The Metabolomics of Desiccation Tolerance in *Xerophyta humilis*

by Halford Dace

Dissertation presented for the degree of Master of Science
in the Department of Molecular and Cell Biology
University of Cape Town
February 2014

Supervisors:
Prof. J. M. Farrant
Dr M. S. Rafudeen

Plagiarism Statement

I know the meaning of Plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Abstract

Resurrection plants are unique in the ability to survive near complete water loss in vegetative tissues without loss of viability. In order to do so, they employ multifaceted strategies which include structural adaptations, antioxidant and photoprotective mechanisms, and the accumulation of proteins and metabolites that stabilise macromolecules. A full understanding of the phenomenon of vegetative desiccation tolerance will require a systems view of these adaptations at the levels of the genome, the control of gene expression, and the control of metabolic pathways.

This dissertation reports a high-throughput metabolomic analysis of the changes that occur in vegetative tissues of resurrection plant Xerophyta humilis during dehydration. A combination of chromatography, mass spectrometry and nuclear magnetic resonance revealed numerous primary and secondary metabolites in the plant. Multivariate statistics identified a subset of metabolites that were significantly up- or down-regulated in response to water deficit stress. These metabolites both confirmed existing observations about the metabolic response of *X. humilis* to drying and revealed compounds not previously known to be associated with this response.

Desiccation-associated metabolites were mapped onto known biochemical pathways, to generate hypotheses concerning possible regulatory schemes in the stress response, inviting deeper investigation in future.

Table of Contents

Chapter One: Introduction and Background 9
Introduction 9
Desiccation Biology 11
Metabolomics 20
This Study 46
Chapter Two: Materials and Methods 50
Plant Source and Maintenance 50
Water Content Determination 51
Metabolite Extraction 53
GC-MS Experiments 54
LC-MS Experiments 57
NMR Experiments 58
Chapter Three: Numerical Analysis 60
Data Pre-Processing 60
Statistical Treatment 63
Compound Identification 66
Pathway Inference 69
Chapter Four: Results and Discussion 70
Drying Kinetics 70
Principal Component Analysis 72
Partial Least Squares Discriminant Analysis 86
Targeted Compounds 94
Untargeted Compounds 98
Compound Pathway Inference 102
Metabolite Changes in the Context of Plant Desiccation 107
Conclusion and Future Prospects 112
Appendices 116
Appendix A: Selection Among Scaling Algorithms 117

Appendix B: Untargeted Peak Lists....121
Appendix C: GC-MS Chromatograms....140
Appendix D: LC-MS Chromatograms....173
Appendix E: NMR Spectra....190
Appendix F: Perl Code for Unit Mass Consolidation....194
Appendix G: R Code for PCA....195
Appendix H: R Code for PLS-DA....197
Appendix I: R Code for Pathway Mapping....202
Bibliography....205

List of Figures

Figure 1. Relative Water Content of *X. humilis* leaves....70
Figure 2. Variance per Principal Component for GC-MS peaks of *X. humilis* leaf extract....73
Figure 3: Score and loading plots for GC-MS leaf extract data. 74
Figure 4: Variance per Principal Component for GC-MS peaks of *X. humilis* root extract,..76
Figure 5: Score and loading plots for GC-MS root extract data. 77
Figure 6: Variance per Principal Component for LC-MS peaks of *X. humilis* leaf extract....79
Figure 7: Score and loading plots for LC-MS leaf extract data. 80
Figure 8: Variance per Principal Component for LC-MS peaks of *X. humilis* root extract. ...83
Figure 9: Score and loading plots for LC-MS root extract data. 84
Figure 10: PLS regression for GC-MS peaks of *X. humilis* leaf extracts 87
Figure 11: PLS regression for GC-MS peaks of *X. humilis* root extracts....89
Figure 12: PLS regression for LC-MS peaks of *X. humilis* leaf extracts....91
Figure 13: PLS regression for LC-MS peaks of *X. humilis* root extracts 93
Figure 14: Untargeted compound peak summary....99
Figure 15: Relative activity scores of inferred pathways....104
Figure 16: PCA with no data scaling....118
Figure 17: PCA with autoscaled data....119
Figure 18: PCA with Pareto-scaled data....120

