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to wounding stress

6.1 Introduction

Plants may not possess the same senses that humans have, but they are rather adept at constructing an image of the environment that surrounds them. They are even able to exert some control over this, in addition to being able to communicate with one another. There are a few different ways that plants can react when they are attacked by herbivores. Many studies reported the wounding [54], [280], [281], [282], [283], [284], [285], [286], [287], [288], herbivore attack [289], [290], [291], [283], plant-insect interaction [292], [293], [294], [295], plant-herbivore interaction [296], [297], [298] and defense mechanisms of plants [299], [300], [301]. The subsequent transcriptome and metabolic alterations give rise to resistance or tolerance to attack by herbivores or pathogens, which enables plants to maintain a high level of fitness despite the presence of predators. When leaves are injured, the metabolic activities that occur in those leaves may not only protect those leaves but also regenerate the affected tissue, such as by suberizing the wound [302]. The identification of individual metabolites or metabolite families has been the primary focus of most research on the metabolic changes and processes that occur in plants in response to herbivore attack. How can plants exchange messages and interact with one another while they are being attacked by herbivores? When herbivores attack, what happens to their morphological, biochemical, and molecular mechanisms? What happens when plants are wounded? These are some of the questions that arise. Unfortunately for plants, their predators

may attack from any direction. The presence of potential threats, such as animals and insects that are feeding nearby, is communicated between plants. However, plants are not defenseless. When a plant detects that it is being attacked, it releases a toxic chemical that alters its flavor and repels herbivores. These chemicals may provide a bitter taste, impede the digestive enzymes of herbivores, disturb their metabolism, or harm them. Furthermore, when plants are attacked by insects, they may react swiftly, often in only a few minutes or hours. Plants communicate with other organisms as well, such as certain fungi and bacteria. When it comes to metabolomic methods, NMR spectroscopy and data mining make it possible to identify a wide range of metabolites and to figure out which metabolites contribute to the response of the plant to a herbivore attack [141], [303], [304]. Studies of metabolic changes in plant leaves after a wounding treatment mimicking herbivore attack can be conducted using NMR metabolomic methods. We examined the metabolic profile of damaged Bougainvillea spectabilis plant leaves. We identified the primary and secondary metabolites and their alterations in response to mechanical wounding of the plants. We identified sugars, amino acids, secondary metabolites which are associated with photosynthetic efficiency, carbon absorption, transamination, and membrane production following plant injury.

6.2 Experimental Methods

6.2.1 Study site, Plant cultivation & Sample collection

IISER Mohali campus site (30.663611 ^{0}N, 76.7275 ^{0}E) was selected for the study. Bougainvillea spectabilis plants leaves were used for the study; these plants were located at the dividers in IISER Mohali. Leaves from the Bougainvillea spectabilis plants were collected from the site and classified into two categories: Control (C: before cutting; plants that were not wounded in any way), and Wounded (W: after cutting; plant leaves that were wounded; mechanical wounding with scissor). The samples were collected in September 2021. These plants growing at the divider show the following features: larger leaves (1.5-2.0 cm length & 1.2-1.6 cm width), darker green color, and a thicker stem. The average temperature at the time of sample collection was between 30-35 ^{0}C. The plant leaves were collected at the time of day & night, freeze-dried with liquid nitrogen, and kept in storage at -80 ^{0}C until needed. The control samples were collected during 15^{th}-16^{th} of August 2021. Sixty-five control leaf samples were made for the thirteen time points that were evenly spaced (each time point has five replicates). The leaves of the plant were initially subjected to a mechanical wounding treatment. The one time wounding began at 6:50 A.M. in the morning on the 17^{th} of September 2021. The first time point for collecting wound samples was at 8:00 A.M. in

the morning on the 17^{th} of September 2021. The all time points for sample collection are as follows: 08:00 A.M., 12:00 P.M., 04:00 P.M., 08:00 P.M. (Day-1), 12:00 A.M., 04:00 A.M., 06:00 A.M., 12:00 P.M., 06:00 P.M. (Day-2), and 12:00 A.M., 10:00 A.M (Day-3). In each and every instance, the samples were immediately frozen in liquid nitrogen. The mechanical wounding treatment of Bougainvillea spectabilis plant leaves simulating the herbivore attack is shown in Fig 6.1, Fig 6.2 and Fig 6.3.

Figure 6.1: Image of a Bougainvillea spectabilis plant after wounding treatment (mechanical wounding with scissor).

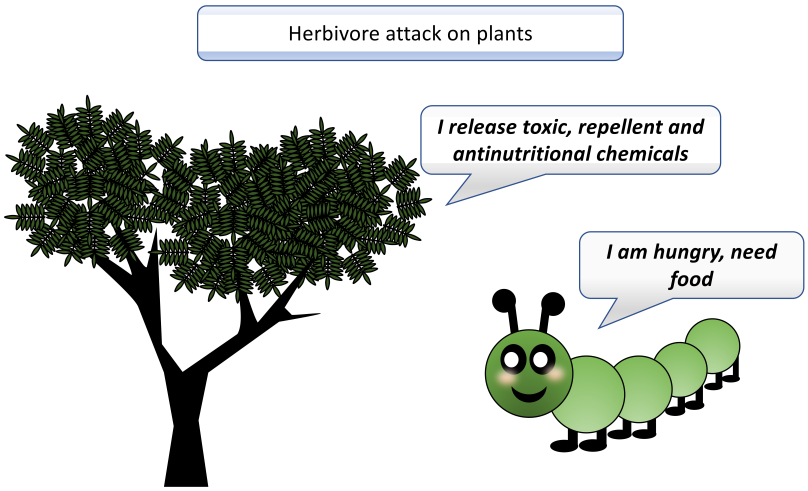

Figure 6.2: Herbivore attack on plant: Plants release toxic, repellent and anti-nutritional chemicals.

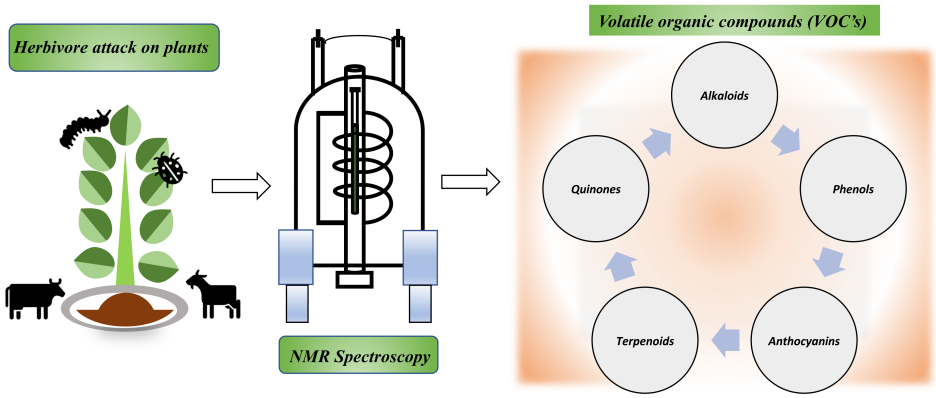

Figure 6.3: Schematic diagram: Using NMR spectroscopy, herbivore attack can be simulated by mechanically injuring plants.

6.2.2 Standards & chemicals

The following items were procured from Sigma-Aldrich (India): CH_3OH-d4 (99.9%), D_2O (99.9%), $CDCl_3$, sodium azide (NaN_3, 99.5%), phosphate buffer solution (2.0 mM, pH 7.4) and trimethylsilyl propionic acid sodium salt-d4 (TMSP-d4). All other compounds were of reagent grade.

6.2.3 Extraction & sample preparation

The polar extraction of the Bougainvillea spectabilis plant's leaves was carried out in accordance with the procedure reported in the papers [178],[305]. In summary, before lyophilization, the frozen leaves were ground in a pestle and mortar chilled with liquid nitrogen and stored at -80 ^{0}C. Freeze-dried sample weighing 30 mg was poured into a 2 mL Eppendorf tube. The extraction solvent was made by combining methanol-D4 (600 μL), D_2O (200 μL), Na_2HPO_4 (0.95 mg), sodium azide (NaN_3, 99.5%) in phosphate buffer solution (2.0 mM, pH 7.4, made in D_2O) with TMSP powder (1.5 mg/mL), which was utilized as an internal NMR reference [306]. For metabolite extraction, 800 μL solvent was added to the leaf samples and allowed to stand for 30 minutes. After that, the samples were centrifuged for 15 minutes at 5000 rcf, and then 500 μL of the supernatant was transferred to a 5 mm NMR tube before the NMR studies were carried out. For non-polar metabolites extraction (30 mg of each sample was combined with 400 μL methanol-d4 + 400 μL D_2O + 800 μL $CDCl_3$ in the ratio of 1:1:2) [306]. After the centrifugation process, we got the two layers in the eppendorf tube. The upper and lower layers in the tube contained the polar and non-polar metabolites. The 600 μL upper part of the eppendorf tube was carefully removed using the micro-pipette and

500 μL of the lower part was used for the analysis. Solvents CH_3OH-d4 and $CDCl_3$ were used as the internal lock.

6.2.4 NMR spectroscopy

^{1}H and 2D NMR spectra were acquired at a temperature of 25 degrees Celsius using a Bruker Biospin Avance-III 600 spectrometer (Fällenden, Switzerland) at a frequency of 600.17 MHz and equipped with a 5 mm QXI quadrupolar resonance pulse field gradient rf probe. 1D ^{1}H NMR spectra were acquired with a Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequence with incorporated water suppression and optimized with a spin-echo delay time t = 300 μs and loop counter n = 400, and a total spin-spin relaxation delay (2nt) time of 240 ms. The proton spectra were recorded using a 90 degree pulse width of 9.95 μs, a relaxation delay of 4 s, 32 scans, 64 K data points in the free induction decay (FID), and a spectral width of 14 ppm. Before performing the Fourier transformation, the data were first given a zero fill of a factor of 2 and then the FIDs were multiplied by an exponential weighting function that was equivalent to a line broadening of 1 Hz. The obtained spectra were manually phased, baseline-corrected, and referred to the internal standard TMSP at 0.0 ppm using Bruker TopSpin 4.2.0 software. The observed NMR resonances multiplicities were labelled as follows: s = singlet; d = doublet; dd = doublet of doublets; t = triplet; and m = multiplet. 2D NMR experiments were performed, including homonuclear ^{1}H-^{1}H correlation spectroscopy (COSY), ^{1}H-^{1}H total correlation spectroscopy (TOCSY), ^{1}H-^{13}C heteronuclear coherence spectroscopy (HSQC), and J-resolved (JRES) to support the identification of metabolites. COSY and TOCSY spectra were collected with a spectral width of 11 ppm in both dimensions, a relaxation delay of 2 s, size of FID 512 in F1 dimension, and 16 number of scans. The HSQC experiments were carried out with a relaxation delay of 2 s, 4K data points, 16 number of scans, 128 t_1 increments, and a spectral width of 14 ppm in proton dimension and 170 ppm in the carbon dimension, respectively. 2D J-resolved NMR spectra were collected utilizing 16 number of scans per 32 increments of F1 and 16384 for F2 with spectral widths of 1802 Hz in F2 (chemical shift axis) and 30 Hz in F1 (spin–spin coupling constant axis).

6.2.5 Metabolite identification

In order to assign metabolites in the ^{1}H NMR spectra, we compared proton chemical shifts to literature values, compared our spectra to those of standard compounds recorded in the same solvent circumstances (own local database), and spiked the samples. The 1D and 2D NMR recorded spectra were used to identify the metabolites, and the acquired chemical shift values were compared with those found in databases

such as the Biological Magnetic Resonance Bank Data (BMRB: www.bmrb.wisc.edu), the Human Metabolome Database (HMDB: www.hmdb.ca) [307], Prime [308], and BioMagResBank (www.bmrb.wisc.edu).

6.2.6 Univariate & Multivariate statistics

The MestReNova software (version 10.0.2, Mestrelab Research, Spain) was utilized to carry out the analysis of the NMR data. Spectral intensities were scaled to TMSP and reduced to integrated areas of equal bin widths (0.01 & 0.04 ppm), corresponding to the δ 0.5-9.0 ppm range. The spectral range δ 4.57-5.12 ppm included residual water peak and was therefore excluded from the analysis. All the ^{1}H NMR data were subjected to one way ANOVA (analysis of variance) and t-test using MetaboAnalyst version 5.0 (https://www.metaboanalyst.ca/) [184] to reveal significant differences between the groups. Multivariate analysis was performed with the SIMCA version 14.1.0-2047 software (Umetrics; Umea, Sweden)) with mean-centred data scaled to univariate & pareto. The data was first analyses using the unsupervised approach of Principal component analysis (PCA) to look for clustering, groupings, outliers, and other patterns within the data. The data points in the PCA score plot that were outside the 95% confidence region of the Hotelling's ellipse were eliminated prior to the subsequent analyses utilizing the supervised approach of orthogonal projections to latent structures discriminant analysis (OPLS-DA). The data was then analyzed using OPLS-DA to optimize group differentiation and find clustering trends.

6.3 Results and Discussions

The wounded Bougainvillea spectabilis plant leaves were subjected to polar and nonpolar extraction as well as ^{1}H NMR analysis so that a metabolic profile could be established as a starting point. The polar and nonpolar metabolic profiles of wounded Bougainvillea spectabilis plant leaves are shown in Fig 6.4, Fig 6.5, Table 6.1, & Table 6.2. Signals between δ 3.2 and 5.9 ppm correspond mostly to sugars. The resonance of α-glucose was found to be δ 5.15(d,3.8) ppm. The singlet at δ 3.22(s) ppm is choline, which is a key osmolyte. Secondary metabolites quinic acid and quercitol were found to show highly intense signals in the aliphatic area between δ 2.0 and 3.65 ppm. Asparagine, one of the amino acids, was recognized by the doublet signal at δ 4.06(d) ppm. In addition, the amino acids glutamate, valine, alanine, isoleucine, and leucine signals were identified. The presence of certain organic acids such formic acid, malic acid, citric acid, and lactic acid were confirmed. The triplet at δ 1.33 (t, J=7.17 Hz) & 4.36(m) ppm typically corresponds to an N-acetyl group. Lastly, signals in the

region ranging from δ 6.5 to 7.4 ppm, which is primarily composed of aromatic chemicals, were also identified. Chains of saturated and unsaturated fatty acids were also identified.

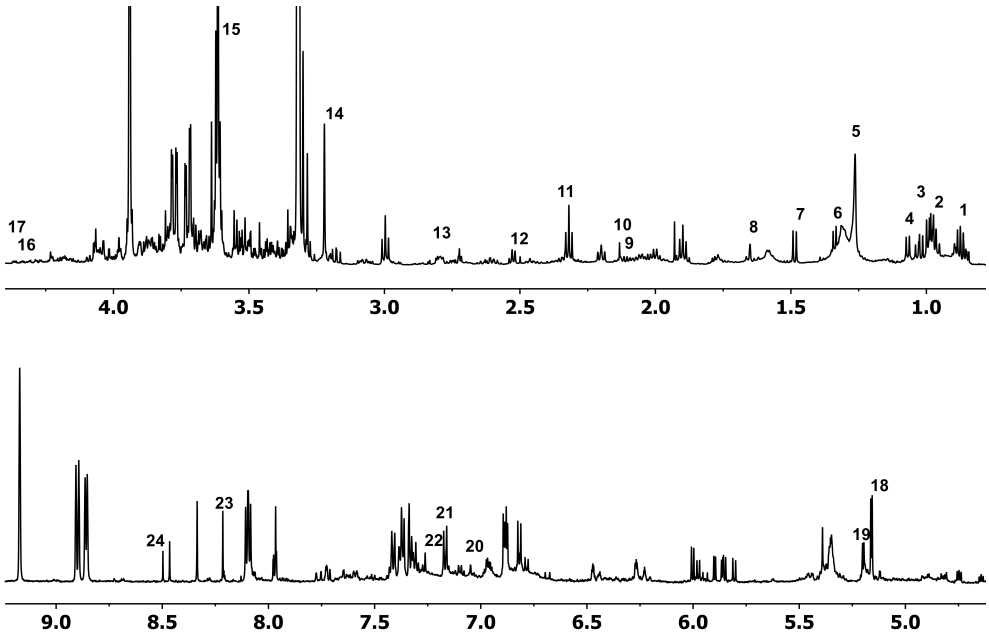

Figure 6.4: ^{1}H NMR polar metabolite profiling in wounded Bougainvillea spectabilis leaves. Peaks numbering: 1, Fatty acid; 2, Isoleucine; 3, Leucine; 4, Valine; 5, Fatty acid; 6, Lactate; 7, Alanine; 8, Fatty acid; 9, Quinic acid; 10, Glutamine; 11, Glutamate; 12, Citrate; 13, Asparagine; 14, Choline; 15, Quercitol; 16, Malate; 17, Acetyl group; 18, Alpha-glucose; 19, Beta-glucose; 20, Gallic acid; 21, Chlorogenic acid; 22, Ellagic acid; 23, NAD; 24, Formate.

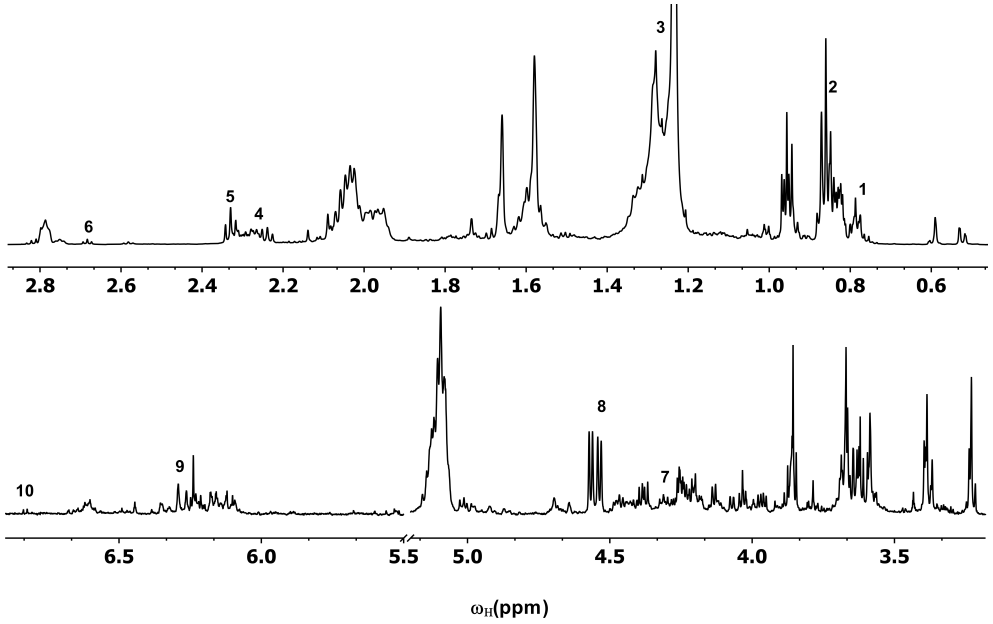

Figure 6.5: ^{1}H NMR non-polar metabolic profiling in wounded Bougainvillea spectabilis leaves. Peaks numbering: 1, fatty acid chains (-CH_3); 2, Linoleyl fatty acid chains (-CH_3); 3, fatty acid chains (-CH_2); 4, Fatty acids chains (-CO-CH_2); 5, Free fatty acids chains (-CO-CH_2); 6, polyunsaturated fatty acids chains (PUFA) (CH-CH_2-CH); 7, acetyl group (CH_3-CH_2-OR); 8, triacylglycerol (TAG) CH_2O; 9, p-coumaric acid derivative 1 PCAD-1(2); 10, PCAD-1(6/8).

Table 6.1: ^{1}H & 2D NMR assignment of the major metabolites identified in polar extracts (MeOD-D_2O) of Bougainvillea spectabilis leaves. Samples were measured at a magnetic field of 600 MHz and at a temp. of 298.0 K. Chemical shifts were referenced to TMSP.

Metabolites	Chemical shifts (m, J)
Fatty acid	0.86(t,6.95), 0.95(t,6.95)
Isoleucine	0.95(t), 1.06(d,6.98)
Leucine	1.03(d), 0.87(t)
Valine	1.06(d)
Fatty acid	1.26 (br,s)
Threonine	1.32(d), 4.24(m)
Lactate	1.33(d)
Alanine	1.48(d,7.23)
Fatty acid	1.64(m)
Quinic acid	2.13(dd)
Glutamine	2.15(m)
Glutamate	2.31(t,7), 2.15(m)
Citrate	2.52(d), 2.60(d), 2.63(d)
Asparagine	2.84(d), 4.06(d)
Choline	3.22(s)
Quercitol/quercetin	3.61(t,9.45)
Malate	4.35(m)
Acetyl group	4.36(q,7.17), 1.33(t,7.17)
Alpha-glucose	5.15(d,3.8)
Beta-glucose	5.19(d,3.8)
Gallic acid	7.06(s)
Ferulic acid	7.16(d,2)
Chlorogenic acid	7.18(d,2.1), 6.98(d,8)
Ellagic acid	7.27(s)
NAD	8.21(s)
AMP	8.21(s)
NAD	8.49(s)
AMP	8.49(s)
ATP	8.49(s)
Formate	8.49(s)

Abbreviations: s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

Table 6.2: ^{1}H & 2D NMR assignment of the metabolites identified in non-polar extracts ($CDCl_3$) of Bougainvillea spectabilis leaves. Samples were measured at a magnetic field of 600 MHz and at a temp. of 298.0 K. Chemical shifts were referenced to TMSP.

Metabolites	Chemical shifts (m, J)
Fatty acid chains; -CH_3	0.76-0.80(t)
Linoleyl fatty acid chains; -CH_3	0.84-0.87(t)
Fatty acid chains; -CH_2	1.20-1.26(br,s)
Fatty acids chains; -CO-CH_2	2.21-2.24(m)
Free fatty acids chains; -CO-CH_2	2.30-2.35(t)
PUFA chains; CH-CH_2-CH (PUFA) polyunsaturated fatty acids chains	2.66-2.70(m)
Acetyl group; CH_3-CH_2-OR;	4.28-4.33(q)
TAG-2; CH_2O (TAG) triacylglycerol	4.51-4.59
PCAD-1(2) p-coumaric acid derivative 1	6.27(d,16)
PCAD-1(6/8)	6.83(d,8)

Abbreviations: s=singlet, d=doublet, dd=doublet of doublet, t=triplet, m=multiplet, q=quartet, br=broad.

6. NMR-based metabolomic study of response of Bougainvillea spectabilis leaves to wounding stress

Leaf samples from wounded Bougainvillea spectabilis plants were analyzed at eleven time points to identify metabolites and see the variations between them. Using the unsupervised approach of PCA, the wounded samples from all eleven time points were first compared, and the results showed good grouping or clustering patterns with one replicate acting as outlier as shown in Fig 6.6 (A). To better visualize the clustering of all the time points, a supervised approach of OPLS-DA score plot was used and depicted in Fig 6.6 (B). Based on the grouping and predicted scores from the OPLS-DA scores plot for each comparison, we picked 08:00 A.M. (Day-1) and 10:00 A.M. (Day-3) for further investigation.

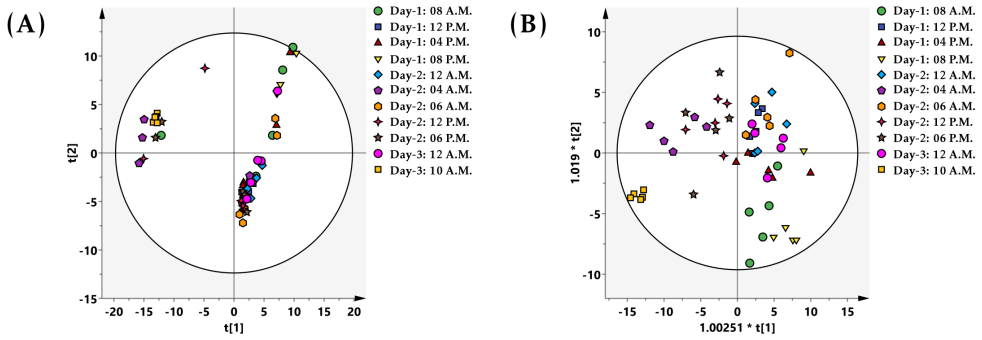

Figure 6.6: (A) PCA score plot was utilized for ^{1}H NMR in order to emphasize variations between all the time points. (B) OPLS-DA score plot derived from ^{1}H NMR spectra of wounded Bougainvillea spectabilis leaves of all eleven time points showing a clear separation.

Fig 6.7 showed the OPLS-DA score plot of the wounded samples for comparison between 08:00 A.M. (Day-1) and 10:00 A.M. (Day-3), with one predictive (T Score[1] 49%) and one orthogonal (Orthogonal T Score[1] 31.7%) component showing a clear separation between the two time points. Metabolites relative concentrations (w/w%) were derived from NMR spectra, and the results are presented in the Table 6.3 below. First, we found an increase in glucose content in wounded plants. Increases in glucose concentration have been observed as a result of a series of defensive reactions to infection and injury that involve modifications in the metabolism of carbohydrates [309]. There was also an increase in the content of C-rich secondary metabolites such as quinic acid [304], [310] and quercitol [311], [312], [313] both related to the shikimic acid pathway and linked to defense against biotic stress. The increase in choline concentration in wounded plants was due to an increase in membrane component synthesis following membrane injury [314]. The observed higher content of Asparagine is related to the higher content of choline through serine that was proved to be the precursor of choline [315].

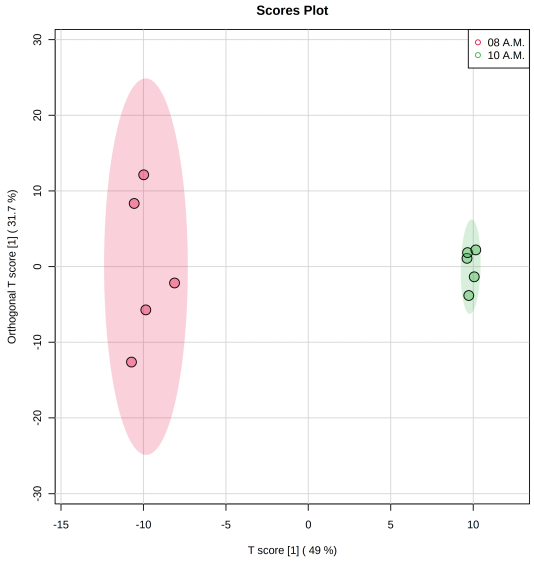

Figure 6.7: OPLS-DA score plot obtained from ^{1}H NMR spectra of wounded Bougainvillea spectabilis leaf samples at two time points 08:00 A.M. (Day-1) and 10:00 A.M. (Day-3).

Table 6.3: Comparisons of peak areas of different metabolites in wounded plants at two time points 08:00 A.M. (Day-1) & 10:00 A.M. (Day-3).

Metabolites	**08:00 A.M. (Day-1)**	**10:00 A.M. (Day-3)**
Leucine	2.91±0.09	1.91±0.01
Valine	5.23±0.16	3.85±0.04
Isoleucine	9.46±0.32	6.53±0.06
Malic acid	4.22±0.00	4.57±0.27
Asparagine	0.18±0.00	0.46±0.01
Citric acid	9.16±0.09	12.82±0.31
Choline	2.26±0.02	2.93±0.07
Quercitol	47.12±0.45	50.55±1.64
Quinic acid	2.97±0.03	3.95±0.05
Glucose	1.98±0.09	2.74±0.02

6.4 Conclusions

The metabolomic profiling of Bougainvillea spectabilis wounded leaves was investigated using 1D and 2D NMR spectroscopy in conjunction with multivariate statistical analysis. The purpose of this study was to determine whether there are significantly distinct metabolite profiles as a result of mechanical wounding stress and to identify the potential metabolites that contribute to changes in levels of metabolite concentration

at different time points. Wounded leaves of the Bougainvillea spectabilis plant show distinct changes in their metabolome with time. Plants defend themselves against herbivory attack at several levels using secondary metabolites/VOC chemicals. A rapid increase was observed in the production of quinic acid, quercitol and choline after the wounding of the plants.

Chapter 7

Studying immunity across ages in a population of Drosophila melanogaster evolved for increased immunity using NMR-based metabolomics

7.1 Introduction

Metabolomics is a cutting-edge tool for studying and exploring the physiology in Drosophila melanogaster. This method allows researchers to detect hundreds of metabolites in a single spectra which can represents the combined contributions of gene, protein and various environmental exposures (lifestyle, temperature, diet etc.) of Drosophila melanogaster. The presence of these metabolites in living organisms demonstrates how the metabolic and physiological processes of those species alter. Metabolites can be detected, identified, and quantified by using various experimental techniques like Nuclear magnetic resonance (NMR), gas or liquid chromatography (GC or LC), and mass spectroscopy. Drosophila melanogaster, often known as the fruit fly, has been the subject of a significant amount of research cited in a review article [316]. Various studies on Drosophila melanogaster cited as: the effect of heat stress [97], hypoxia [98], cold shock [99], hypoxia in the heart [100], hypoxia and age [317], temperature on inbred and outbred lines [318], acute hypoxia [319], tolerance to cold, heat, starvation, and desiccation [320], effect of cold (0^0) on lines selected for fast or slow recovery from chill coma [321], circadian regulation of metabolites [101], young and old flies expressing amyloid beta peptide in their brain [322], effects of infection with S. aureus

[323], effect of different temperatures during development [324], and long-term effects of repeated mild heat treatment [325].

A common sign of aging is immunosenescence, which is the age-related loss in immune system function [326]. Previous research on Drosophila melanogaster has attempted to examine the functional response of the immune system as it ages, but the findings have been generally inconclusive. The inherent potential to generate antimicrobial peptide genes diminishes with age, while the net activity of the immune system upon infection with live bacteria is higher in older flies [327]. The ability to withstand a bacterial infection declines with age, as shown by functional study of the immune response in Drosophila melanogaster [328]. Drosophila melanogaster have also been shown to have a decline in their phagocytic capacity with increasing age [329]. Some research, however, shows that Drosophila melanogaster immunity improves with age, with older flies (13 days old) having a lower bacterial burden than younger flies (3 days old) [330]. Studies also report a significant variation in the effect of age on immune response between lines, with some lines exhibiting an improvement in immune response with aging, some lines exhibiting no change in immune response with aging, and some lines exhibiting a decline in immune response with aging [331]. Since immunity changes with age and given that immunity and metabolism are intimately connected, it makes sense to investigate the changing metabolomes of immune-selective Drosophila melanogaster populations at different ages. This will allow one to better understand how immunity affects metabolism.

First, we studied immature flies (2 days old as adults) to determine metabolome evolution in immune-selective groups [254]. The authors examined the usefulness of NMR-based metabolomics approaches for researching the evolutionary process and to demonstrates the role performed by selection in modifying the metabolome in Drosophila melanogaster fruit flies. Metabolic variations between a fly population chosen for increased immunity and its control reveal that populations develop different metabolomes in response to selection. Additionally, the metabolic response to prick injury or infection is evaluated for immune selected groups as well as control populations, and variations in response are observed. This research extends the previously reported work [254] to account for aging-related metabolome alterations in immune-selected and control groups. This study aims to discover whether the metabolome reacts similarly in young (5-day-old), middle-aged (20-day-old) and elderly flies (35 days old as adults). In this research, we compared immune-selected populations and their controls at baseline and in response to prick damage and bacterial infection at all three ages. It is observed that even though the immune selected population ages to resemble its control population by old age, it still retains its ability to mount a better immune response to prick injury/infection as compared to its control population. This shows that both selection and aging play their individual roles in defining the ability of an organism to

mount an immune response.

7.2 Materials and Methods

7.2.1 Study protocol

The generation of the immune selected populations (I) and their controls (S) from the fly stocks were categorized under three different treatments namely: unhandled control treatment (u), injury by pricking with a needle containing a buffer (s) and infection by a pathogen (i). Thus, we had six different combinations of treated flies (Iu, Is and Ii for I population and Su, Ss and Si for S population). The flies from both the immune selected populations (I) and their controls (S) were further selected at three different ages- young flies (5 days old as adults), middle aged flies (20 days old as adults) and old flies (35 days old as adults) and labeled as Age 1, Age 2 and Age 3 respectively, used for this study. The fly stocks, selection regime used for creating the populations, the bacteria used for infections and the standardization procedures have been described previously and reported [254].

7.2.2 Metabolite extraction

The frozen samples were mechanically homogenized using a battery run homogenizer in 300 μL of ice-cold acetonitrile (50 $\%$) and centrifuged at 12,000 rpm for 10 min at 4 ^{0}C. The supernatant was then lyophilized using a freeze dryer at a pressure of 30 Pa for 16 h and stored at -80 ^{0}C until the NMR experiments were performed. For NMR spectroscopy, the samples were rehydrated in 500 μL of D_2O, containing 0.5 mg/ml of an NMR reference 3-(trimethylsilyl) propionicacid-D4 sodium salt (TMSP), and then supernatant was transferred to a 5 mm NMR tubes. All the extraction solvents and chemicals including D_2O (with 99.96 $\%$ deuterium-enriched), acetonitrile (99.8 $\%$) and TMSP (98 $\%$) were of analytical grade with purity greater than 90 $\%$ and purchased from Sigma-Aldrich Inc. (St. Louis, Missouri, United States).

7.2.3 NMR spectroscopy

All NMR spectra were acquired at 25 ^{0}C on a Bruker Biospin Avance-III 400 MHz spectrometer with a 5 mm BBO (broad-band observe) probe, operating at a ^{1}H frequency of 400.12 MHz. Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence was used for acquiring 1D ^{1}H NMR spectra in order to attenuate broad signals from high

molecular weight molecules (proteins and lipids). ^{1}H CPMG experiments were optimized for a spin echo delay time t of 300 μs (n = 400) and a total spin-spin relaxation delay (2nt) time of 240 ms. CPMG experiment provides spectra with smooth baseline and phase which simplify multivariate analyses. All experiments were collected with uniform parameters with a 90^{0} pulse width of 12.6 μs, a relaxation delay of 1 s, 16 scans, 32 K data points and a spectral width of 12 ppm. All ^{1}H NMR spectra were automatically phased and baseline-corrected with 0.30 Hz line broadening using Bruker Topspin 4.0.6 software (Fllenden, Switzerland). For assignment purposes, two-dimensional homonuclear correlation spectroscopy (^{1}H–^{1}H COSY) and total correlation spectroscopy (^{1}H-^{1}H TOCSY) NMR spectra were obtained with a spectral width of 12 ppm in both the dimensions, 4 k data points, a relaxation decay of 2 s, 16 scans and 512 t_1 increments. The 2D heteronuclear single quantum coherence spectroscopy (^{1}H-^{13}C HSQC) NMR spectra were recorded with spectral width of 12 and 200 ppm in proton and carbon dimensions respectively, 4 k data points, 32 scans, relaxation delay of 2 s and 128 t_1 increments.

7.2.4 Metabolite assignments

Using 1D and 2D NMR spectra data, metabolites were assigned by comparing their chemical shift and scalar coupling values with standard NMR metabolite databases as Human metabolome database (HMDB) https://www.hmdb.ca/, Biological Magnetic Resonance Data Bank (BMRB) http://www.bmrb.wise.edu, the Madison Metabolomics Consortium Database (MMCD) http://mmcd.nmrfam.wise.edu and published literature [99], [321], [254]. We also recorded the pure compounds to confirm the metabolites peaks. MetaboHunter [182] and MetaboMiner [183] web-server applications were also used for the identification of metabolites in ^{1}H and 2D NMR (TOCSY, HSQC) spectra. The identification of compounds is done according to recommendations by the Metabolomics Standards Initiative (MSI) [118]. Highest confidence MSI level 1 is assigned for those compounds which satisfy the two or more independent and orthogonal parameters of the standard compounds measured in the same laboratory under the same conditions. Putatively annotated compounds from databases and literatures marked as MSI level 2.

7.2.5 Data analysis

Bruker Topspin 3.5pl7 and MestReNova (version 10.0.2, Mestrelab Research, Spain) software were used for the processing of free induction decay (FID), spectral processing, the estimation of the signal-to-noise ratio, and spectral data reduction (binning). Each FID was zero-filled to 128 K data points and multiplied by an exponential

window function with 0.3 Hz line broadening. All the ^{1}H NMR spectra were then automatically zero-order phased and manually calibrated on TMSP signal (0.0 ppm) and converted to ASCII format. For data analysis, spectra were aligned by using the "icoshift" (interval correlation shifting) algorithm [119]. The spectral region from δ 0.5-9.0 ppm were normalized with respect to TMSP signal and region from δ 4.7–4.9 ppm contained residual water peak was removed in order not to compromise with the analysis. PCA (Principal component analysis), PLS-DA (Partial least square discriminant analysis) and OPLS-DA (orthogonal partial least square discriminant analysis) for classification were performed on unit-variance and Pareto-scaled data by using the SIMCA 14.1.0-2047 software (Umetrics, Umea, Sweden) and Metaboanalyst 4.0 (http://www.metaboanalyst.ca/) web-based platform [332]. The unsupervised model PCA was carried out on data to see the outlier and patterns within the data [333]. Outliers in the PCA model was identified as samples located far away from the 95 $\%$ confidence area of the Hotelling's T2 and excluded from further analysis. The validity of PCA was performed by R2X and R2Y values, indicate the percentage of the variance in X and Y matrices explained the goodness of fit, respectively, while Q2Y is the predictive capability of the model [334]. The data were further examined by supervised model PLS-DA and OPLS-DA to identify the important features between the groups [67]. The importance of each metabolite in OPLS-DA was evaluated by variable importance in the projection (VIP) score, where metabolites with VIP score > 1 were considered important. The loading plots were used to identify the significant metabolites responsible for the maximum separation in the OPLS-DA. In addition, s-line plot was used to visualize the covariance (p(ctr)[1]) and correlation coefficient (abs(p(corr)[1]) between the variables and the classification score in the model. Metabolites with (abs(p(corr)[1]) > 0.6) in the s-line plot were considered to be statistically significant. The significance test of the OPLS-DA model was performed by using CV-ANOVA (Cross-validated ANOVA) [252] in the SIMCA software. The data were analyzed by one-way ANOVA (analysis of variance) and Student's t-test for comparison between three and two groups respectively with p-value threshold of 0.05. The polar dendrogram was created using a MATLAB script (Sam Roberts 2022, Draw a Polar Dendrogram (https://www.mathworks.com/matlabcentral/fileexchange/21983-draw-a-polar-dendrogram), MATLAB Central File Exchange. Retrieved September 12, 2022.) to better visualize the clustering and patterns [101], [254].

7.3 Results and Discussions

7.3.1 Metabolite fingerprinting

Metabolite fingerprinting was done as explained in our previous study [254]. Both 1D and 2D NMR spectra were recorded, and metabolites identified by matching their chemical shift, J-coupling values obtained with those of pure compounds recorded in databases such as HMDB, BMRB and MMCD. Furthermore, these assignments were confirmed by 2D NMR spectra including J-resolved, COSY, TOCSY, HSQC and by the comparison of reference compounds. Representative ^{1}H NMR spectra of Drosophila melanogaster fly is displayed in Fig 7.1 with metabolites labeled. Table 7.1 lists all the metabolites identified using their chemical shift values and scalar coupling patterns. Aliphatic (δ 0.8–2.8 ppm), sugar and amino acids (δ 2.6–4.6 ppm), and aromatic (δ 5.5–9.5 ppm) areas were identified as the principal groups of chemicals. In brief, proton nuclear magnetic resonance spectroscopy of Drosophila melanogaster extracts led to the identification of 36 different metabolites: 1 sterol, 3 lipids, 14 amino acids, 5 organic compounds, 4 carbohydrates, 3 carboxylic acids, 1 choline and 5 other compounds. A low intense peak at δ 0.76 (s) was found as a signal of sterol. We identified a downfield-shifted terminal methyl signal at δ 0.89 ppm. The amino acids regions δ 0.93-3.60 & 3.70-7.80 ppm in the spectra showed the signals of leucine, valine, isoleucine, alanine, arginine, proline, glutamine, lysine, threonine, glutamic acid, histidine, tyrosine, phenylalanine and tryptophan. Signals found in the sugar region were attributable to sucrose at δ 5.40 (d, J = 3.8 Hz), α-glucose at 5.22 (d, J = 3.8 Hz), β-glucose at δ 4.64 (d, J = 7.9 Hz), and galactose at δ 4.07 (t, J = 5.98 Hz). Organic compounds lactic acid, citric acid, 3-hydroxykynurenine, creatine, and fumaric acid peaks were detected at δ 1.32 (d, J = 6.9 Hz), δ 2.53 (d), δ 3.69 (d, J = 5.5 Hz), δ 3.92 (s) and δ 6.51 (s) ppm respectively. The identified energy compounds were NAD, AMP and ADP in the proton spectra.

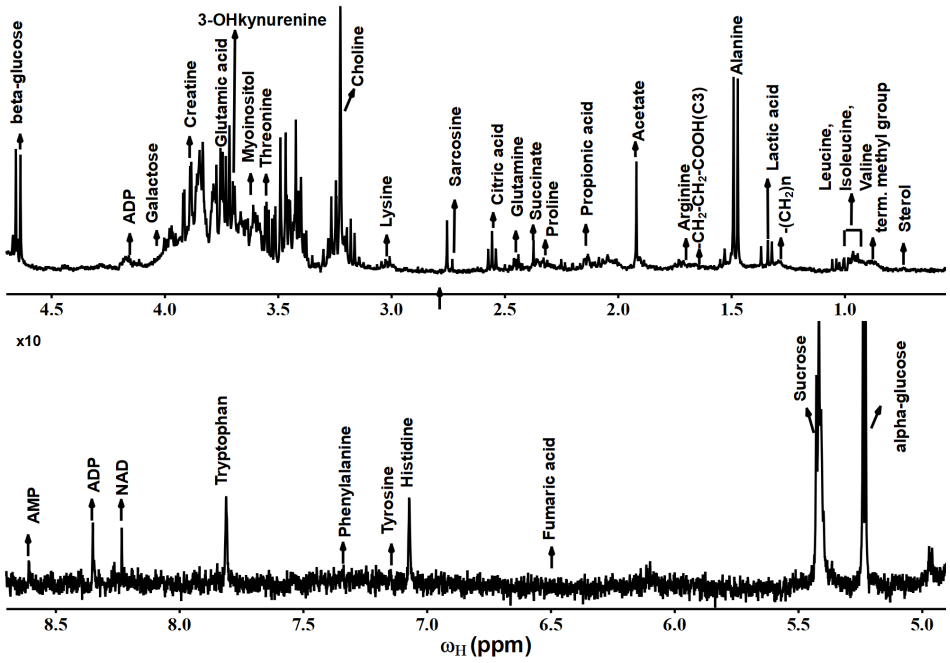

Figure 7.1: A representative one dimensional ^{1}H CPMG NMR spectra collected from Drosophila melanogaster flies extract on Bruker Biospin-Avance 400 MHz spectrometer. The signals from the assigned metabolites have been shown in the spectra by using the standard metabolite libraries and previous reported literatures.