List of Tables

Table 1: A Selection of metabolite classes, with examples and properties 27
Table 2: Chromatographic modes and their uses 33
Table 3: Mass Analysers in common MS instruments 34
Table 4: Common NMR experiments and their metabolomic applications 38
Table 5: Standard substances for peak identification 56
Table 6: Experimental NMR parameters 59
Table 7: Targeted sugars up-regulated on desiccation 95
Table 8: Targeted sugars down-regulated on desiccation 95
Table 9: Targeted amino acids up-regulated on desiccation 96
Table 10: Targeted organic acids up-regulated on desiccation 96
Table 11: Targeted organic acids down-regulated on desiccation 96
Table 12: Targeted sugar alcohols up-regulated on desiccation 96
Table 13: Targeted sugar alcohols down-regulated on desiccation 96
Table 14: Targeted compounds significantly regulated in roots 97
Table 15: Summary of peaks found per experiment 100
Table 16: Pathway activities inferred from metabolite data. 105
Table 17: Peaks predominantly observed in the hydrated state in leaves 121
Table 18: Peaks predominantly observed in the desiccated state in leaves 122
Table 19: Peaks predominantly observed in the hydrated state in roots 124
Table 20: Peaks predominantly observed in the desiccated state in roots 124
Table 21: Ions predominantly observed in the hydrated state in leaves 125
Table 22: Ions predominantly observed in the desiccated state in leaves 128
Table 23: Ions predominantly observed in the hydrated state in roots 137
Table 24: Ions predominantly observed in the desiccated state in roots 137

Acknowledgements

Most of this MSc project was conducted while I worked full-time in an unrelated field. Known as "The Phantom of the Laboratory", to my laboratory colleagues I was a shadowy figure who haunted the department at night and over weekends.

Completing a major project in this way would not have been possible without the moral and material support of numerous people, who deserve so much more than this acknowledgement.

First, my supervisors, Prof Jill Farrant and Dr Suhail Rafudeen, have been unstinting in their support, encouragement and enthusiasm as this project has unfolded.

The Department of Molecular and Cell Biology and the Faculty of Science have been supportive and patient with a student who they may not have seen as often as they would have liked, and I am very grateful for the faith that they have placed in me during the course of the work.

Keren Cooper, the Senior Scientific Officer who runs our laboratory, has been a saint, both as a friend and colleague, and has often gone far beyond the call of duty to be helpful and supportive.

My fellow student Millie Hilgart put in many hours in the work of optimising LC-MS data processing, providing considerable advice on operations in MZmine. She has been an enthusiastic pillar of peer support as we share the pain and joy of pioneering metabolomic work in the MCB Plant Stress lab.

Dr Zac McDonald, who ran the MCB Mass Spectrometry lab during the course of this project, provided many hours of training, help and support with the LC-MS and GC-MS instruments.

Pete Roberts, of the UCT Chemistry Department, put up with my incessant demands for more than his instrument could give, patiently running numerous NMR spectra for me.

Dr Neil Ravenscroft, also of Chemistry, provided enthusiastic support in NMR strategies and interpretation.

Natalie Hoffman kindly measured light flux intensities in the plant growth chambers that we were sharing, providing the value reported in this dissertation.

Finally, of course, my friends and family have been encouraging and supportive -- vital parts of what made this project possible.

Thank you all.

Halford Dace
Cape Town, February 2014

Chapter One: Introduction and Background

Introduction

This dissertation describes an analysis of the metabolomic response of the desiccation-tolerant vascular plant *Xerophyta humilis* to water loss. This metabolomic study forms part of an ongoing programme of research into the biology of plant desiccation tolerance, and is intended as a step toward the construction of a fuller model of the systems biology of this desiccation tolerant model organism.

In this light, the goals of this study go beyond simply producing a list of metabolites and their up- or down-regulation in response to environment but include the following: to master and refine laboratory and data analysis techniques for investigating the metabolomes of resurrection plants; to place the observed metabolomic response (as far as is practical) in the context of the known biology of the organism; and, if possible, to use these new data to generate testable hypotheses concerning the sensory, regulatory and metabolic pathways involved in the plant's response to desiccation stress.

This dissertation therefore includes a brief discussion of the biology of desiccation tolerance, and a review of prior work concerning the transcriptomic, proteomic, and targeted metabolite responses of *X.humilis* to water loss.

This project includes the first untargteted investigation of the *X. humilis* metabolome of which we are aware, and although numerous statistically significant metabolites remain unidentified, their mass spectra are now available for further analysis by our or other laboratories.

Desiccation Biology

Desiccation Throughout Nature

Although desiccation tolerance is rare in complex multicellular organisms, it has been observed in numerous clades, including among fungi, lichens, bryophytes, higher plants and animals ((Alpert 2006), (Gaff & Oliver 2013)). In higher plants, desiccation tolerance is a complex trait, involving the co-expression of many genes ((Collett et al. 2004), (Farrant & Moore 2011)), and the complex coordination of systems at every level of biological organisation (Moore et al. 2009). It is unsurprising that such an exotic evolutionary adaptation has evolved rarely, and that it is observed more frequently among simpler than among more complex organisms. Nevertheless, despite the complexity of the trait, desiccation tolerance has been observed widely among terrestrial and intertidal multicellular organisms, with distinctive variations in mechanism among various biological clades(Oliver et al. 2000).

This distribution of such a rare and complex trait reflects a potential selective advantage for organisms that can tolerate periodic environmental water deficits. Indeed, extreme water-stress tolerance is relatively common in ecological niches exposed to intermittent cycles of water abundance and scarcity. By contrast, it is rare in persistently dry environments where adaptive strategies focusing on water conservation (for example by succulents) are more successful.