Table 7.1: The metabolites identified from ^{1}H NMR and 2D NMR spectra of Drosophila melanogaster, with chemical shifts in ppm and the corresponding multiplicity m and scalar coupling J values (in Hz).

Metabolites	Chemical shifts (m, J)	MSI Level
Sterol	0.76 (s)	Level 2
Term. methyl group	0.89(t,6.9)	Level 2
(-CH_2)n	1.27(m)	Level 2
-$CH_2CH_2COOH(C_3)$	1.64(m)	Level 2
-CH_2COOH (C_2)	2.36(t,7.5)	Level 2
Leucine	0.94(t,5.9), 1.70(m)	Level 2
Valine	0.97(d,7), 1.02 (d,7.1), 2.26(m,14,7)	Level 1
Alanine	1.48 (d,7.2), 3.76(q,7.2)	Level 1
Glutamine	2.15(m), 2.44(m)	Level 2
Glutamic acid	2.15(m), 2.44(m), 3.75(dd,7.18,4.72)	Level 2
Phenylalanine	3.19(m), 7.36(m)	Level 1
Tyrosine	6.86(m), 7.17(m)	Level 2
Tryptophan	7.31(s), 7.72(d,8)	Level 1
Threonine	3.57(d,4.8), 1.31(d)	Level 2
Isoleucine	0.99(d,6.9), 0.93(t,7.62)	Level 2
Histidine	3.16(dd,15.5,7.7), 7.09(d,0.58)	Level 2
Lysine	3.00(t), 3.71(t,6.09)	Level 2
Arginine	1.68(m), 3.76(t,6.5)	Level 2
Proline	4.11(dd,8.56,6.4), 2.34(m)	Level 2
Sarcosine	2.73(s), 3.60(s)	Level 2
Beta-glucose	4.64(d,7.9)	Level 2
Alpha-glucose	4.63(d,7.9), 5.22(d,3.8)	Level 2
Sucrose	4.20(d,8.75), 5.4(d,3.8)	Level 1
Galactose	4.07(t,5.98), 5.26(d,3.66)	Level 2
Fumaric acid	6.51(s)	Level 1
Acetic acid	1.91(s)	Level 2
Succinic acid	2.37(s)	Level 1
Lactic acid	1.32(d,6.9), 4.1(q)	Level 2
Propionic acid	2.16 (q,7.15), 1.04(t)	Level 2
Citric acid	2.53(d), 2.65(d)	Level 2
Pyruvate	2.46(s)	Level 2
Myo-inositol	3.26(t,9.3), 4.05(t,2.8), 3.61(t,9.7)	Level 2
Choline	3.22(s)	Level 2
Creatine	3.92(s), 3.02(s)	Level 2
NAD	8.20(m), 8.40(s)	Level 2
3-hydroxykynurenine	3.69(d,5.5), 7.02(d)	Level 2
AMP	8.23(s), 8.58(s)	Level 2
ADP	8.29(s), 8.54(s)	Level 2
GABA	2.33(t,7.2), 1.90(m)	Level 1

Abbreviations: AMP (Adenosine monophosphate), ADP (Adenosine diphosphate), GABA (gamma-aminobutyric acid), s=singlet, d=doublet, dd=doublet of doublet, t=triplet, q=quartet, br=broad

7.3.2 Statistical analysis

It was established from our previous study that the immune selected (I) population evolved to be metabolically different from the control (S) population at the basal, uninfected level [254]. Our results also confirmed that both I and S populations respond in a similar fashion to both prick injury and bacterial infection. However, the changes in the metabolite levels in S population were more severe than I population, suggesting the immune (I) population to be more robust to treatments and have a better metabolic homeostasis during infections. Our previous study was performed on young flies (flies were 2-3 days old as adults). In this study, flies from three different ages were analyzed

using multivariate statistical analysis to answer two important questions: 1. How does aging impact the basal metabolome of immune selected population? And 2. How does the metabolomic response to prick injury and bacterial infection change with aging in both I and S populations? Fig 7.2 shows the experimental design, where three blocks were analyzed individually and independently. Each block had both I (immune selected) and S (control) populations, with each population undergoing three different treatments so as to have a total of six differently treated populations namely Iu, Is, Ii, Su, Ss and Si in the study.

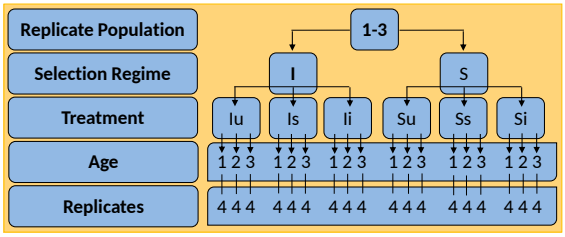

Figure 7.2: Schematic experimental design for the NMR metabolomics experiments. Three independent blocks, each having two regimes; immune selected populations (I) and their controls (S) were subjected to three treatments mainly: unhandled control treatment (u), injury by pricking with a needle containing a buffer (s) and infection by a pathogen (i). Three ages were selected per treatment and each treatment had 4 replicates.

Further, flies from all the six treatments were analyzed at three different ages – Age 1 (young flies; 5 days old as adults), Age 2 (middle-ages flies; 20 days old as adults) and Age 3 (old flies; 35 days old as adults). An initial analysis of all six treatments from both I and S regimes, at three different ages, was done using the unsupervised statistical method of PCA in order to identify the patterns such as groupings and outliers shown in Fig 7.3. A supervised PLS-DA method (Partial least squares-discriminant analysis) was also used to identify metabolic differences between all the groups in the present study depicted in Fig 7.4. Each age was then analyzed individually using the supervised pattern recognition method of OPLS-DA. Fig 7.5 (a-c) shows cross-validated OPLS-DA scores plot, for the three ages, showing that irrespective of the type of treatment given, both I and S selection regimes continue to remain inherently different from each other throughout their life span.

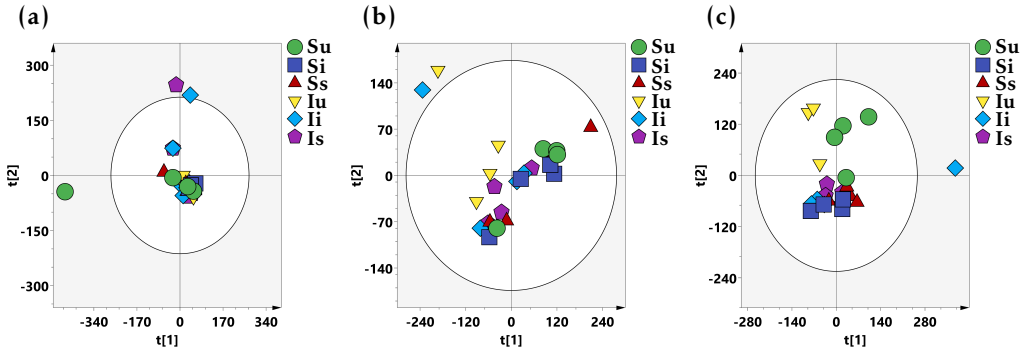

Figure 7.3: PCA score plot on ^{1}H NMR spectra of all the six treatments with three different ages (a) Age-1 (PC1=29.3%, PC2=20.2%), (b) Age-2 (PC1=32.3%, PC2=17.4%) and (c) Age-3 (PC1=43.4%, PC2=29.3%). The PCA score plot also showing the outliers which is located outside the 95% confidence area of the Hotelling's T2 ellipse.

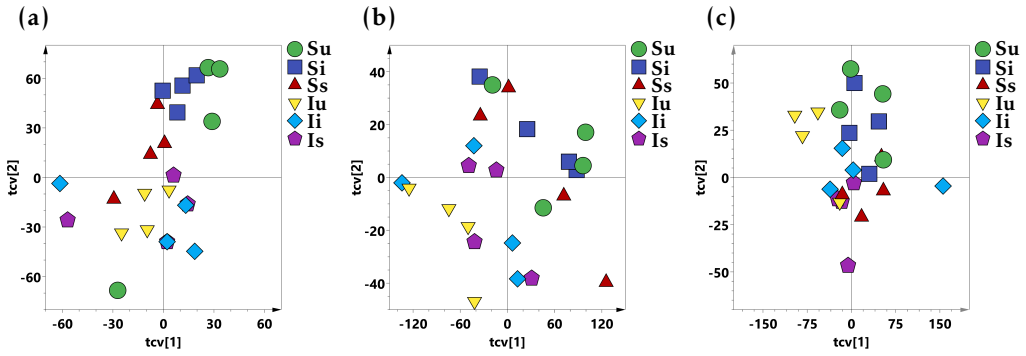

Figure 7.4: PLS-DA (Partial least squares-discriminant analysis) score plot on ^{1}H NMR spectra of all the six treatments with three different ages (a) Age-1 (Component 1=21.9%, Component 2=17.5%), (b) Age-2 (Component 1=26.9%, Component 2=15.5%) and (c) Age-3 (Component 1=26.4%, Component 2=15%).

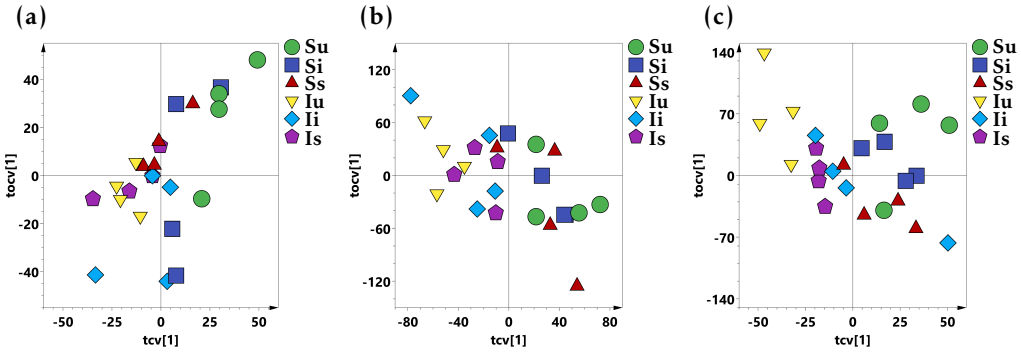

Figure 7.5: Cross validated orthogonal partial least-squares-discriminant analysis (CV OPLS-DA) on ^{1}H NMR spectra of all the six treatments with three different ages (a) Age-1, (b) Age-2 and (c) Age-3.

The impact of aging was confirmed from a polar dendrogram as shown in Figure 3, plotted for all six treatments at three ages. Fig 7.6 shows a clear separation of Age 1 and 2 from Age 3 by a main node, while a further node also separates both Ages 1 and 2. Further smaller nodes are shown to separate treatments within individual ages. This confirms that the metabolome of both I and S regimes does change with aging and further at any particular age, the metabolome of the treated populations is different from the untreated ones.

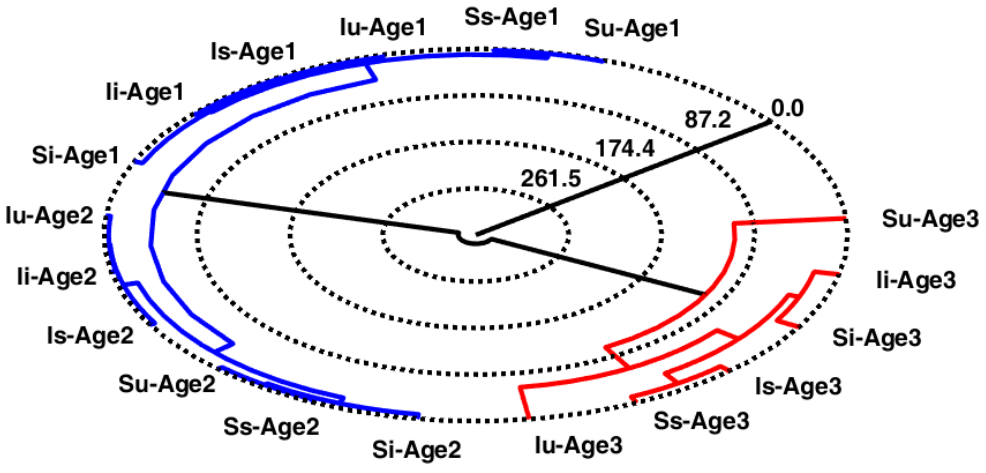

Figure 7.6: Polar dendrogram showing the differences between all the treatments (I: Iu, Is, Ii and S: Su, Ss, Si) with three different ages.

Does aging have an impact on the metabolome of I and S populations? We

then moved to answer our first question of whether I population, which had evolved to become metabolically different from the S population at a young age, continues to remain so throughout the lifespan of the fly? To answer this, we compared the unhandled control fly treatments from both I and S regimes (Iu and Su) at all three ages independently. Fig 7.7 (a-c) shows cross-validated OPLS-DA scores plot for comparison between Su and Iu treatments at the three ages. As can be seen from the figure, both Su and Iu are clearly separated from each other at all three ages. Table 7.2 lists the variance explained R2 and the variance predicted Q2 by the model, along with the predicted and orthogonal scores explaining the percentage variance observed between the two groups. As can be seen from the scores, there is a decrease in variation between Su and Iu populations with aging. This shows that even though the basal metabolome of the immune selected population (Iu) continues to remain different from the control population (Su) at all ages, the differences between them decrease with time. The variance explained R2 and the variance predicted Q2 confirm the model to be an effective one, with good predictive accuracy. Testing with CV-ANOVA with α value < 0.05 and validating with a permutation test (p-value < 0.05) also showed the model to be statistically significant and robust in nature. The significant metabolites were identified and confirmed by first preselecting them using the VIP score parameter (VIP > 1), followed by identifying the metabolites from the linear S-line plot (abs(p(corr) [1]) $>$ 0.6) as shown in Fig 7.8 (a-c).

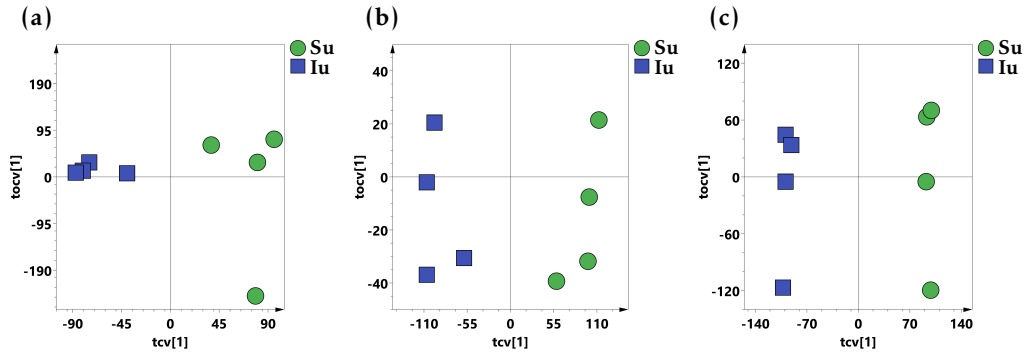

Figure 7.7: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Su and Iu populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

Table 7.2: Summary of parameters for assessing model quality: Su-Iu Age-1, 2, 3

Comparison	No. of components	R2X (cum)	R2Y (cum)	Q2 (cum)	T score [1] %	Orthogonal T score [1] %
Su-Iu (Age-1)	1+3	0.762	1	0.902	56.1	16.2
Su-Iu (Age-2)	1+4	0.818	1	0.951	49.7	19.2
Su-Iu (Age-3)	1+4	0.924	1	1	30.1	50.1

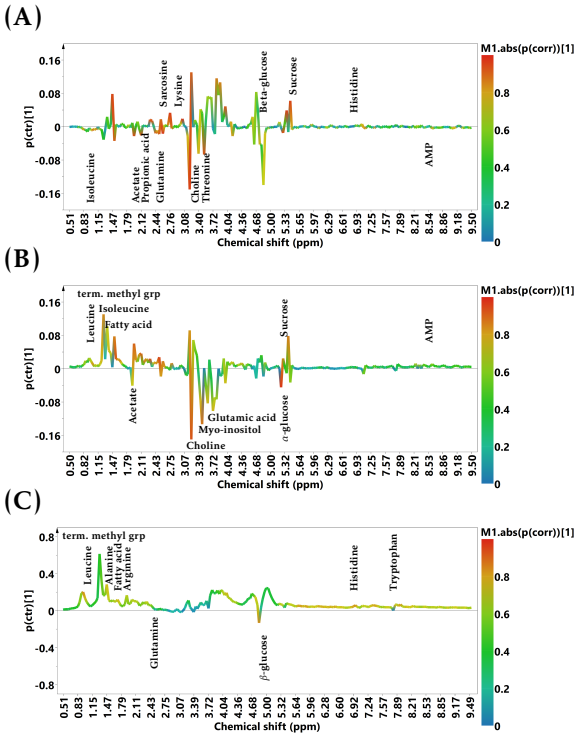

Figure 7.8: Loading S-line plot from OPLS-DA showing ^{1}H NMR metabolomics data for comparison between Su and Iu populations in Drosophila melanogaster at the three Ages: (A) Age-1; (B) Age-2; (C) Age-3. Variations of bucket intensities are represented using a line plot from 0.5 to 9.5 ppm and color of the line indicates the significance of correlation.

Similarly, a clear separation observed between Ss and Is across all the ages. Fig 7.9 (a-c) shows cross-validated OPLS-DA scores plot for comparison between Ss and Is treatments at the three ages. Table 7.3 lists the variance explained R2 and the variance predicted Q2 by the model, along with the predicted and orthogonal scores explaining the percentage variance observed between the two groups. The significant metabolites were identified and confirmed by first preselecting them using the VIP score parameter (VIP $>$ 1), followed by identifying the metabolites from the linear S-line plot (abs(p(corr) [1]) $>$ 0.6) as shown in Fig 7.10(a-c).

Table 7.3: Summary of parameters for assessing model quality: Ss-Is Age-1, 2, 3

Comparison	No. of components	R2X (cum)	R2Y (cum)	Q2 (cum)	T score [1] %	Orthogonal T score [1] %
Ss-Is (Age-1)	1+4	0.798	1	0.728	23.2	37
Ss-Is (Age-2)	1+4	0.808	1	0.759	35.5	27.3
Ss-Is (Age-3)	1+4	0.813	1	0.81	16.4	50.8

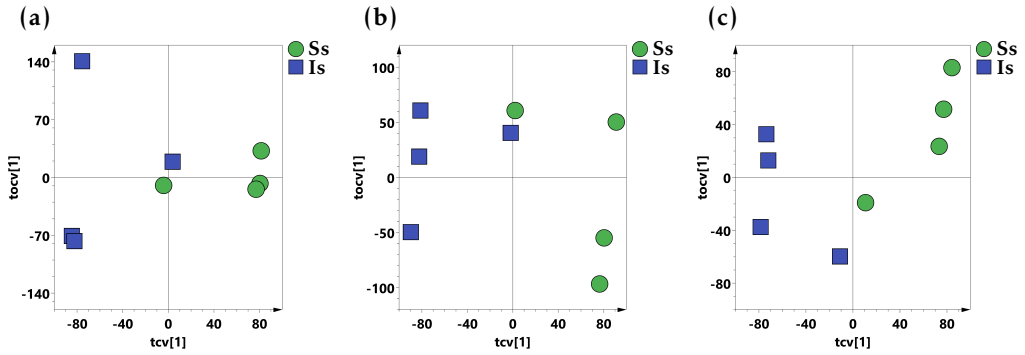

Figure 7.9: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Ss and Is populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

Table 7.4: Summary of parameters for assessing model quality: Si-Ii Age-1, 2, 3

Comparison	No. of components	R2X (cum)	R2Y (cum)	Q2 (cum)	T score [1] %	Orthogonal T score [1] %
Si-Ii (Age-1)	1+4	0.809	1	0.863	19.3	37.4
Si-Ii (Age-2)	1+4	0.821	1	0.858	24.4	33.6
Si-Ii (Age-3)	1+1	0.618	0.983	0.466	42.2	29.9

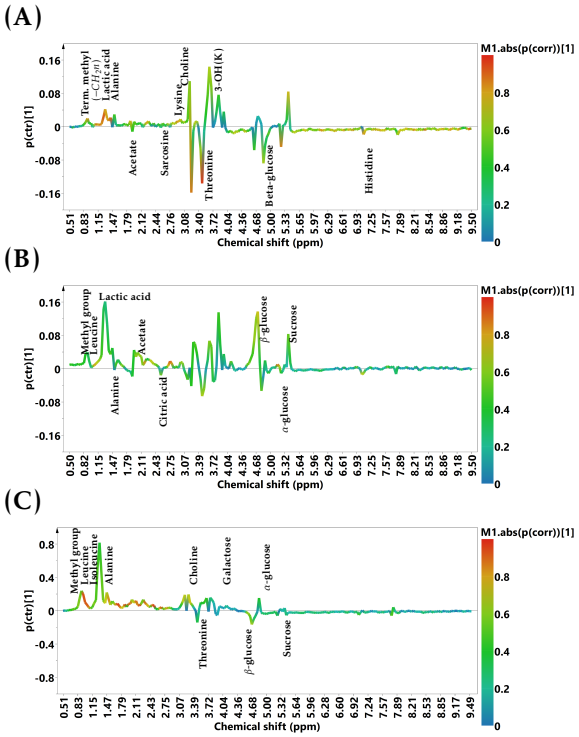

Figure 7.10: Loading S-line plot from OPLS-DA showing ^{1}H NMR metabolomics data for comparison between Ss and Is populations in Drosophila melanogaster at the three Ages: (A) Age-1; (B) Age-2; (C) Age-3. Variations of bucket intensities are represented using a line plot from 0.5 to 9.5 ppm and color of the line indicates the significance of correlation.

Similarly, a clear separation observed between Si and Ii across all the ages. Fig 7.11 (a-c) shows cross-validated OPLS-DA scores plot for comparison between Si and Ii treatments at the three ages. Table 7.4 lists the variance explained R2 and the variance predicted Q2 by the model, along with the predicted and orthogonal scores explaining the percentage variance observed between the two groups.

The significant metabolites were identified and confirmed by first preselecting them using the VIP score parameter (VIP $>$ 1), followed by identifying the metabolites from

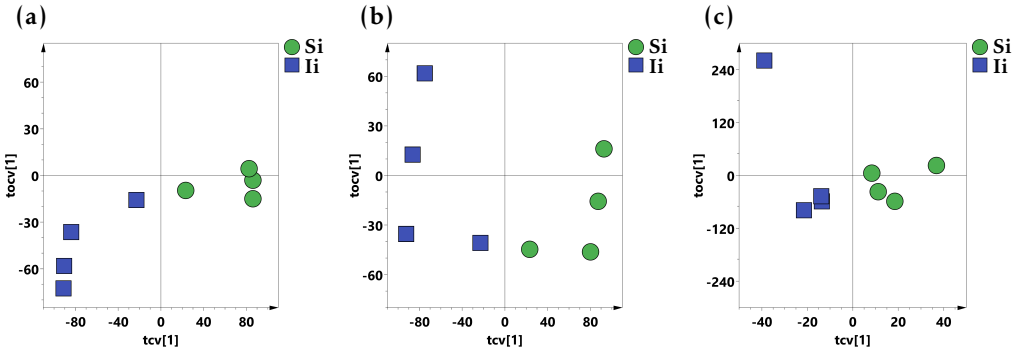

Figure 7.11: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Si and Ii populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

the linear S-line plot (abs(p(corr) [1]) $>$ 0.6) as shown in Fig 7.12 (a-c).

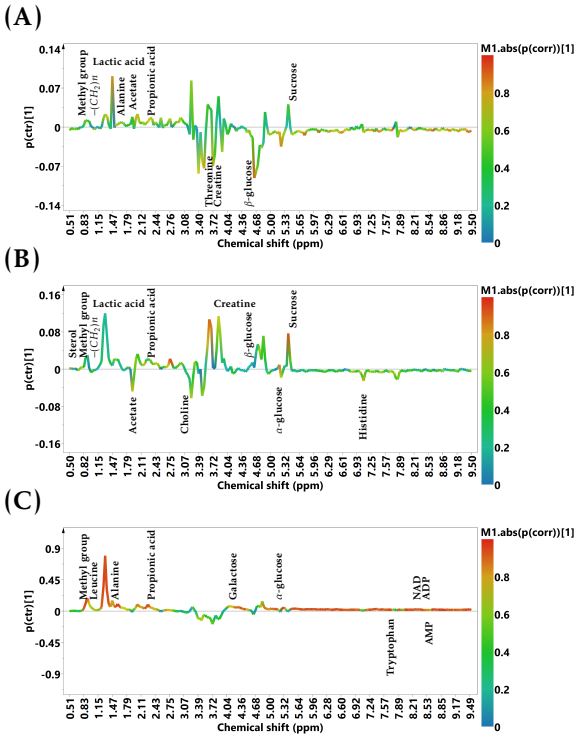

Figure 7.12: Loading S-line plot from OPLS-DA showing ^{1}H NMR metabolomics data for comparison between Si and Ii populations in Drosophila melanogaster at the three Ages: (A) Age-1; (B) Age-2; (C) Age-3. Variations of bucket intensities are represented using a line plot from 0.5 to 9.5 ppm and color of the line indicates the significance of correlation.

How does the metabolomic response to prick injury and bacterial infection change with aging in both I and S populations? To answer our next question, we performed OPLS-DA comparison between prick injury treated population Is & its control Iu; Ss population & its control Su, at all three ages. Similarly, to identify significant metabolic changes due to bacterial infection, we compared Ii with Iu population and Si with Su population at all three ages. Fig 7.17 (a,b,c) shows cross-validated OLPS-DA scores plot for Iu-Is comparison and Fig 7.18 (a,b,c) shows cross-validated OPLS-DA scores plot for Su-Ss comparison. Similarly, Fig 7.19 (a,b,c) and Fig 7.20 (a,b,c) shows cross-validated OPLS-DA scores plot for Iu-Ii and Su-Si comparisons respectively. S-plots for the above comparisons have been included in the Fig 7.17, Fig 7.18, Fig 7.19 and Fig 7.20. As can be seen from the predicted scores, both S and I populations undergo metabolic changes in response prick injury and bacterial infection. As can be inferred from the predicted scores, in case of both prick injury and bacterial infec-

tion, both I (Is and Ii) and S (Ss and Si) populations tend to have a greater separation from their controls (Iu and Su respectively) with aging. This suggests that there is an increased metabolic response with aging in both I and S populations in response to both prick injury and bacterial infection. All the above models were validated using CV-ANOVA and permutation test as explained before. All the metabolites identified in different comparisons were also confirmed by univariate statistical analysis of t-test. A multiple hypothesis test correction method of Benjamini-Hochberg (BH) was applied to confirm the statistical significance of the identified metabolites.

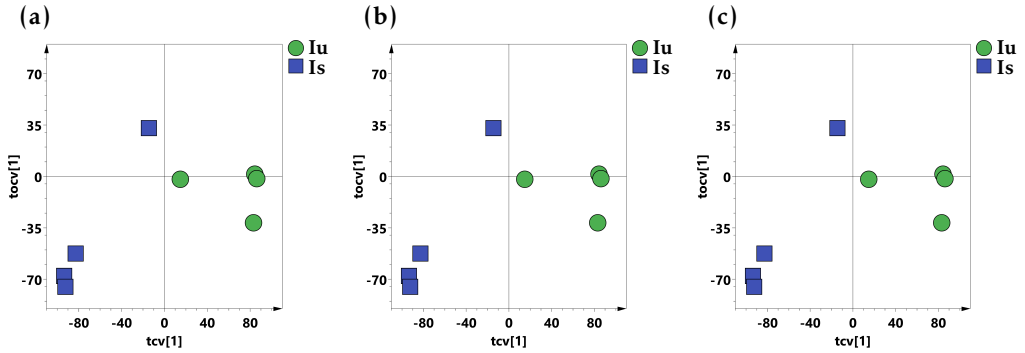

Figure 7.13: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Iu and Is populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

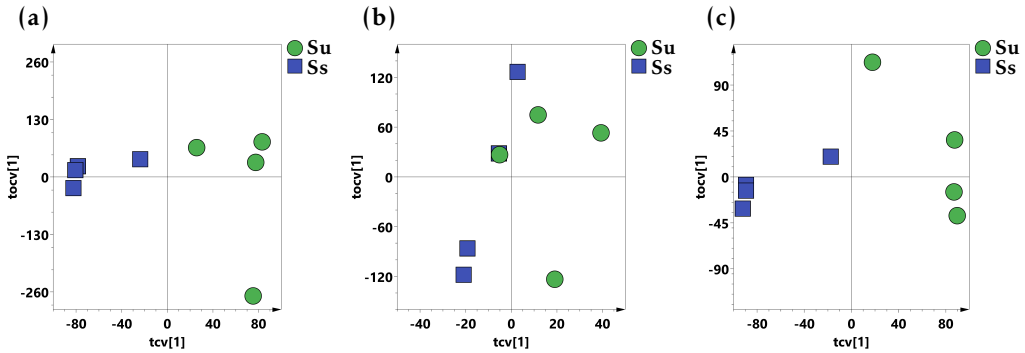

Figure 7.14: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Su and Ss populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

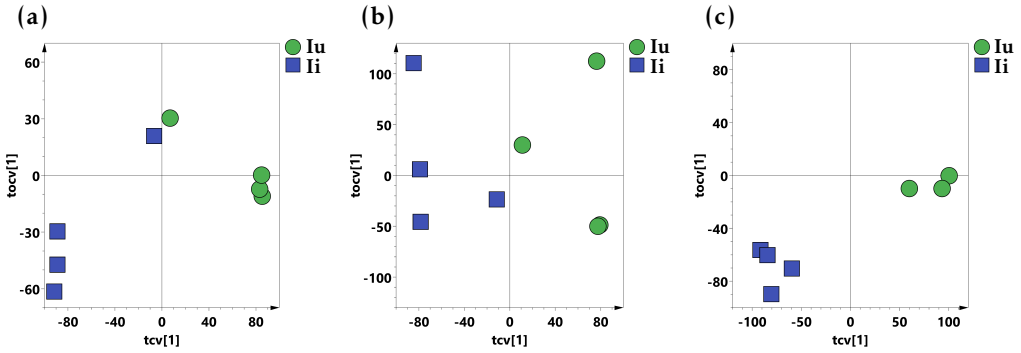

Figure 7.15: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Iu and Ii populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

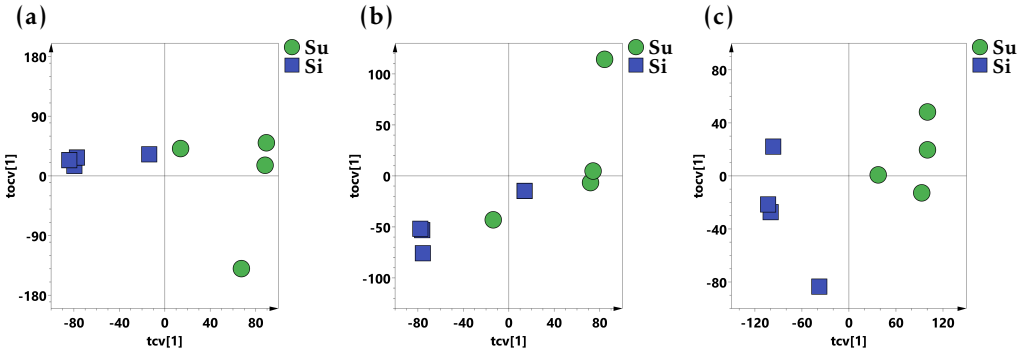

Figure 7.16: CV OPLS-DA scores plot with one predictive and one orthogonal component, for comparison between Su and Si populations (a) at Age 1 (5 days old as adults), (b) at Age 2 (20 days old as adults) and (c) at Age 3 (35 days old as adults).

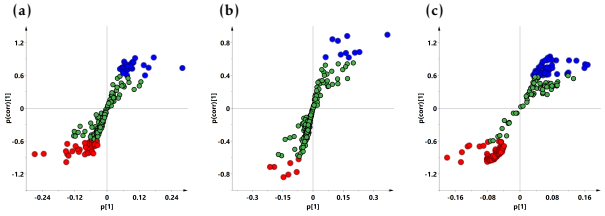

Figure 7.17: (a) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Is across the Age-1. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (b) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Is across the Age-2. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (c) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Is across the Age-3. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot.

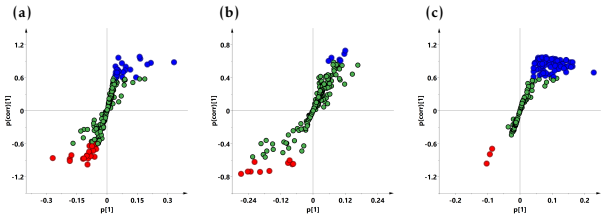

Figure 7.18: (a) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Ss across the Age-1. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (b) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Ss across the Age-2. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (c) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Ss across the Age-3. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot.

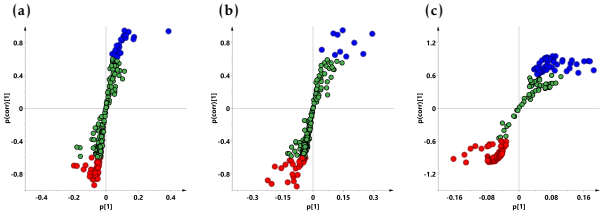

Figure 7.19: (a) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Ii across the Age-1. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (b) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Ii across the Age-2. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (c) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Iu and Ii across the Age-3. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot.

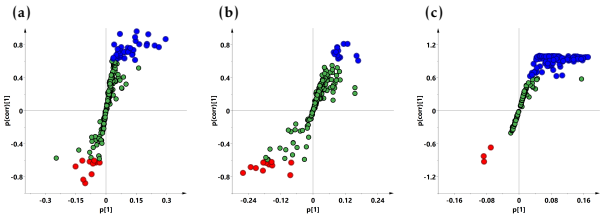

Figure 7.20: (a) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Si across the Age-1. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (b) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Si across the Age-2. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot. (c) In the corresponding S-plot, the variables with p(corr)[1] $\geq \pm$ 0.6 values were more important to discriminate between two groups Su and Si across the Age-3. These blue and red circles were relevant variables shown in upper-right and lower-left quadrants of the S-plot.

7.3.3 Metabolites concentration variation with individual treatments at different ages

We report the findings of metabolic alterations in Drosophila melanogaster in response to prick injury and bacterial infection with respect to control populations across the three ages. Temporal variations in relative concentrations of these metabolites were tracked across the treatments at three ages. The metabolome of the chosen populations has evolved to be distinct from that of the control (Su) populations even under un-infected (Iu) populations, as can be shown from the comparison between the un-infected selected (Iu) and control (Su) populations in Fig 7.7 (a) Age 1 (young flies; 5 days old as adults), (b) Age 2 (middle-ages flies; 20 days old as adults) and (c) Age 3 (old flies; 35 days old as adults) and in Fig 7.21 for the relative concentration. It has been observed that for the young flies when compared to the control population Su, the chosen un-infected population Iu has evolved to have much greater concentration of fatty acids, carbohydrates (glucose, galactose), organic acids (citric acid), amino acids (leucine, proline, lysine, arginine, alanine and threonine), NAD, and AMP. Similarly, for the Age 2 (middle-ages flies) the chosen un-infected population Iu has evolved to have much less concentration of fatty acid, an increase concentration of carbohydrates (glucose, galactose) and organic acid (citric acid), amino acids (leucine, proline, lysine, arginine, alanine, and threonine), NAD, and AMP with respect to control populations Su in comparison to Age-1. Similarly, for the Age 3 (old-ages flies) the chosen un-infected population Iu has evolved to have much more concentration of fatty acid, carbohydrates (glucose, galactose) and organic acid (citric acid), amino acids (leucine, proline, lysine, arginine, alanine, and threonine), NAD, and AMP with respect to control populations Su in respect to with Age-2.

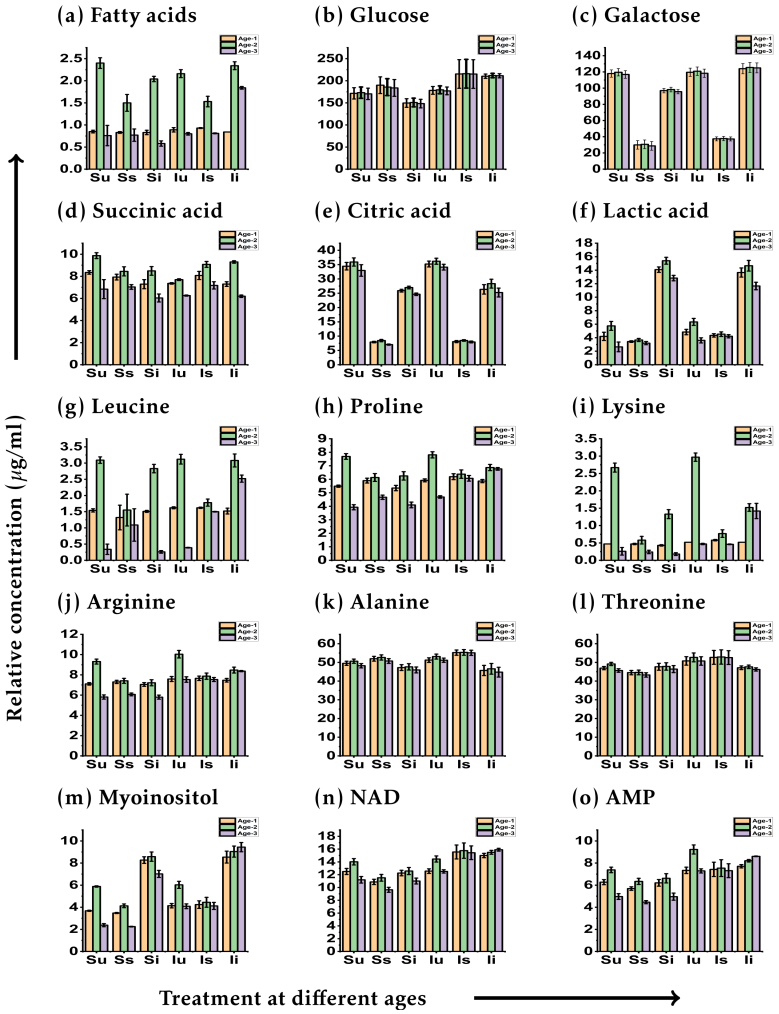

Figure 7.21: Figure depicting the relative concentrations of key metabolites across all of the I and S treatments at three different ages. The treatment (untreated u, sham-infected s, and bacterial infection i is shown on the X-axis, while the relative concentrations of the metabolites are indicated on the Y-axis.)

7.4 Conclusions

According to the conclusions drawn from the predictive scores of the OPLS-DA model, the metabolome of young and middle-aged flies that had been selected for immunity was different at the basal level and continued to be different in response to prick injury and infection. This was in comparison to the control populations. However, in terms

of the metabolome at the basal level, the old age flies from both the immune-selected group and the control population were quite similar. Even though there was less of a difference in the metabolic profiles of both populations at the basal level at the old age, the response to prick injury and infection was different in immune selected populations, as evidenced by large predictive scores. This indicates that the immune-selected populations were still better able to deal with injury or infections. Even though the basal level metabolic differences decrease with age, the ability to induce an immune response is still retained in populations that have evolved for better immunity. As a consequence of this, they continue to respond better to injuries or infections even when they are older when compared to flies that come from control populations.

Chapter 8

Summary and Future Outlook

The first project in this thesis deals with the investigation of the phytomedicinal compounds in the pericarp, skin & seeds of Momordica charantia fruit using NMR & Mass spectroscopy. We were able to determine the differential distributions of phytomedically significant metabolites in the pericarp, skin, and seeds of Momordica charantia fruit. We found that multivariate statistical analysis not only revealed a clustering of metabolic profiles, but also made it evident that the metabolic profile of the skin was distinct from that of the seeds and pericarp. Bio-assays were used to determine the total phenolic and flavonoid content of the fruit extracts. The presence of considerable levels of phenolic and flavonoid chemicals in the pericarp and seeds suggests that they are a valuable source of anti-oxidants. Anti-oxidant property was also confirmed by the bio-assays DPPH & ABTS. Anti-diabetic activity was evaluated by the alpha-glucosidase test. We found pericarp has higher anti-diabetic activity than its skin extracts.

The second project in this thesis discusses the importance of the phytomedicinal compounds in Phyllanthus emblica and Tinospora cordifolia. Phyllanthus emblica (Amla) fruit is used to treat several disease as in diabetes, swelling of the pancreas, eye problems, obesity, cancer and diarrhea. The Tinospora cordifolia (giloy) stem has been shown to facilitate better digestion as well as a reduction in constipation, acidity, gas, and bloating. We performed one- and two-dimensional NMR experiments to confirm the primary and secondary metabolites in the amla & giloy extracts. We confirmed the presence of fatty acids, amino acids, carbohydrates, organic acids, phenolic and flavonoid compounds peaks in these extracts. Multivariate statistics analysis was performed to identify the metabolic differences between the organic raw fruit and juice (Patanjali product) of Phyllanthus emblica (Amla). Supervised OPLS-DA model was used to identify the metabolic difference between the raw stem and juice (Patanjali product) of Tinospora cordifolia (Giloy). We identified the preservative and contaminant peaks in the Patanjali product to check the product genuine medicinal value.

The third project in this thesis focuses on identifying the cycling metabolites in young Helianthus annuus (sunflower) stems during the circadian cycle. We reported several primary metabolites including sugars, amino acids and tri-carboxylic acid (TCA) intermediates of the young sunflower stems using one- and two-dimensional NMR spectroscopy. These metabolites primarily support cell growth by providing energy and building blocks for the synthesis of essential bio-molecules, as well as serving as signaling molecules. The most typical visual clues for a plant's activity are the sections of the plant moving in different directions, such as bending, turning, twisting, or elongating. Sunflowers have an internal circadian clock that is controlled by growth hormones, and we studied how this clock allows sunflower to track the sun throughout the day. We confirmed the presence of auxin (Indole-3-acetic acid, IAA) hormonal molecule that regulates various aspects of plant growth, development and to stimulate cell elongation using mass-spectrometric methods.

The fourth project in this thesis describe circadian rhythms in Bougainvillea spectabilis leaves grown under controlled conditions with no exposure to vehicular emissions at the IISER Mohali campus. We identified the primary metabolites necessary to sustain plant life activities, growth and development. Several of the metabolites identified in our study showed a consistent rhythmic pattern during the circadian cycle, indicating that circadian rhythms are a strong influence on plant metabolism. We also studied the response to air pollution stress in Bougainvillea spectabilis leaves which resulted in changes in concentrations of amino acids, tricarboxylic acid cycle (TCA) intermediates and sugars. We observed that the plant circadian clock is intimately connected with plant response and tolerance to abiotic stress.

The fifth project in this thesis was to identify the metabolic changes in Bougainvillea plant leaves after a wounding treatment simulating a herbivore attack. We found that plants release toxic, repellent and anti-nutritional chemicals. These are secondary metabolites generated using primary metabolites that are closely related to plant disease, stress resistance and herbivory attack at several levels. We observed a rapid increase in the production of quinic acid, quercitol and choline after the wounding of the plants.

The sixth project in this thesis is devoted to the metabolic variations between a fly Drosophila melanogaster population chosen for increased immunity and its control population. Our studies reveal that populations develop different metabolomes in response to selection at three different ages. An investigation into the metabolic shifts that accompany the development of an immunological response to bacterial infection or prick damage was carried out on the fruit fly Drosophila melanogaster at three different ages. The metabolome of young and middle-aged flies that had been selected for immunity was different at the basal level and continued to be different in response to prick injury and infection. We observed the aged flies from both the immune-selected

group and the control population were quite similar in their metabolic response.

The findings of all of the studies presented in this thesis point to the potential that NMR-based metabolomics studies hold in integrating other techniques of systems biology, with the goal of providing answers to questions originating from a variety of subfields within the scientific community, including plant physiology, nutrition research, pharmacology, human disease, drug development, toxicology and cancer research.

Appendix A

NMR data processing tools

A.1 Data processing tools

A.1.1 Bruker TopSpin Software-NMR Data Analysis

Bruker's TopSpin software is utilised to control the operation and processing of the raw data in 600 MHz NMR spectrometer. The commands listed below are useful for processing data.

- command ej: eject the sample
- command ij: ij to lower your sample
- command new or edc: create a new experiment
- command lock: locks the spectrometer to the deuterium signal in your solvent choose
- command atma or atmm: matches and tunes to the correct frequency for your solvent and nucleus
- command ro: rotate or spin sample
- command topshim: shimming improves the homogeneity of the field
- command ased: acquisition parameters for the pulse sequence being run
- command eda: edit acquisition parameters, all acquisition parameters are shown
- command getprosol: reads in parameters from table specific for the probe and solvent

- command rga: sets receiver gain automatically
- command zg: zeros (overwrites) current data set and starts acquisition
- command tr: saves the data of the current number of scans while the experiment is still running.
- command halt: stops the acquisition run and saves the data.
- command stop: stops the acquisition without saving any data.
- command efp: Basic 1D Fourier transform command
- command apk: Automatic phase correction
- command abs: Automatic baseline correction and integration
- command sref: Automatic referencing of spectrum to TMS/TSP (if present)
- command ppf: Peak picking
- command proc2d: 2D automated processing
- command xfb: Basic 2D processing command
- command pp2d: peak picking in 2D spectra
- command int2d: automatic integration of 2D spectra
- command abs1: corrects F1 direction, abs2: corrects F2 direction

A.1.2 Mnova software

For combining NMR, LC/GC/MS, and electronic & vibrational spectroscopic techniques, there is a multivendor software suite called Mnova.

- Mnova software is utilised for the visualisation, processing, analysis, and reporting of 1D and 2D NMR data.
- High quality analysis of spectra (deconvolution, peak picking, integration, multiplet analysis, etc.
- Automated processing of your data (^{1}H, ^{13}C, DEPTs or any other 1D NMR as well as any 2D NMR correlations, such as HSQC, HMBC, NOESY, COSY, TOCSY, etc.).
- Used for NMR metabolomics data: Preprocessing, Binning, Normalization, Referencing, Scaling and so on.

Appendix B

Metabolomics data analysis

B.1 MetaboAnalyst (V5.0)

MetaboAnalyst is a web-based application that provides a straightforward and intuitive user interface, and it is designed specifically for the analysis of metabolomics data. The most recent version of MetaboAnalyst (V5.0) offers support for processing raw MS/NMR spectra, thorough data normalisation, statistical analysis, functional analysis, meta-analysis, and integrative analysis with other omics data. The goal is to reduce the time it takes to go from raw spectra to biological insights, allowing for high-throughput analysis in both targeted and untargeted metabolomics.
MetaboAnalyst is able to process a wide range of data types that are typically generated during metabolomic research. These data types include compound concentration data, binned NMR/MS spectra data, NMR/MS peak list data, and MS spectra data (NetCDF, mzXML, mzDATA). MetaboAnalyst provides users with access to a variety of methodologies that are frequently applied while analysing metabolomic data. They are as follows:

B.1.1 Univariate Analysis: For two groups

Exploratory data analysis almost always makes use of univariate analysis methods. These methods are the most prevalent. MetaboAnalyst offers a fold change (FC) analysis, t-tests, and a volcano plot, which is a combination of the first two approaches. These analyses can be performed on data with two groups. Each of these three approaches is capable of doing analyses with and without paired data.

- **Fold Change Analysis**
- **T-tests**

- **Volcano plot**

B.1.2 Univariate Analysis: For multiple groups

MetaboAnalyst offers two different types of studies for multi-group study: a one-way analysis of variance (ANOVA) with related post-hoc analyses, and a correlation analysis to discover significant molecules that follow a given pattern.

- **One-way Analysis of Variance (ANOVA) & post-hoc analysis**

B.1.3 Multivariate analysis methods & Chemometrics Analysis

- **Principal Component Analysis (PCA)** PCA is an unsupervised method that finds the directions that best describe data set (X) variation without class labels (Y).
- **Partial Least Squares-Discriminant Analysis (PLS-DA)** PLS is a supervised method that makes use of multivariate regression techniques to extract, via linear combination of the original variables (X), the information that might predict the class membership (Y).
- **Orthogonal Partial Least Squares - Discriminant Analysis (OPLS-DA) (For two groups)** Similar to PLS-DA, OPLS-DA is a powerful tool that can be used to reduce the number of dimensions and identify spectral features that are responsible for driving group separation.

B.1.4 R Command History: Example test data

mSet←InitDataObjects(”specbin”, ”stat”, FALSE)
mSet←Read.TextData(mSet, ”Replacing with your file path”, ”rowu”, ”disc”);
mSet←SanityCheckData(mSet)
mSet←ReplaceMin(mSet);
mSet←SanityCheckData(mSet)
mSet←FilterVariable(mSet, ”none”, ”F”, 25)
mSet←PreparePrenormData(mSet)
mSet←Normalization(mSet, ”SumNorm”, ”LogNorm”, ”ParetoNorm”, ratio=FALSE, ratioNum=20)
mSet←PlotNormSummary(mSet, ”norm 0 ”, ”png”, 72, width=NA)
mSet←PlotSampleNormSummary(mSet, ”snorm 0 ”, ”png”, 72, width=NA)
mSet←PCA.Anal(mSet)

mSet←PlotPCAPairSummary(mSet, ”pca pair 0 ”, ”png”, 72, width=NA, 5)
mSet←PlotPCAScree(mSet, ”pca scree 0 ”, ”png”, 72, width=NA, 5)
mSet←PlotPCA2DScore(mSet, ”pca score2d 0 ”, ”png”, 72, width=NA, 1,2,0.95,0,0)
mSet←PlotPCALoading(mSet, ”pca loading 0 ”, ”png”, 72, width=NA, 1,2);
mSet←PlotPCABiplot(mSet, ”pca biplot 0 ”, ”png”, 72, width=NA, 1,2)
mSet←PlotPCA3DLoading(mSet, ”pca loading3d 0 ”, ”json”, 1,2,3)

B.2 SIMCA - Multivariate Data Analysis Software - Sartorius

The SIMCA Multivariate Data Analysis software enables you to view trends and clusters derived from a variety of data sources, including batch process groups, time-series data, and multiple sources.

- **PCA & loading plot analysis**
- **PLS & VIP plot analysis**
- **OPLS-DA & S-plot analysis**

B.2.1 Command History

The work process in SIMCA consists of the following steps: Create a new project by importing a dataset and specifying data properties in the SIMCA import spreadsheet.

- Quick Access Toolbar: A customizable toolbar that displays frequently used commands.
- Regular tabs File, Home, Data, Batch, Analyze, Predict, Plot/List, and View.
- New: Creating a new project (To import from a database).
- Add data: Import more than one dataset at the same time in SIMCA Import.
- Finish: To import all currently open spreadsheets.
- Primary ID: Indicating file contents (desired for both variables and observations).
- Secondary ID: Indicating file contents (desired for both variables and observations).

- Finish: To import all currently open spreadsheets.
- Dataset: To open any available dataset (Viewing and preprocessing the data).
- Quick Info: The Quick Info pane holds interactive plots tied to the dataset displaying variables or observations in the time or frequency domain.
- Modifying the workset: To fit a model different from the default model, with maybe other or more datasets included, excluded variables and observations, or transformations, or different scaling, it is necessary to modify the default Workset. Defining variables as X or Y.
- Variables: Excluding & including variables
- Observations: Including & excluding observations. Grouping observations into classes for classification.
- Transform: Applying transformations for variables.
- Lag: Defining lag structure.
- Expand: Expanding terms with cross, square, and cubic terms.
- Scale: Scaling variables.
- Statistics: Workset statistics – Displays statistics of selected variables of the active model.
- Statistics: Correlation matrix – Displays the correlations between the variables for the active model.
- Change model type - Displays the available model types:
- Overview - PCA-X, PCA-Y, PCA-X&Y, O2PLS.
- Regression - PLS, OPLS.
- Discriminant analysis - PLS-DA, OPLS-DA, O2PLS-DA.
- Class - PCA, PLS, OPLS, O2PLS.
- Clustering - PLS-Tree.
- Autofit: Rule based fitting of the marked model or group of models (Fitting the model).

- Two first: Calculates two components whether they are significant or not. Often used to get a quick overview of the data.
- Add: Calculate one component at a time. Here it is possible to force components to be calculated regardless of significance rules.
- Remove: Remove the last component.

References

[1] A. Derome, The use of NMR spectroscopy in the structure determination of natural products: two-dimensional methods, Nat. Prod. Rep. **6**(2), 111–141 (1989). 1

[2] R. Claramunt, C. López, M. Santa María, D. Sanz, and J. Elguero, The use of NMR spectroscopy to study tautomerism, Prog. Nucl. Magn. Reson. Spectrosc. **49**(3-4), 169–206 (2006). 1

[3] G. K. Radda, The use of NMR spectroscopy for the understanding of disease, Science **233**(4764), 640–645 (1986). 1

[4] J. L. Ward, J. M. Baker, and M. H. Beale, Recent applications of NMR spectroscopy in plant metabolomics, FEBS J. **274**(5), 1126–1131 (2007). 1

[5] L. Cardoza, A. Korir, W. Otto, C. Wurrey, and C. Larive, Applications of NMR spectroscopy in environmental science, Prog. Nucl. Magn. Reson. Spectrosc. **45**(3-4), 209–238 (2004). 1

[6] D. S. Wishart, Applications of metabolomics in drug discovery and development, Drugs in R & D **9**(5), 307–322 (2008). 1

[7] A.-H. Emwas, R. Roy, R. T. McKay, L. Tenori, E. Saccenti, G. N. Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko, et al., NMR spectroscopy for metabolomics research, Metabolites **9**(7), 123 (2019). 1, 11

[8] D. S. Wishart, NMR metabolomics: A look ahead, J. Magn. Reson. **306**, 155–161 (2019). 1

[9] D. S. Wishart, L. L. Cheng, V. Copié, A. S. Edison, H. R. Eghbalnia, J. C. Hoch, G. J. Gouveia, W. Pathmasiri, R. Powers, T. B. Schock, et al., NMR and Metabolomics—A Roadmap for the Future, Metabolites **12**(8), 678 (2022). 1

[10] K. Dettmer, P. A. Aronov, and B. D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. **26**(1), 51–78 (2007). 1

[11] Z. Lei, D. V. Huhman, and L. W. Sumner, Mass spectrometry strategies in metabolomics, J. Biol. Chem. **286**(29), 25435–25442 (2011). 1

[12] G. Gowda and D. Djukovic, Overview of mass spectrometry-based metabolomics: opportunities and challenges, Mass Spectrometry in Metabolomics , 3–12 (2014). 1

[13] S. G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, Mass spectrometry in metabolome analysis, Mass Spectrom. Rev. **24**(5), 613–646 (2005). 1

[14] A.-H. M. Emwas, The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research, in Metabonomics, pages 161–193, Springer, 2015. 1

[15] O. Corcoran and M. Spraul, LC–NMR–MS in drug discovery, Drug Discov. Today **8**(14), 624–631 (2003). 1

[16] I. I. Rabi, S. Millman, P. Kusch, and J. R. Zacharias, The molecular beam resonance method for measuring nuclear magnetic moments. the magnetic moments of li 6 3, li 7 3 and f 19 9, Phys. Rev. **55**(6), 526 (1939). 1

[17] F. Bloch, Nuclear induction, Phys. Rev. **70**(7-8), 460 (1946). 1

[18] E. M. Purcell, H. C. Torrey, and R. V. Pound, Resonance absorption by nuclear magnetic moments in a solid, Phys. Rev. **69**(1-2), 37 (1946). 1

[19] M. A. Shampo, R. A. Kyle, and D. P. Steensma, Richard Ernst—Nobel Prize for nuclear magnetic resonance spectroscopy, in Mayo Clin. Proc., volume 87, page e109, Elsevier, 2012. 1

[20] K. Nagayama, K. Wüthrich, P. Bachmann, and R. Ernst, Two-dimensional J-resolved 1H nmr spectroscopy for studies of biological macromolecules, Biochem. Biophys. Res. Commun. **78**(1), 99–105 (1977). 2

[21] P. C. Lauterbur, Image formation by induced local interactions: examples employing nuclear magnetic resonance, nature **242**(5394), 190–191 (1973). 2

[22] C. Kittel, P. McEuen, and P. McEuen, Introduction to solid state physics, volume 8, Wiley New York, 1996. 2

[23] I. Oliveira, R. Sarthour Jr, T. Bonagamba, E. Azevedo, and J. C. Freitas, NMR quantum information processing, Elsevier, 2011. 2, 3, 6

[24] F. Reif, Fundamentals of statistical and thermal physics, Waveland Press, 2009. 3

[25] J. J. Sakurai and E. D. Commins, Modern quantum mechanics, revised edition, 1995. 4

[26] R. Ernst, G. Bodenhausen, and A. Wokaun, Principles of NMR in one and two dimensions, Clarendon, 1987. 5

[27] C. P. Slichter, Principles of magnetic resonance, volume 1, Springer Science & Business Media, 2013. 6

[28] J. D. Jackson, Classical electrodynamics 3rd ed john wiley & sons, Inc., NewYork, NY (1999). 7

[29] M. H. Levitt, Spin dynamics: basics of nuclear magnetic resonance, John Wiley & Sons, 2013. 8

[30] J. Keeler, Understanding NMR spectroscopy, John Wiley & Sons, 2010. 8

[31] J. Kruk, M. Doskocz, E. Jodłowska, A. Zacharzewska, J. Łakomiec, K. Czaja, and J. Kujawski, NMR techniques in metabolomic studies: A quick overview on examples of utilization, Appl. Magn. Reson. **48**(1), 1–21 (2017). 10

[32] A. Le Guennec, F. Tayyari, and A. S. Edison, Alternatives to nuclear overhauser enhancement spectroscopy presat and carr–purcell–meiboom–gill presat for NMR-based metabolomics, Anal. Chem. **89**(17), 8582–8588 (2017). 10

[33] A. C. Dona, M. Kyriakides, F. Scott, E. A. Shephard, D. Varshavi, K. Veselkov, and J. R. Everett, A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments, Comput. Struct. Biotechnol. J. **14**, 135–153 (2016). 10

[34] C. Ludwig and M. R. Viant, Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox, Phytochem Anal. **21**(1), 22–32 (2010). 10

[35] T. De Meyer, D. Sinnaeve, B. Van Gasse, E.-R. Rietzschel, M. L. De Buyzere, M. R. Langlois, S. Bekaert, J. C. Martins, and W. Van Criekinge, Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum 1H-NMR spectra, Anal. Bioanal. Chem. **398**(4), 1781–1790 (2010). 10

[36] D. S. Wishart, Metabolomics for investigating physiological and pathophysiological processes, Physiol. Rev. **99**(4), 1819–1875 (2019). 10

[37] P. G. Takis, V. Ghini, L. Tenori, P. Turano, and C. Luchinat, Uniqueness of the NMR approach to metabolomics, Trends Anal. Chem. **120**, 115300 (2019). 11

[38] A. Vignoli, V. Ghini, G. Meoni, C. Licari, P. G. Takis, L. Tenori, P. Turano, and C. Luchinat, High-throughput metabolomics by 1D NMR, Angew. Chem. Int. Ed. **58**(4), 968–994 (2019). 11

[39] A. Smolinska, L. Blanchet, L. M. Buydens, and S. S. Wijmenga, NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review, Anal. Chim. Acta **750**, 82–97 (2012). 11

[40] S. K. Bharti and R. Roy, Quantitative 1H NMR spectroscopy, Trends Anal. Chem. **35**, 5–26 (2012). 11

[41] D. S. Wishart, Quantitative metabolomics using NMR, Trends Anal. Chem. **27**(3), 228–237 (2008). 11

[42] G. G. Harrigan and R. Goodacre, Metabolic profiling: its role in biomarker discovery and gene function analysis: its role in biomarker discovery and gene function analysis, Springer Science & Business Media, 2003. 11, 12

[43] J. C. Lindon, J. K. Nicholson, and E. Holmes, The handbook of metabonomics and metabolomics, Elsevier, 2011. 11

[44] W. J. Griffiths, Metabolomics, metabonomics and metabolite profiling, Royal Society of Chemistry, 2007. 11

[45] J. G. Bundy, M. P. Davey, and M. R. Viant, Environmental metabolomics: a critical review and future perspectives, Metabolomics **5**(1), 3–21 (2009). 11

[46] P. Krishnan, N. Kruger, and R. Ratcliffe, Metabolite fingerprinting and profiling in plants using NMR, J. Exp. Bot. **56**(410), 255–265 (2005). 11

[47] Y. Sekiyama, E. Chikayama, and J. Kikuchi, Evaluation of a semipolar solvent system as a step toward heteronuclear multidimensional NMR-based metabolomics for 13C-labeled bacteria, plants, and animals, Anal. Chem. **83**(3), 719–726 (2011). 11

[48] W. J. Griffiths, T. Koal, Y. Wang, M. Kohl, D. P. Enot, and H.-P. Deigner, Targeted metabolomics for biomarker discovery, Angew. Chem. Int. Ed. **49**(32), 5426–5445 (2010). 12

[49] S. A. Aherne and N. M. O'Brien, Dietary flavonols: chemistry, food content, and metabolism, Nutrition **18**(1), 75–81 (2002). 12

[50] F. Bourgaud, A. Gravot, S. Milesi, and E. Gontier, Production of plant secondary metabolites: a historical perspective, Plant Sci. **161**(5), 839–851 (2001). 12

[51] G. Valentino, V. Graziani, B. D'Abrosca, S. Pacifico, A. Fiorentino, and M. Scognamiglio, NMR-based plant metabolomics in nutraceutical research: an overview, Molecules **25**(6), 1444 (2020). 12

[52] V. Shulaev, D. Cortes, G. Miller, and R. Mittler, Metabolomics for plant stress response, Physiol. Plant. **132**(2), 199–208 (2008). 12, 95

[53] F. Tugizimana, L. Piater, and I. Dubery, Plant metabolomics: A new frontier in phytochemical analysis, S. Afr. J. Sci. **109**(5-6), 01–11 (2013). 12

[54] J. Sardans, A. Gargallo-Garriga, M. Pérez-Trujillo, T. Parella, R. Seco, I. Filella, and J. Penuelas, Metabolic responses of Quercus ilex seedlings to wounding analysed with nuclear magnetic resonance profiling, Plant Biol. **16**(2), 395–403 (2014). 12, 131

[55] A. Singh, R. K. Sharma, M. Chagtoo, G. Agarwal, N. George, N. Sinha, and M. M. Godbole, 1H NMR metabolomics reveals association of high expression of inositol 1, 4, 5 trisphosphate receptor and metabolites in breast cancer patients, PLoS One **12**(1), e0169330 (2017). 12

[56] S. Suman, R. K. Sharma, V. Kumar, N. Sinha, and Y. Shukla, Metabolic fingerprinting in breast cancer stages through 1H NMR spectroscopy-based metabolomic analysis of plasma, J. Pharm. Biomed. Anal. **160**, 38–45 (2018). 12

[57] R. Ranjan and N. Sinha, Nuclear magnetic resonance (NMR)-based metabolomics for cancer research, NMR Biomed. **32**(10), e3916 (2019). 12

[58] S. S. Ahmed, W. Santosh, S. Kumar, and H. T. T. Christlet, Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection, J. Biomed. Sci. **16**(1), 1–12 (2009). 12

[59] A. J. Sinclair, M. R. Viant, A. K. Ball, M. A. Burdon, E. A. Walker, P. M. Stewart, S. Rauz, and S. P. Young, NMR-based metabolomic analysis of cerebrospinal fluid and serum in neurological diseases–a diagnostic tool?, NMR Biomed. **23**(2), 123–132 (2010). 12

[60] J.-H. Shin, J.-Y. Yang, B.-Y. Jeon, Y. J. Yoon, S.-N. Cho, Y.-H. Kang, D. H. Ryu, and G.-S. Hwang, 1H NMR-based metabolomic profiling in mice infected with Mycobacterium tuberculosis, J. Proteome Res. **10**(5), 2238–2247 (2011). 12

[61] R. Teng, P. R. Junankar, W. A. Bubb, C. Rae, P. Mercier, and K. Kirk, Metabolite profiling of the intraerythrocytic malaria parasite Plasmodium falciparum by 1H NMR spectroscopy, NMR Biomed. **22**(3), 292–302 (2009). 12

[62] C. M. Slupsky, NMR-based analysis of metabolites in urine provides rapid diagnosis and etiology of pneumonia, Biomark. Med. **4**(2), 195–197 (2010). 12

[63] R. Wedeking, M. Maucourt, C. Deborde, A. Moing, Y. Gibon, H. E. Goldbach, and M. A. Wimmer, 1H-NMR metabolomic profiling reveals a distinct metabolic recovery response in shoots and roots of temporarily drought-stressed sugar beets, PloS one **13**(5), e0196102 (2018). 12

[64] S. S. Datir, S. Yousf, S. Sharma, M. Kochle, A. Ravikumar, and J. Chugh, Cold storage reveals distinct metabolic perturbations in processing and non-processing cultivars of potato (Solanum tuberosum L.), Sci. Rep. **10**(1), 1–13 (2020). 12

[65] E. Saccenti, H. C. Hoefsloot, A. K. Smilde, J. A. Westerhuis, and M. M. Hendriks, Reflections on univariate and multivariate analysis of metabolomics data, Metabolomics **10**(3), 361–374 (2014). 14

[66] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, and O. Yanes, A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data, Metabolites **2**(4), 775–795 (2012). 14

[67] B. Worley and R. Powers, Multivariate analysis in metabolomics, Curr. Metabolomics **1**(1), 92–107 (2013). 14, 82, 147

[68] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, and C. Vikström, Multi-and megavariate data analysis basic principles and applications, volume 1, Umetrics Academy, 2013. 14

[69] M. Kortesniemi, C. M. Slupsky, T. Ollikka, L. Kauko, A. R. Spevacek, O. Sjövall, B. Yang, and H. Kallio, NMR profiling clarifies the characterization of Finnish honeys of different botanical origins, Food Res. Int. **86**, 83–92 (2016). 14

[70] H. Hotelling, Analysis of a complex of statistical variables into principal components., J. Educ. Psychol. **24**(6), 417 (1933). 14

[71] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, Lond. Edinb. Dublin philos. mag. j. sci. **2**(11), 559–572 (1901). 14

[72] R. G. Brereton and G. R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. **28**(4), 213–225 (2014). 14

[73] E. Szymańska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics **8**(1), 3–16 (2012). 14

[74] P. S. Gromski, H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner, and R. Goodacre, A tutorial review: Metabolomics and partial least squares-discriminant analysis–a marriage of convenience or a shotgun wedding, Anal. Chim. Acta **879**, 10–23 (2015). 14

[75] M. N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge, and P. Savarin, PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters, Mol. Biosyst. **11**(1), 13–19 (2015). 14

[76] B. Worley and R. Powers, PCA as a practical indicator of OPLS-DA model reliability, Curr. Metabolomics **4**(2), 97–103 (2016). 14

[77] M. Bylesjö, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes, and J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, J. Chemom. **20**(8-10), 341–351 (2006). 14, 15

[78] J. K. Nicholson, J. C. Lindon, and E. Holmes, 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, Xenobiotica **29**(11), 1181–1189 (1999). 15

[79] M. Grootveld, Metabolic Profiling: Disease and Xenobiotics, Royal Society of Chemistry, 2014. 15

[80] J. Xia, N. Psychogios, N. Young, and D. S. Wishart, MetaboAnalyst: a web server for metabolomic data analysis and interpretation, Nucleic Acids Res. **37**(suppl_2), W652–W660 (2009). 15

[81] J. Trygg and S. Wold, Orthogonal projections to latent structures (O-PLS), J. Chemom. **16**(3), 119–128 (2002). 15

[82] D. Balmer, V. Flors, G. Glauser, and B. Mauch-Mani, Metabolomics of cereals under biotic stress: current knowledge and techniques, Front. Plant Sci. **4**, 82 (2013). 16

[83] K. T. X. Vo, M. M. Rahman, M. M. Rahman, K. T. T. Trinh, S. T. Kim, and J.-S. Jeon, Proteomics and metabolomics studies on the biotic stress responses of rice: an update, Rice **14**(1), 1–16 (2021). 16

[84] R. Nakabayashi and K. Saito, Integrated metabolomics for abiotic stress responses in plants, Curr. Opin. Plant Biol. **24**, 10–16 (2015). 16, 95

[85] T. Obata and A. R. Fernie, The use of metabolomics to dissect plant responses to abiotic stresses, Cell. Mol. Life Sci. **69**(19), 3225–3243 (2012). 16, 95

[86] H. Winning, N. Viereck, B. Wollenweber, F. Larsen, S. Jacobsen, I. Søndergaard, and S. Engelsen, Exploring abiotic stress on asynchronous protein metabolism in single kernels of wheat studied by NMR spectroscopy and chemometrics, J. Exp. Bot. **60**(1), 291–300 (2009). 16

[87] Z. Feng, C. Ding, W. Li, D. Wang, and D. Cui, Applications of metabolomics in the research of soybean plant under abiotic stress, Food Chem. **310**, 125914 (2020). 16

[88] P. C. Bueno and N. P. Lopes, Metabolomics to characterize adaptive and signaling responses in legume crops under abiotic stresses, ACS omega **5**(4), 1752–1763 (2020). 16

[89] R. Mittler, Abiotic stress, the field environment and stress combination, Trends Plant Sci. **11**(1), 15–19 (2006). 16

[90] G. P. Bolwell, L. V. Bindschedler, K. A. Blee, V. S. Butt, D. R. Davies, S. L. Gardner, C. Gerrish, and F. Minibayeva, The apoplastic oxidative burst in response to biotic stress in plants: a three-component system, J. Exp. Bot. **53**(372), 1367–1376 (2002). 16

[91] L. Taiz, E. Zeiger, I. M. Møller, A. Murphy, et al., Plant physiology and development., Sinauer Associates Incorporated, 2015. 16

[92] C. R. McClung, Circadian rhythms in plants, Annu. Rev. Plant Biol. **52**(1), 139–162 (2001). 16, 75

[93] C. Weissman, The metabolic response to stress: an overview and update., Anesthesiology **73**(2), 308–327 (1990). 16

[94] C. Zipfel and G. E. Oldroyd, Plant signalling in symbiosis and immunity, Nature **543**(7645), 328–336 (2017). 16

[95] M. Heil and J. Ton, Long-distance signalling in plant defence, Trends Plant Sci. **13**(6), 264–272 (2008). 16

[96] H. M. Heyman and J. J. M. Meyer, NMR-based metabolomics as a quality control tool for herbal products, S. Afr. J. Bot. **82**, 21–32 (2012). 16

[97] A. Malmendal, J. Overgaard, J. G. Bundy, J. G. Sørensen, N. C. Nielsen, V. Loeschcke, and M. Holmstrup, Metabolomic profiling of heat stress: hardening and recovery of homeostasis in Drosophila, Am. J. Physiol. Regul. Integr. Comp. Physiol. **291**(1), R205–R212 (2006). 17, 143

[98] J. D. Feala, L. Coquin, A. D. McCulloch, and G. Paternostro, Flexibility in energy metabolism supports hypoxia tolerance in Drosophila flight muscle: metabolomic and computational systems analysis, Mol. Syst. Biol. **3**(1), 99 (2007). 17, 143

[99] J. Overgaard, A. Malmendal, J. G. Sørensen, J. G. Bundy, V. Loeschcke, N. C. Nielsen, and M. Holmstrup, Metabolomic profiling of rapid cold hardening and cold shock in Drosophila melanogaster, J. Insect Physiol. **53**(12), 1218–1232 (2007). 17, 143, 146

[100] J. D. Feala, L. Coquin, G. Paternostro, and A. D. McCulloch, Integrating metabolomics and phenomics with systems models of cardiac hypoxia, Prog. Biophys. Mol. Biol. **96**(1), 209–225 (2008), Cardiovascular Physiome. 17, 143

[101] N. Gogna, V. J. Singh, V. Sheeba, and K. Dorai, NMR-based investigation of the Drosophila melanogaster metabolome under the influence of daily cycles of light and temperature, Mol. BioSyst. **11**, 3305–3315 (2015). 17, 143, 147

[102] R. Horax, N. Hettiarachchy, and P. Chen, Extraction, quantification, and antioxidant activities of phenolics from pericarp and seeds of bitter melons (Momordica charantia) harvested at three maturity stages (immature, mature, and ripe), J. Agric. Food Chem. **58**(7), 4428–4433 (2010). 19

[103] K. Zhou, H. Wang, W. Mei, X. Li, Y. Luo, and H. Dai, Antioxidant activity of papaya seed extracts, Molecules **16**(8), 6179–6192 (2011). 19, 23

[104] P. Nerurkar and R. B. Ray, Bitter melon: antagonist to cancer, Pharmaceutical research **27**(6), 1049–1053 (2010). 19

[105] L. Guan, Synthesis and anti-tumour activities of sulphated polysaccharide obtained from Momordica charantia, Nat. Prod. Res. **26**(14), 1303–1309 (2012). 19

[106] Z. Yang, W. Shi, Z. Shen, Y. Zhang, and J. Yang, Preventive effect of Momordica charantia L. saponins on osteoporosis in ovariectomized rats, Food Sci **31**(7), 272–5 (2010). 19

[107] G. Yaldız, N. Sekeroglu, M. Kulak, and G. Demirkol, Antimicrobial activity and agricultural properties of bitter melon (Momordica charantia L.) grown in northern parts of Turkey: a case study for adaptation, Nat. Prod. Res. **29**(6), 543–545 (2015). 19

[108] P.-C. Hsiao, C.-C. Liaw, S.-Y. Hwang, H.-L. Cheng, L.-J. Zhang, C.-C. Shen, F.-L. Hsu, and Y.-H. Kuo, Antiproliferative and hypoglycemic cucurbitane-type glycosides from the fruits of Momordica charantia, J. Agric. Food Chem. **61**(12), 2979–2986 (2013). 19

[109] F. Zhang, L. Lin, and J. Xie, A mini-review of chemical and biological properties of polysaccharides from Momordica charantia, Int. J. Biol. Macromol. **92**, 246–253 (2016). 19

[110] Y. Liu, N. T. Nyberg, A. K. Jäger, and D. Staerk, Facilitated Visual Interpretation of Scores in Principal Component Analysis by Bioactivity-Labeling of 1H-NMR Spectra—Metabolomics Investigation and Identification of a New α-Glucosidase Inhibitor in Radix Astragali, Molecules **22**(3), 411 (2017). 19

[111] A. A. B. Sajak, A. Mediani, N. S. M. Dom, C. Machap, M. Hamid, A. Ismail, A. Khatib, F. Abas, et al., Effect of Ipomoea aquatica ethanolic extract in streptozotocin (STZ) induced diabetic rats via 1H NMR-based metabolomics approach, Phytomedicine **36**, 201–209 (2017). 19

[112] L. Harinantenaina, M. Tanaka, S. Takaoka, M. Oda, O. Mogami, M. Uchida, and Y. Asakawa, Momordica charantia constituents and antidiabetic screening of the isolated major compounds, Chem. Pharm. Bull. **54**(7), 1017–1021 (2006). 19

[113] M. F. Mahmoud, F. E. Z. Z. El Ashry, N. N. El Maraghy, and A. Fahmy, Studies on the antidiabetic activities of Momordica charantia fruit juice in streptozotocin-induced diabetic rats, Pharm. Biol. **55**(1), 758–765 (2017). 19

[114] Y.-c. Li, X.-j. Xu, J. Yang, X.-g. Wu, and Q.-y. Fu, One new 19-nor cucurbitane-type triterpenoid from the stems of Momordica charantia, Nat. Prod. Res. **30**(8), 973–978 (2016). 19

[115] P. Kulkarni, S. Lohidasan, and K. Mahadik, Isolation, characterisation and investigation of in vitro antidiabetic and antioxidant activity of phytoconstituents from fruit of Momordica charantia Linn, Nat. Prod. Res. **35**(6), 1035–1037 (2021). 19

[116] C. Han, Q. Hui, and Y. Wang, Hypoglycaemic activity of saponin fraction extracted from Momordica charantia in PEG/salt aqueous two-phase systems, Nat. Prod. Res. **22**(13), 1112–1119 (2008). 19

[117] R. M. Hafizur, N. Kabir, and S. Chishti, Modulation of pancreatic β-cells in neonatally streptozotocin-induced type 2 diabetic rats by the ethanolic extract of Momordica charantia fruit pulp, Nat. Prod. Res. **25**(4), 353–367 (2011). 19

[118] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, et al., Proposed minimum reporting standards for chemical analysis, Metabolomics **3**(3), 211–221 (2007). 20, 22, 74, 146

[119] F. Savorani, G. Tomasi, and S. B. Engelsen, icoshift: A versatile tool for the rapid alignment of 1D NMR spectra, J. Magn. Reson. **202**(2), 190–202 (2010). 20, 22, 57, 147

[120] G. Fan, W.-Z. Luo, S.-H. Luo, Y. Li, X.-L. Meng, X.-D. Zhou, and Y. Zhang, Metabolic discrimination of Swertia mussotii and Swertia chirayita known as "Zangyinchen" in traditional Tibetan medicine by 1H NMR-based metabolomics, J. Pharm. Biomed. Anal. **98**, 364–370 (2014). 20, 56

[121] N. Gogna, N. Hamid, and K. Dorai, Metabolomic profiling of the phytomedicinal constituents of Carica papaya L. leaves and seeds by 1H NMR spectroscopy and multivariate statistical analysis, J. Pharm. Biomed. Anal. **115**, 74–85 (2015). 20, 23, 99, 101

[122] E. A. Mahrous and M. A. Farag, Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review, J. Adv. Res. **6**(1), 3–15 (2015). 20

[123] A. Meda, C. E. Lamien, M. Romito, J. Millogo, and O. G. Nacoulma, Determination of the total phenolic, flavonoid and proline contents in Burkina Fasan

honey, as well as their radical scavenging activity, Food Chem. **91**(3), 571–577 (2005). 20, 23, 102

[124] N. S. Rajurkar and S. Hande, Estimation of phytochemical content and antioxidant activity of some selected traditional Indian medicinal plants, Indian J. Pharm. Sci. **73**(2), 146 (2011). 20, 24, 103

[125] S. Poovitha and M. Parani, In vitro and in vivo α-amylase and α-glucosidase inhibiting activities of the protein extracts from two varieties of bitter gourd (Momordica charantia L.), BMC Complement. Altern. Med. **16**(1), 1–8 (2016). 20, 25

[126] B. Joseph and D. Jini, Antidiabetic effects of Momordica charantia (bitter melon) and its medicinal potency, Asian Pac. J. Trop. Dis. **3**(2), 93–102 (2013). 30

[127] O. Kenny, T. Smyth, C. Hewage, and N. Brunton, Antioxidant properties and quantitative UPLC-MS analysis of phenolic compounds from extracts of fenugreek (Trigonella foenum-graecum) seeds and bitter melon (Momordica charantia) fruit, Food Chem. **141**(4), 4295–4302 (2013). 47

[128] L. Aksoy, E. Kolay, Y. Ağılönü, Z. Aslan, and M. Kargıoğlu, Free radical scavenging activity, total phenolic content, total antioxidant status, and total oxidant status of endemic Thermopsis turcica, Saudi J. Biol. Sci. **20**(3), 235–239 (2013). 47

[129] A. Upadhyay, P. Agrahari, and D. Singh, A review on salient pharmacological features of Momordica charantia, Int. J. Pharmacol. **11**(5), 405–413 (2015). 49

[130] D. M. Cuong, J. Jeon, A. M. Morgan, C. Kim, J. K. Kim, S. Y. Lee, and S. U. Park, Accumulation of charantin and expression of triterpenoid biosynthesis genes in bitter melon (Momordica charantia), J. Agric. Food Chem. **65**(33), 7240–7249 (2017). 49

[131] S. Kakkar and S. Bais, A review on protocatechuic acid and its pharmacological potential, Int. Sch. Res. Notices **2014** (2014). 49

[132] J. C. H. Singh, R. M. Kakalij, R. P. Kshirsagar, B. H. Kumar, S. S. B. Komakula, and P. V. Diwan, Cognitive effects of vanillic acid against streptozotocin-induced neurodegeneration in mice, Pharm. Biol. **53**(5), 630–636 (2015). 49

[133] J. Liggins, L. J. Bluck, S. Runswick, C. Atkinson, W. A. Coward, and S. A. Bingham, Daidzein and genistein content of fruits and nuts, J. Nutr. Biochem. **11**(6), 326–331 (2000). 49

[134] S. Ou and K.-C. Kwok, Ferulic acid: pharmaceutical functions, preparation and applications in foods, J. Sci. Food Agric. **84**(11), 1261–1269 (2004). 49

[135] A. Suzuki, N. Yamamoto, H. Jokura, M. Yamamoto, A. Fujii, I. Tokimitsu, and I. Saito, Chlorogenic acid attenuates hypertension and improves endothelial function in spontaneously hypertensive rats, J. Hypertens. **24**(6), 1065–1073 (2006). 49

[136] K. C. Ong and H.-E. Khoo, Biological effects of myricetin, General pharmacology: The vascular system **29**(2), 121–126 (1997). 49

[137] L. G. Menon, R. Kuttan, and G. Kuttan, Anti-metastatic activity of curcumin and catechin, Cancer Lett. **141**(1-2), 159–165 (1999). 49

[138] J. L. Pang, D. A. Ricupero, S. Huang, N. Fatma, D. P. Singh, J. R. Romero, and N. Chattopadhyay, Differential activity of kaempferol and quercetin in attenuating tumor necrosis factor receptor family signaling in bone cells, Biochem. Pharmacol. **71**(6), 818–826 (2006). 49

[139] G. Seelinger, I. Merfort, and C. M. Schempp, Anti-oxidant, anti-inflammatory and anti-allergic activities of luteolin, Planta Med. **74**(14), 1667–1677 (2008). 49

[140] D. Miller, Health benefits of lecithin and choline, Cereal foods world **47**(5), 178 (2002). 49

[141] R. Verpoorte, Y. Choi, and H. Kim, NMR-based metabolomics at work in phytochemistry, Phytochem. Rev. **6**(1), 3–14 (2007). 53, 132

[142] Y.-J. Zhang, T. Tanaka, Y. Iwamoto, C.-R. Yang, and I. Kouno, Phyllaemblic acid, a novel highly oxygenated norbisabolane from the roots of Phyllanthus emblica, Tetrahedron Lett. **41**(11), 1781–1784 (2000). 53

[143] L. Anila and N. Vijayalakshmi, Antioxidant action of flavonoids from Mangifera indica and Emblica officinalis in hypercholesterolemic rats, Food Chem. **83**(4), 569–574 (2003). 53

[144] X. Liu, M. Zhao, J. Wang, B. Yang, and Y. Jiang, Antioxidant activity of methanolic extract of emblica fruit (Phyllanthus emblica L.) from six regions in China, J. Food Compos. Anal. **21**(3), 219–228 (2008). 53

[145] L. Anila and N. Vijayalakshmi, Beneficial effects of flavonoids from Sesamum indicum, Emblica officinalis and Momordica charantia, Phytother Res. **14**(8), 592–595 (2000). 53

[146] S. Mirunalini and M. Krishnaveni, Therapeutic potential of Phyllanthus emblica (amla): the ayurvedic wonder, J. Basic Clin. Physiol. Pharmacol. **21**(1), 93–105 (2010). 53

[147] K. J. Abesundara, T. Matsui, and K. Matsumoto, α-Glucosidase inhibitory activity of some Sri Lanka plant extracts, one of which, Cassia auriculata, exerts a strong antihyperglycemic effect in rats comparable to the therapeutic drug acarbose, J. Agric. Food Chem. **52**(9), 2541–2545 (2004). 53

[148] E. Singh, S. Sharma, A. Pareek, J. Dwivedi, S. Yadav, and S. Sharma, Phytochemistry, traditional uses and cancer chemopreventive activity of Amla (Phyllanthus emblica): The Sustainer, J. Appl. Pharm. Sci. **02(01)**(Issue), 176–183 (2012). 53

[149] C. Deng, Q. Xie, and Z. QU, Bacteriostatic action of 4 ethanol extracts of the fruit of Phyllanthus emblica, China Pharm. (1991). 53

[150] Y.-J. Zhang, T. Nagao, T. Tanaka, C.-R. Yang, H. Okabe, and I. Kouno, Antiproliferative activity of the main constituents from Phyllanthus emblica, Biol. Pharm. Bull. **27**(2), 251–255 (2004). 53

[151] S. Panda and A. Kar, Fruit extract of Emblica officinalis ameliorates hyperthyroidism and hepatic lipid peroxidation in mice, Pharmazie **58**(10), 753–755 (2003). 53

[152] N. Barthakur and N. Arnold, Chemical analysis of the emblic (Phyllanthus emblica L.) and its potential as a food source, Sci. Hortic. **47**(1-2), 99–105 (1991). 54

[153] M. Rajalakshmi, J. Eliza, C. E. Priya, A. Nirmala, P. Daisy, et al., Anti-diabetic properties of Tinospora cordifolia stem extracts on streptozotocin-induced diabetic rats, Afr. J. Pharm. Pharmacol. **3**(5), 171–180 (2009). 54

[154] P. Baghel, Plant of versatile properties: A review of Tinospora Cordifolia (Guduchi), Int. J. Agric. Innov. Res. **5**(5), 2319–1473 (2017). 54

[155] L. Hussain, M. S. Akash, N.-U. Ain, K. Rehman, and M. Ibrahim, The analgesic, anti-inflammatory and anti-pyretic activities of Tinospora cordifolia, Adv. Clin. Exp. Med. **24**(6), 957–964 (2015). 54

[156] D. K. Verma, P. Kumar, M. El-Shazly, et al., Unmasking the many faces of giloy (Tinospora cordifolia L.): A fresh look on its phytochemical and medicinal properties, Curr. Pharm. Des. **27**(22), 2571–2581 (2021). 54

[157] R. K. Singh, Tinospora cordifolia as an adjuvant drug in the treatment of hyper-reactive malarious splenomegaly-case reports, J. Vector Borne Dis. **42**(1), 36 (2005). 54

[158] X. Mao, L.-F. Wu, H.-L. Guo, W.-J. Chen, Y.-P. Cui, Q. Qi, S. Li, W.-Y. Liang, G.-H. Yang, Y.-Y. Shao, et al., The genus Phyllanthus: an ethnopharmacological, phytochemical, and pharmacological review, Evid. Based Complementary Altern. Med. **2016** (2016). 66

[159] O. Fernandez, M. Urrutia, T. Berton, S. Bernillon, C. Deborde, D. Jacob, M. Maucourt, P. Maury, H. Duruflé, Y. Gibon, et al., Metabolomic characterization of sunflower leaf allows discriminating genotype groups or stress levels with a minimal set of metabolic markers, Metabolomics **15**(4), 1–14 (2019). 69, 73

[160] H. Mano and M. Hasebe, Rapid movements in plants, J. Plant Res. **134**(1), 3–17 (2021). 69

[161] J. P. Vandenbrink, E. A. Brown, S. L. Harmer, and B. K. Blackman, Turning heads: the biology of solar tracking in sunflower, Plant Sci. **224**, 20–26 (2014). 69

[162] U. Kutschera and W. R. Briggs, Phototropic solar tracking in sunflower plants: an integrative perspective, Ann. Bot. **117**(1), 1–8 (2016). 69

[163] W. R. Briggs, How do sunflowers follow the Sun—and to what end?, Science **353**(6299), 541–542 (2016). 69

[164] H. S. Atamian, N. M. Creux, E. A. Brown, A. G. Garner, B. K. Blackman, and S. L. Harmer, Circadian regulation of sunflower heliotropism, floral orientation, and pollinator visits, Science **353**(6299), 587–590 (2016). 69, 86

[165] A. Nardini, S. Salleo, and S. Andri, Circadian regulation of leaf hydraulic conductance in sunflower (Helianthus annuus L. cv Margot), Plant Cell Environ. **28**(6), 750–759 (2005). 69

[166] T. J. Knight and G. S. Weissman, Rhythms in glutamine synthetase activity, energy charge, and glutamine in sunflower roots, Plant Physiol. **70**(6), 1683–1688 (1982). 69

[167] F. Cellier, G. Conéjéro, and F. Casse, Dehydrin transcript fluctuations during a day/night cycle in drought-stressed sunflower, J. Exp. Bot. **51**(343), 299–304 (2000). 69

[168] M. Fambrini, E. Bonsignori, F. Rapparini, G. Cionini, V. Michelotti, D. Bertini, R. Baraldi, and C. Pugliesi, stem fasciated, a recessive mutation in sunflower (Helianthus annuus), alters plant morphology and auxin level, Ann. Bot. **98**(4), 715–730 (2006). 69

[169] G. Lobello, M. Fambrini, R. Baraldi, B. Lercari, and C. Pugliesi, Hormonal influence on photocontrol of the protandry in the genus Helianthus, J. Exp. Bot. **51**(349), 1403–1412 (2000). 69

[170] A. S. van Wyk and G. Prinsloo, Challenging current interpretation of sunflower movements, J. Exp. Bot. **70**(21), 6049–6056 (2019). 69

[171] J. C. Dunlap, Molecular bases for circadian clocks, Cell **96**(2), 271–290 (1999). 69

[172] F. Halberg, Implications of biologic rhythms for clinical practice, Hosp. Pract. **12**(1), 139–149 (1977). 70

[173] C. R. McClung, Plant circadian rhythms, The Plant Cell **18**(4), 792–803 (2006). 70

[174] M. E. Eriksson and A. J. Millar, The circadian clock. A plant's best friend in a spinning world, Plant Physiol. **132**(2), 732–738 (2003). 70

[175] T. L. Rost, M. G. Barbour, R. M. Thornton, T. E. Weier, and C. R. Stocking, Botany; a brief introduction to plant biology, (1979). 70

[176] J. A. Ozga, H. Kaur, R. P. Savada, and D. M. Reinecke, Hormonal regulation of reproductive growth under normal and heat-stress conditions in legume and other model crop species, J. Exp. Bot. **68**(8), 1885–1894 (2017). 70

[177] H. K. Kim, Y. H. Choi, and R. Verpoorte, NMR-based metabolomic analysis of plants, Nat. Protoc. **5**(3), 536–549 (2010). 72

[178] S. Mishra, N. Gogna, and K. Dorai, NMR-based investigation of the altered metabolic response of Bougainvillea spectabilis leaves exposed to air pollution stress during the circadian cycle, Environ. Exp. Bot. **164**, 58–70 (2019). 72, 134

[179] J. L. Griffin, L. A. Walker, R. F. Shore, and J. K. Nicholson, Metabolic profiling of chronic cadmium exposure in the rat, Chem. Res. Toxicol. **14**(10), 1428–1434 (2001). 72

[180] S. Mishra, Ankit, R. Sharma, N. Gogna, and K. Dorai, NMR-based metabolomic profiling of the differential concentration of phytomedicinal compounds in pericarp, skin and seeds of Momordica charantia (bitter melon), Nat. Prod. Res. **36**(1), 390–395 (2021). 73

[181] N. Gogna, N. Hamid, and K. Dorai, Metabolomic profiling of the phytomedicinal constituents of Carica papaya L. leaves and seeds by 1H NMR spectroscopy and multivariate statistical analysis, J. Pharm. Biomed. **115**, 74–85 (2015). 73

[182] D. Tulpan, S. Léger, L. Belliveau, A. Culf, and M. Čuperlović-Culf, MetaboHunter: an automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures, BMC Bioinform. **12**(1), 1–22 (2011). 74, 100, 146

[183] J. Xia, T. C. Bjorndahl, P. Tang, and D. S. Wishart, MetaboMiner–semi-automated identification of metabolites from 2D NMR spectra of complex biofluids, BMC Bioinform. **9**(1), 1–16 (2008). 74, 100, 146

[184] Z. Pang, J. Chong, G. Zhou, D. A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, and J. Xia, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic acids Res. **49**(W1), W388–W396 (2021). 74, 136

[185] A. Liaw, M. Wiener, et al., Classification and regression by randomForest, R news **2**(3), 18–22 (2002). 74, 90

[186] T. Chen, Y. Cao, Y. Zhang, J. Liu, Y. Bao, C. Wang, W. Jia, and A. Zhao, Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection, Evid.-Based Complementary Altern. Med. **2013** (2013). 74

[187] M. Mehrpour, A. Kyani, M. Tafazzoli, F. Fathi, and M.-T. Joghataie, A metabonomics investigation of multiple sclerosis by nuclear magnetic resonance, Magn. Reson. Chem. **51**(2), 102–109 (2013). 74

[188] F. Fathi, L. Majari-Kasmaee, A. Mani-Varnosfaderani, A. Kyani, M. Rostami-Nejad, K. Sohrabzadeh, N. Naderi, M. R. Zali, M. Rezaei-Tavirani, M. Tafazzoli, et al., 1H NMR based metabolic profiling in Crohn's disease by random forest methodology, Magn. Reson. Chem. **52**(7), 370–376 (2014). 74

[189] T. Barri, J. Holmer-Jensen, K. Hermansen, and L. O. Dragsted, Metabolic fingerprinting of high-fat plasma samples processed by centrifugation-and

filtration-based protein precipitation delineates significant differences in metabolite information coverage, Anal. Chim. Acta. **718**, 47–57 (2012). 75

[190] S. L. Harmer, J. B. Hogenesch, M. Straume, H.-S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, and S. A. Kay, Orchestrated transcription of key pathways in Arabidopsis by the circadian clock, Science **290**(5499), 2110–2113 (2000). 75

[191] A. A. Webb, The physiology of circadian rhythms in plants, New Phytol. **160**(2), 281–303 (2003). 75

[192] A. N. Dodd, N. Salathia, A. Hall, E. Kévei, R. Tóth, F. Nagy, J. M. Hibberd, A. J. Millar, and A. A. Webb, Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage, Science **309**(5734), 630–633 (2005). 75, 86, 96

[193] E. Bünning, The physiological clock: endogenous diurnal rhythms and biological chronometry, Springer, 2013. 75

[194] Z. Ramadan, D. Jacobs, M. Grigorov, and S. Kochhar, Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms, Talanta **68**(5), 1683–1691 (2006). 82

[195] J. Bartel, J. Krumsiek, and F. J. Theis, Statistical methods for the analysis of high-throughput metabolomics data, Comput. Struct. Biotechnol. J. **4**(5), e201301009 (2013). 82

[196] M. Bylesjö, Extracting meaningful information from metabonomic data using multivariate statistics, in Metabonomics, pages 137–146, Springer, 2015. 82

[197] A. Alonso, S. Marsal, and A. Julià, Analytical methods in untargeted metabolomics: state of the art in 2015, Front. Bioeng. Biotechnol. **3**, 23 (2015). 82

[198] A. Samach and G. Coupland, Time measurement and the control of flowering in plants, Bioessays **22**(1), 38–47 (2000). 86

[199] S. Barak, E. M. Tobin, R. M. Green, C. Andronis, and S. Sugano, All in good time: the Arabidopsis circadian clock, Trends Plant Sci. **5**(12), 517–522 (2000). 86

[200] J. A. Kreps and S. A. Kay, Coordination of plant metabolism and development by the circadian clock., The Plant Cell **9**(7), 1235 (1997). 86

[201] R. Amir, G. Galili, and H. Cohen, The metabolic roles of free amino acids during seed development, Plant Science **275**, 11–18 (2018). 86

[202] Q.-Q. Yang, D.-S. Zhao, C.-Q. Zhang, H.-Y. Wu, Q.-F. Li, M.-H. Gu, S. S.-M. Sun, and Q.-Q. Liu, A connection between lysine and serotonin metabolism in rice endosperm, Plant physiology **176**(3), 1965–1980 (2018). 86

[203] J. Sheen, L. Zhou, and J.-C. Jang, Sugars as signaling molecules, Curr. Opin. Plant Biol. **2**(5), 410–418 (1999). 86

[204] O. E. Bläsing, Y. Gibon, M. Günther, M. Höhne, R. Morcuende, D. Osuna, O. Thimm, B. Usadel, W.-R. Scheible, and M. Stitt, Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis, The Plant Cell **17**(12), 3257–3281 (2005). 86

[205] Y. Gibon, B. Usadel, O. E. Blaesing, B. Kamlage, M. Hoehne, R. Trethewey, and M. Stitt, Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsisrosettes, Genome Biol. **7**(8), 1–23 (2006). 86

[206] M. J. Haydon, T. J. Hearn, L. J. Bell, M. A. Hannah, and A. A. Webb, Metabolic regulation of circadian clocks, in Semin. Cell Dev. Biol., volume 24, pages 414–421, Elsevier, 2013. 86

[207] G. Chen, Y. Chi, X. Wu, J. Duan, and N. Li, Chemical oxidation of p-hydroxyphenylpyruvic acid in aqueous solution by capillary electrophoresis with an electrochemiluminescence detection system, Anal. Chem. **75**(23), 6602–6607 (2003). 87

[208] Y. Huang, X. Zhang, L. Xu, H. Chen, and G. Chen, Characterization of keto-enol tautomerism of p-hydroxyphenylpyruvic acid using CE with amperometric detection and spectrometric analysis, J. Sep. Sci. **32**(23-24), 4155–4160 (2009). 87

[209] S. H. Zeisel, Dietary choline: biochemistry, physiology, and pharmacology, Annu. Rev. Nutr. **1**(1), 95–121 (1981). 87

[210] S. Maatta, B. Scheu, M. R. Roth, P. Tamura, M. Li, T. D. Williams, X. Wang, and R. Welti, Levels of Arabidopsis thaliana leaf phosphatidic acids, phosphatidylserines, and most trienoate-containing polar lipid molecular species increase during the dark period of the diurnal cycle, Front. Plant Sci. **3**, 49 (2012). 87, 119

[211] A. Sakamoto and N. Murata, The role of glycine betaine in the protection of plants from stress: clues from transgenic plants, Plant Cell Environ. **25**(2), 163–171 (2002). 87

[212] J. Huang, K. Rozwadowski, V. Bhinu, U. Schäfer, and A. Hannoufa, Manipulation of sinapine, choline and betaine accumulation in Arabidopsis seed: towards improving the nutritional value of the meal and enhancing the seedling performance under environmental stresses in oilseed crops, Plant Physiol. Biochem. **46**(7), 647–654 (2008). 87

[213] D. K. Stumpf and R. Burris, Organic acid contents of soybean: age and source of nitrogen, Plant Physiology **68**(5), 989–991 (1981). 87

[214] D. W. Chia, T. J. Yoder, W.-D. Reiter, and S. I. Gibson, Fumaric acid: an overlooked form of fixed carbon in Arabidopsis and other plant species, Planta **211**(5), 743–751 (2000). 87, 118

[215] M. Cercós, G. Soler, D. J. Iglesias, J. Gadea, J. Forment, and M. Talón, Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric acid utilization, Plant Mol. Biol. **62**(4), 513–527 (2006). 87

[216] A. Degu, B. Hatew, A. Nunes-Nesi, L. Shlizerman, N. Zur, E. Katz, A. R. Fernie, E. Blumwald, and A. Sadka, Inhibition of aconitase in citrus fruit callus results in a metabolic shift towards amino acid biosynthesis, Planta **234**(3), 501–513 (2011). 87

[217] N. Bouche and H. Fromm, GABA in plants: just a metabolite?, Trends Plant Sci. **9**(3), 110–115 (2004). 87

[218] S. Michaeli and H. Fromm, Closing the loop on the GABA shunt in plants: are GABA metabolism and signaling entwined?, Front. Plant Sci. **6**, 419 (2015). 87

[219] T. A. Smith, Tryptamine and related compounds in plants, Phytochemistry **16**(2), 171–175 (1977). 88

[220] H. Maeda and N. Dudareva, The shikimate pathway and aromatic amino acid biosynthesis in plants, Annu. Rev. Plant Biol. **63**, 73–105 (2012). 88

[221] S. Murch, S. KrishnaRaj, and P. Saxena, Tryptophan is a precursor for melatonin and serotonin biosynthesis in in vitro regenerated St. John's wort (Hypericum perforatum L. cv. Anthos) plants, Plant Cell Rep. **19**(7), 698–704 (2000). 89

[222] A. W. Woodward and B. Bartel, Auxin: regulation, action, and interaction, Ann. Bot. **95**(5), 707–735 (2005). 89

[223] J. E. Sherwin and W. K. Purves, Tryptophan as an auxin precursor in cucumber seedlings, Plant Physiol. **44**(9), 1303–1309 (1969). 89

[224] S. E. O'Connor and J. J. Maresh, Chemistry and biology of monoterpene indole alkaloid biosynthesis, Nat. Prod. Rep. **23**(4), 532–547 (2006). 89

[225] Y. Kokubo, M. Nishizaka, N. Ube, Y. Yabuta, S.-i. Tebayashi, K. Ueno, S. Taketa, and A. Ishihara, Distribution of the tryptophan pathway-derived defensive secondary metabolites gramine and benzoxazinones in Poaceae, Biosci. Biotechnol. Biochem. **81**(3), 431–440 (2017). 89

[226] H. Stotz, P. Brown, and J. Tokuhisa, Glucosinolate biosynthesis from amino acids., in Amino acids in higher plants, pages 436–447, CAB International Wallingford UK, 2015. 89

[227] B. Bartel, Auxin biosynthesis, Annu. Rev. Plant Biol. **48**(1), 51–66 (1997). 92

[228] Y. Zhao, Auxin biosynthesis and its role in plant development, Annu. Rev. Plant Biol. **61**, 49 (2010). 92

[229] Y. Mano and K. Nemoto, The pathway of auxin biosynthesis in plants, J. Exp. Bot. **63**(8), 2853–2872 (2012). 92

[230] G. Taylor Jr, Plant and leaf resistance to gaseous air pollution stress, New Phytol. **80**(3), 523–534 (1978). 95

[231] E.-L. Viskari, J. Surakka, P. Pasanen, A. Mirme, S. Kössi, J. Ruuskanen, and J. Holopainen, Responses of spruce seedlings (Picea abies) to exhaust gas under laboratory conditions—I plant–insect interactions, Environ. Pollut. **107**(1), 89–98 (2000). 95

[232] P. Saxena and U. C. Kulshrestha, The impact of gasoline emission on plants–a review, Chem Ecol. **32**(4), 378–405 (2016). 95

[233] X. Wei, S. Lyu, Y. Yu, Z. Wang, H. Liu, D. Pan, and J. Chen, Phylloremediation of air pollutants: exploiting the potential of plant leaves and leaf-associated microbes, Front. Plant Sci. **8**, 1318 (2017). 95

[234] C. Caldana, T. Degenkolbe, A. Cuadros-Inostroza, S. Klie, R. Sulpice, A. Leisse, D. Steinhauser, A. R. Fernie, L. Willmitzer, and M. A. Hannah, High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions, Plant J. **67**(5), 869–884 (2011). 95

[235] J. Bowne, A. Bacic, M. Tester, and U. Roessner, Abiotic stress and metabolomics, Annu Plant Rev **43**, 61–85 (2011). 95

[236] R. Kooke and J. Keurentjes, Multi-dimensional regulation of metabolic networks shaping plant development and performance, J. Exp. Bot. **63**(9), 3353–3365 (2012). 95

[237] K. Urano, Y. Kurihara, M. Seki, and K. Shinozaki, 'Omics' analyses of regulatory networks in plant abiotic stress responses, Curr. Opin. Plant Biol. **13**(2), 132–138 (2010). 95

[238] M. Bastawrous, A. Jenne, M. Tabatabaei Anaraki, and A. J. Simpson, In-vivo NMR spectroscopy: A powerful and complimentary tool for understanding environmental toxicity, Metabolites **8**(2), 35 (2018). 95

[239] S. S. Gill and N. Tuteja, Polyamines and abiotic stress tolerance in plants, Plant Signal. Behav. **5**(1), 26–33 (2010). 95

[240] C. Deborde, A. Moing, L. Roch, D. Jacob, D. Rolin, and P. Giraudeau, Plant metabolism as studied by NMR spectroscopy, Prog. Nucl. Magn. Reson. Spectrosc. **102**, 61–97 (2017). 96

[241] J. Kikuchi, K. Ito, and Y. Date, Environmental metabolomics with data science for investigating ecosystem homeostasis, Prog. Nucl. Magn. Reson. Spectrosc. **104**, 56–88 (2018). 96

[242] S. Sahar and P. Sassone-Corsi, Regulation of metabolism: the circadian clock dictates the time, Trends Endocrinol. Metab. **23**(1), 1–8 (2012). 96

[243] A. Sanchez, J. Shin, and S. J. Davis, Abiotic stress and the plant circadian clock, Plant Signal. Behav. **6**(2), 223–231 (2011). 96

[244] D. Augustijn, U. Roy, R. Van Schadewijk, H. De Groot, and A. Alia, Metabolic profiling of intact Arabidopsis thaliana leaves during circadian cycle using 1H high resolution magic angle spinning NMR, PLoS One **11**(9), e0163258 (2016). 96

[245] J. Krasensky and C. Jonak, Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks, J. Exp. Bot. **63**(4), 1593–1608 (2012). 96

[246] P. E. Verslues, M. Agarwal, S. Katiyar-Agarwal, J. Zhu, and J.-K. Zhu, Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status, Plant J. **45**(4), 523–539 (2006). 96

[247] W. E. Winner and C. J. Atkinson, Absorption of air pollution by plants, and consequences for growth, Trends Ecol. Evol. **1**(1), 15–18 (1986). 96

[248] S. Singh, D. Rao, M. Agrawal, J. Pandey, and D. Naryan, Air pollution tolerance index of plants, J. Environ. Manage. **32**(1), 45–55 (1991). 96

[249] P. S. Lakshmi, K. L. Sravanti, and N. Srinivas, Air pollution tolerance index of various plant species growing in industrial areas, The Ecoscan **2**(2), 203–206 (2009). 96

[250] P. Saxena and C. Ghosh, Ornamental plants as sinks and bioindicators, Environ. Technol. **34**(23), 3059–3067 (2013). 96

[251] A. Verma and S. Singh, Biochemical and ultrastructural changes in plant foliage exposed to auto-pollution, Environ. Monit. Assess. **120**(1), 585–602 (2006). 96

[252] L. Eriksson, J. Trygg, and S. Wold, CV-ANOVA for significance testing of PLS and OPLS® models, J. Chemom. **22**(11-12), 594–600 (2008). 101, 147

[253] N. Gogna, V. J. Singh, V. Sheeba, and K. Dorai, NMR-based investigation of the Drosophila melanogaster metabolome under the influence of daily cycles of light and temperature, Mol. Biosyst. **11**(12), 3305–3315 (2015). 101

[254] N. Gogna, R. Sharma, V. Gupta, K. Dorai, and N. Prasad, Evolution of the metabolome in response to selection for increased immunity in populations of Drosophila melanogaster, PloS one **12**(11), e0188089 (2017). 101, 144, 145, 146, 147, 148, 150

[255] R. Sharma, N. Gogna, H. Singh, and K. Dorai, Fast profiling of metabolite mixtures using chemometric analysis of a speeded-up 2D heteronuclear correlation NMR experiment, RSC Adv. **7**(47), 29860–29870 (2017). 101

[256] Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc., B: Stat. Methodol. **57**(1), 289–300 (1995). 102

[257] O. E. Bläsing, Y. Gibon, M. Günther, M. Höhne, R. Morcuende, D. Osuna, O. Thimm, B. Usadel, W.-R. Scheible, and M. Stitt, Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis, Plant Cell **17**(12), 3257–3281 (2005). 118

[258] B. G. Forde and P. J. Lea, Glutamate in plants: metabolism, regulation, and signalling, J. Exp. Bot. **58**(9), 2339–2358 (2007). 118

[259] S. Michaeli, A. Fait, K. Lagor, A. Nunes-Nesi, N. Grillich, A. Yellin, D. Bar, M. Khan, A. R. Fernie, F. J. Turano, et al., A mitochondrial GABA permease connects the GABA shunt and the TCA cycle, and is essential for normal carbon metabolism, Plant J. **67**(3), 485–498 (2011). 118

[260] M. Groppa and M. Benavides, Polyamines and abiotic stress: recent advances, Amino acids **34**(1), 35–45 (2008). 119

[261] A. F. Page, S. Mohapatra, R. Minocha, and S. C. Minocha, The effects of genetic manipulation of putrescine biosynthesis on transcription and activities of the other polyamine biosynthetic enzymes 1, Physiol. Plant. **129**(4), 707–724 (2007). 119

[262] J. C. Cuevas, R. López-Cobollo, R. Alcázar, X. Zarza, C. Koncz, T. Altabella, J. Salinas, A. F. Tiburcio, and A. Ferrando, Putrescine is involved in Arabidopsis freezing tolerance and cold acclimation by regulating abscisic acid levels in response to low temperature, Plant Physiol. **148**(2), 1094–1105 (2008). 119

[263] P. V. Minorsky, The hot and the classic: Trigonelline: A diverse regulator in plants, Plant Physiol. **128**(1), 7 (2002). 119

[264] M. A. Farag, A. Porzel, and L. A. Wessjohann, Unraveling the active hypoglycemic agent trigonelline in Balanites aegyptiaca date fruit using metabolite fingerprinting by NMR, J. Pharm. Biomed. Anal. **115**, 383–387 (2015). 119

[265] L. R. Rajasekaran, D. Aspinall, G. Jones, and L. Paleg, Stress metabolism. IX. Effect of salt stress on trigonelline accumulation in tomato, Can. J. Plant Sci. **81**(3), 487–498 (2001). 119

[266] M. Ueda, M. Niwa, and S. Yamamura, Trigonelline, a leaf-closing factor of the nyctinastic plant, Aeschynomene indica, Phytochemistry **39**(4), 817–819 (1995). 119

[267] S. Y. Woo and S. M. Je, Photosynthetic rates and antioxidant enzyme activity ofPlatanus occidentalis growing under two levels of air pollution along the streets of Seoul, J. Plant Biol. **49**(4), 315–319 (2006). 126

[268] M. Kusano, T. Tohge, A. Fukushima, M. Kobayashi, N. Hayashi, H. Otsuki, Y. Kondou, H. Goto, M. Kawashima, F. Matsuda, et al., Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light, Plant J. **67**(2), 354–369 (2011). 126

[269] G. Kalbin, A. B. Ohlsson, T. Berglund, J. Rydström, and Å. Strid, Ultraviolet-B-radiation-induced changes in nicotinamide and glutathione metabolism and gene expression in plants, Eur. J. Biochem. **249**(2), 465–472 (1997). 126

[270] J.-H. Liu, H. Kitashiba, J. Wang, Y. Ban, and T. Moriguchi, Polyamines and their ability to provide environmental stress tolerance to plants, Plant Biotechnol. **24**(1), 117–126 (2007). 126

[271] I. Zeid and Z. Shedeed, Response of alfalfa to putrescine treatment under drought stress, Biol. Plant. **50**(4), 635–640 (2006). 126

[272] D. Kley, M. Kleinmann, H. Sanderman, and S. Krupa, Photochemical oxidants: state of the science, Environ. Pollut. **100**(1-3), 19–42 (1999). 126

[273] K. Urano, K. Maruyama, Y. Ogata, Y. Morishita, M. Takeda, N. Sakurai, H. Suzuki, K. Saito, D. Shibata, M. Kobayashi, et al., Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics, Plant J. **57**(6), 1065–1078 (2009). 126

[274] F. Kaplan, J. Kopka, D. Y. Sung, W. Zhao, M. Popp, R. Porat, and C. L. Guy, Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content, Plant J. **50**(6), 967–981 (2007). 127

[275] A. M. Kinnersley and F. J. Turano, Gamma aminobutyric acid (GABA) and plant responses to stress, Crit. Rev. Plant Sci. **19**(6), 479–509 (2000). 127

[276] V. Joshi, J.-G. Joung, Z. Fei, and G. Jander, Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress, Amino acids **39**(4), 933–947 (2010). 127

[277] S. Seyyednejad, M. Niknejad, and H. Koochak, A review of some different effects of air pollution on plants, Res. J. Environ. Sci. **5**(4), 302 (2011). 127

[278] I. Kumar and R. K. Sharma, Production of secondary metabolites in plants under abiotic stress: an overview, Significances bioeng. biosci. **2**(4), 196–200 (2018). 127

[279] D. K. Chandawat, P. Verma, H. Solanki, and Y. Patel, Role of total phenol in the resistance mechanism of plants against air pollution, Biolife **2**(2), 586–592 (2014). 128

[280] M. L. Orozco-Cárdenas, J. Narváez-Vásquez, and C. A. Ryan, Hydrogen peroxide acts as a second messenger for the induction of defense genes in tomato plants in response to wounding, systemin, and methyl jasmonate, Plant Cell **13**(1), 179–191 (2001). 131

[281] J. León, E. Rojo, and J. J. Sánchez-Serrano, Wound signalling in plants, J. Exp. Bot. **52**(354), 1–9 (2001). 131

[282] C. A. Ryan and D. S. Moura, Systemic wound signaling in plants: a new perception, Proc. Natl. Acad. Sci. **99**(10), 6519–6520 (2002). 131

[283] S. G. Ralph, H. Yueh, M. Friedmann, D. Aeschliman, J. A. Zeznik, C. C. Nelson, Y. S. Butterfield, R. Kirkpatrick, J. Liu, S. J. Jones, et al., Conifer defence against insects: microarray gene expression profiling of Sitka spruce (Picea sitchensis) induced by mechanical wounding or feeding by spruce budworms (Choristoneura occidentalis) or white pine weevils (Pissodes strobi) reveals large-scale changes of the host transcriptome, Plant Cell Environ. **29**(8), 1545–1570 (2006). 131

[284] E. Grata, J. Boccard, G. Glauser, P.-A. Carrupt, E. E. Farmer, J.-L. Wolfender, and S. Rudaz, Development of a two-step screening ESI-TOF-MS method for rapid determination of significant stress-induced metabolome modifications in plant leaf extracts: The wound response in Arabidopsis thaliana as a case study, J. Sep. Sci. **30**(14), 2268–2278 (2007). 131

[285] E. Grata, J. Boccard, D. Guillarme, G. Glauser, P.-A. Carrupt, E. E. Farmer, J.-L. Wolfender, and S. Rudaz, UPLC–TOF-MS for plant metabolomics: a sequential approach for wound marker analysis in Arabidopsis thaliana, J. Chromatogr. B **871**(2), 261–270 (2008). 131

[286] G. Glauser, E. Grata, L. Dubugnon, S. Rudaz, E. E. Farmer, and J.-L. Wolfender, Spatial and temporal dynamics of jasmonate synthesis and accumulation in Arabidopsis in response to wounding, J. Biol. Chem. **283**(24), 16400–16407 (2008). 131

[287] G. Glauser, J. Boccard, S. Rudaz, and J.-L. Wolfender, Mass spectrometry-based metabolomics oriented by correlation analysis for wound-induced molecule discovery: identification of a novel jasmonate glucoside, Phytochem Anal. **21**(1), 95–101 (2010). 131

[288] A. Thiocone, E. E. Farmer, and J.-L. Wolfender, Screening for wound-induced oxylipins in Arabidopsis thaliana by differential HPLC-APCI/MS profiling of

crude leaf extracts and subsequent characterisation by capillary-scale NMR, Phytochem Anal. **19**(3), 198–205 (2008). 131

[289] G. A. Howe and A. Schaller, Direct defenses in plants and their induction by wounding and insect herbivores, in Induced plant resistance to herbivory, pages 7–29, Springer, 2008. 131

[290] G. L. De Bruxelles and M. R. Roberts, Signals regulating multiple responses to wounding and herbivores, CRC Crit. Rev. Plant Sci. **20**(5), 487–521 (2001). 131

[291] J. Wu and I. T. Baldwin, Herbivory-induced signalling in plants: perception and action, Plant Cell Environ. **32**(9), 1161–1174 (2009). 131

[292] M. O. Mello and M. C. Silva-Filho, Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms, Braz. J. Plant Physiol. **14**, 71–81 (2002). 131

[293] M. E. Maffei, A. Mithöfer, and W. Boland, Before gene expression: early events in plant–insect interaction, Trends Plant Sci. **12**(7), 310–316 (2007). 131

[294] H. U. Stotz, J. Kroymann, and T. Mitchell-Olds, Plant-insect interactions, Curr. Opin. Plant Biol. **2**(4), 268–272 (1999). 131

[295] N. Ferry, M. G. Edwards, J. A. Gatehouse, and A. M. Gatehouse, Plant–insect interactions: molecular approaches to insect resistance, Curr. Opin. Plant Biol. **15**(2), 155–161 (2004). 131

[296] D. Maag, M. Erb, and G. Glauser, Metabolomics in plant–herbivore interactions: challenges and applications, Entomol. Exp. Appl. **157**(1), 18–29 (2015). 131

[297] M. Macel, N. M. Van Dam, and J. J. Keurentjes, Metabolomics: the chemistry between ecology and genetics, Mol. Ecol. Resour. **10**(4), 583–593 (2010). 131

[298] H. T. Widarto, E. Van Der Meijden, A. W. Lefeber, C. Erkelens, H. K. Kim, Y. H. Choi, and R. Verpoorte, Metabolomic differentiation of Brassica rapa following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy, J. Chem. Ecol. **32**(11), 2417–2428 (2006). 131

[299] M. Mazid, T. Khan, F. Mohammad, et al., Role of secondary metabolites in defense mechanisms of plants., Biol. Med. **3**(2), 232–249 (2011). 131

[300] I. P. De León and M. Montesano, Activation of defense mechanisms against pathogens in mosses and flowering plants, Int. J. Mol. Sci. **14**(2), 3178–3200 (2013). 131

[301] T. Belete et al., Defense mechanisms of plants to insect pests: from morphological to biochemical approach, Trends Tech. Sci. Res **2**, 30–38 (2018). 131

[302] W.-L. Yang and M. A. Bernards, Metabolite profiling of potato (Solanum tuberosum L.) tubers during wound-induced suberization, Metabolomics **3**(2), 147–159 (2007). 131

[303] K. A. Leiss, Y. H. Choi, I. B. Abdel-Farid, R. Verpoorte, and P. G. Klinkhamer, NMR metabolomics of thrips (Frankliniella occidentalis) resistance in Senecio hybrids, J. Chem. Ecol. **35**(2), 219–229 (2009). 132

[304] K. A. Leiss, F. Maltese, Y. H. Choi, R. Verpoorte, and P. G. Klinkhamer, Identification of chlorogenic acid as a resistance factor for thrips in chrysanthemum, Plant Physiol. **150**(3), 1567–1575 (2009). 132, 140

[305] A. Moing, M. Maucourt, C. Renaud, M. Gaudillère, R. Brouquisse, B. Lebouteiller, A. Gousset-Dupont, J. Vidal, D. Granot, B. Denoyes-Rothan, et al., Quantitative metabolic profiling by 1-dimensional 1H-NMR analyses: application to plant genetics and functional genomics, Funct. Plant Biol. **31**(9), 889–902 (2004). 134

[306] A. Rivas-Ubach, M. Pérez-Trujillo, J. Sardans, A. Gargallo-Garriga, T. Parella, and J. Penuelas, Ecometabolomics: optimized NMR-based method, Methods Ecol. Evol. **4**(5), 464–473 (2013). 134

[307] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, et al., HMDB: the human metabolome database, Nucleic Acids Res. **35**(suppl_1), D521–D526 (2007). 136

[308] K. Akiyama, E. Chikayama, H. Yuasa, Y. Shimada, T. Tohge, K. Shinozaki, M. Y. Hirai, T. Sakurai, J. Kikuchi, and K. Saito, PRIMe: a Web site that assembles tools for metabolomics and transcriptomics, In Silico Biol. **8**(3-4), 339–345 (2008). 136

[309] R. Ehness, M. Ecker, D. E. Godt, and T. Roitsch, Glucose and stress independently regulate source and sink metabolism and defense mechanisms via signal transduction pathways involving protein phosphorylation., Plant Cell **9**(10), 1825–1841 (1997). 140

[310] J. J. Koskimäki, J. Hokkanen, L. Jaakola, M. Suorsa, A. Tolonen, S. Mattila, A. M. Pirttilä, and A. Hohtola, Flavonoid biosynthesis and degradation play a role in early defence responses of bilberry (Vaccinium myrtillus) against biotic stress, Eur. J. Plant Pathol. **125**(4), 629–640 (2009). 140

[311] M. A. Adams, A. Richter, A. K. Hill, and T. D. Colmer, Salt tolerance in Eucalyptus spp.: identity and response of putative osmolytes, Plant Cell Environ. **28**(6), 772–787 (2005). 140

[312] S. K. Arndt, S. J. Livesley, A. Merchant, T. M. Bleby, and P. F. Grierson, Quercitol and osmotic adaptation of field-grown Eucalyptus under seasonal drought stress, Plant Cell Environ. **31**(7), 915–924 (2008). 140

[313] A. Merchant, S. K. Arndt, A. Callister, and M. A. Adams, Quercitol plays a key role in stress tolerance of Eucalyptus leptophylla (F. Muell) in naturally occurring saline conditions, Environ. Exp. Bot. **65**(2-3), 296–303 (2009). 140

[314] S. D. McNeil, M. L. Nuccio, M. J. Ziemak, and A. D. Hanson, Enhanced synthesis of choline and glycine betaine in transgenic tobacco plants that overexpress phosphoethanolamine N-methyltransferase, Proc. Natl. Acad. Sci. **98**(17), 10001–10005 (2001). 140

[315] H.-M. Lam, K. Coschigano, C. Schultz, R. Melo-Oliveira, G. Tjaden, I. Oliveira, N. Ngai, M.-H. Hsieh, and G. Coruzzi, Use of Arabidopsis mutants and genes to study amide amino acid biosynthesis., Plant Cell **7**(7), 887 (1995). 140

[316] J. E. Cox, C. S. Thummel, and J. M. Tennessen, Metabolomic Studies in Drosophila, Genetics **206**(3), 1169–1185 (07 2017). 143

[317] L. Coquin, J. D. Feala, A. D. McCulloch, and G. Paternostro, Metabolomic and flux-balance analysis of age-related decline of hypoxia tolerance in Drosophila muscle tissue, Mol. Syst. Biol. **4**(1), 233 (2008). 143

[318] K. S. Pedersen, T. N. Kristensen, V. Loeschcke, B. O. Petersen, J. o. Duus, N. C. Nielsen, and A. Malmendal, Metabolomic Signatures of Inbreeding at Benign and Stressful Temperatures in Drosophila melanogaster, Genetics **180**(2), 1233–1243 (10 2008). 143

[319] J. D. Feala, L. Coquin, D. Zhou, G. G. Haddad, G. Paternostro, and A. D. McCulloch, Metabolism as means for hypoxia adaptation: metabolic profiling and flux balance analysis, BMC Syst. Biol. **3**(1), 91 (Sep 2009). 143

[320] A. Malmendal, J. G. Sørensen, J. Overgaard, M. Holmstrup, N. C. Nielsen, and V. Loeschcke, Metabolomic analysis of the selection response of Drosophila melanogaster to environmental stress: are there links to gene expression and phenotypic traits?, Naturwissenschaften **100**(5), 417–427 (May 2013). 143

[321] C. M. Williams, M. Watanabe, M. R. Guarracino, M. B. Ferraro, A. S. Edison, T. J. Morgan, A. F. B. Boroujerdi, and D. A. Hahn, Cold adaptation shapes the robustness of metabolic networks in Drosophila melanogaster, Evolution **68**(12), 3505–3523 (2014). 143, 146

[322] S. Ott, A. Vishnivetskaya, A. Malmendal, and D. C. Crowther, Metabolic changes may precede proteostatic dysfunction in a Drosophila model of amyloid beta peptide toxicity, Neurobiol. Aging **41**, 39–52 (2016). 143

[323] V. Bakalov, R. Amathieu, M. N. Triba, M.-J. Clément, L. Reyes Uribe, L. Le Moyec, and A. M. Kaynar, Metabolomics with Nuclear Magnetic Resonance Spectroscopy in a Drosophila melanogaster Model of Surviving Sepsis, Metabolites **6**(4), 47 (2016). 144

[324] M. F. Schou, T. N. Kristensen, A. Pedersen, B. G. Karlsson, V. Loeschcke, and A. Malmendal, Metabolic and functional characterization of effects of developmental temperature in Drosophila melanogaster, Am. J. Physiol. Regul. Integr. Comp. Physiol. **312**(2), R211–R222 (2017), PMID: 27927623. 144

[325] P. Sarup, S. M. M. Petersen, N. C. Nielsen, V. Loeschcke, and A. Malmendal, Mild heat treatments induce long-term changes in metabolites associated with energy metabolism in Drosophila melanogaster, Biogerontology **17**(5), 873–882 (Nov 2016). 144

[326] T. M. Felix, K. A. Hughes, E. A. Stone, J. M. Drnevich, and J. Leips, Age-Specific Variation in Immune Response in Drosophila melanogaster Has a Genetic Basis, Genetics **191**(3), 989–1002 (07 2012). 144

[327] M. Zerofsky, E. Harel, N. Silverman, and M. Tatar, Aging of the innate immune response in Drosophila melanogaster, Aging Cell **4**(2), 103–108 (2005). 144

[328] S. Ramsden, Y. Y. Cheung, and L. Seroude, Functional analysis of the Drosophila immune response during aging, Aging Cell **7**(2), 225–236 (2008). 144

[329] L. Horn, J. Leips, and M. Starz-Gaiano, Phagocytic ability declines with age in adult Drosophila hemocytes, Aging Cell **13**(4), 719–728 (2014). 144

[330] I. Khan and N. G. Prasad, The Aging of the Immune Response in Drosophila melanogaster, J. Gerontol. A Biol. Sci. Med. Sci. **68**(2), 129–135 (08 2012). 144

[331] K. J. Lesser, I. C. Paiusi, and J. Leips, Naturally occurring genetic variation in the age-specific immune response of Drosophila melanogaster, Aging Cell **5**(4), 293–295 (2006). 144

[332] J. Chong, D. S. Wishart, and J. Xia, Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis, Curr. Protoc. Bioinformatics **68**(1), e86 (2019). 147

[333] R. Bro and A. K. Smilde, Principal component analysis, Anal. Methods **6**, 2812–2831 (2014). 147

[334] J. Shao, Linear Model Selection by Cross-validation, J. Am. Stat. Assoc. **88**(422), 486–494 (1993). 147