The diversity of desiccation tolerant organisms includes species from among the prokaryotes, animals (notably nematodes and tardigrades), fungi, lichen, bryophytes and angiosperms. Among plants, many bryophytes and several pteridophytes are desiccation

tolerant. Although the trait is rare among seed-bearing plants (and unknown among gymnosperms), several examples of vegetative desiccation tolerance are known among angiosperms (Farrant et al. 2007).

The Challenge of Water Loss

Water is the *sine qua non* of the processes of life. Not only is it the solvent in which biochemical reactions take place, it also plays a vital structural role in living systems at every level of organisation, and in contributes to their stability through homeostasis. As a highly polar solvent, water provides the matrix in which biochemical reactions occur. Its strong capacity for polar interactions gives it a crucial role in the folding and structural stabilisation of membranes, proteins, chromatin and other biological macromolecules and structures. In addition to this role in stabilising protein structure, water is an ideal solvent for the charged or polar compounds of primary energy metabolism.

The biological functions of water are so diverse and so fundamental that its loss represents a special class of extreme abiotic stress. For an organism to survive the total loss of water requires a complex and coordinated suite of adaptations, corresponding to each of its roles in normal physiology ((Bohnert et al. 1995), (Vertucci & Farrant 1995), (Vertucci & Leopold 1987), (Walters et al. 2002)).

When water is lost from a cellular system, the cell's constituents become more concentrated, promoting non-enzymatic reactions that can be deleterious to cellular function. These include, among many other possible reactions, the cross-linking of amines (including peptides and proteins) by concentrated aldehydes (such as reducing sugars), and many types of oxidative damage to membrane lipids, nucleic acids and other components ((Bray 1993), (Chaves 2002), (Farrant et al. 2007), (Wang et al. 2003)).

A special case of such disordered chemistries is observed in relation to the Reactive Oxygen Species (ROS) associated with photosynthetic metabolism and the excitation of light harvesting complexes by light ((Farrant et al. 2007), (Sherwin & Farrant 1998), (Collett 2003)). These chemical species cause damage to cellular constituents, including structural lipids and nucleic acids. Photosynthetic tissues face the special challenge of ameliorating this light-dependent oxidative damage in the face of water loss (see (Farrant et al. 2007), (Halliwell 1987), (Oracz et al. 2007) and (Wang et al. 2003)).

On a larger, physiological scale, water is vital to both structural integrity and homeostasis in all organisms. Severe water loss results in cytoplasmic shrinkage and ultimately plasmolysis, which represents a major cause of plant death from water loss ((Munns 2002), (Moore et al. 2008)). In woody plants, decreasing soil water potential can cause xylem cavitation and the irreversible loss of vascular conductivity, resulting in plant death (Sherwin et al. 1998).

As a result of these stress mechanisms, most plants cannot survive a Relative Water Content (RWC) of less than 50% for any significant period. Additionally, it has been observed in many plants that fatal cellular damage, due primarily to plasmolysis and to oxidative damage, occurs at intermediate water contents of 50% to 70% RWC (Smirnoff 1993). Many resurrection plants display the distinctive physiological behaviour of rapid drying through these "dangerous" intermediate water contents ((Farrant et al. 2007), (Beckett 2010)), unlike most sensitive plants, which respond to water loss through various water retention strategies (such as closing stomata).

For any organism to survive water loss to these or lower levels, it must deploy a suite of adaptations to the biochemical, structural and physiological challenges that they inevitably experience.

Adaptations in Desiccation Tolerance

Plant desiccation tolerance is associated with a variety of adaptive responses within cells and tissues, that together serve to ameliorate the many types of damage associated with severe water loss ((Farrant et al. 2007), (Farrant et al. 2012), (Leprince & Buitink 2010), (Smirnoff 1998), (Vicré et al. 2004)). Some of these responses are associated with the stress response itself -- the metabolic and structural changes initiated by the plant in response to falling water potentials -- and some may be associated with the ability of the plant to mount such a response -- the state of tolerance that in some species (such as *Mohria caffrorum*) may be gained or lost at different times (Farrant et al. 2009), while others acquire tolerance developmentally or in response to water stress (Oliver et al. 1998).

To date, the reported measurable responses to water loss include changes in mRNA transcript ((Dace et al. 1998), (Collett 2003), (Collett et al. 2004), (Illing et al. 2005), (Oliver et al. 2004), (Oliver et al. 2009), (Neale et al. 2000), (Iturriaga et al. 2006)) and protein expression ((Ingle et al. 2007), (Röhrig et al. 2006), (Jiang et al. 2006)), lipid and soluble metabolite abundances ((Farrant et al. 2007), (Lehner et al. 2008), (Peters et al. 2007), (Whittaker et al. 2001), (Yobi et al. 2013)), cell wall composition ((Moore et al. 2008), (Moore & Farrant 2012), (Moore et al. 2013)), cell ultrastructure and organelle organisation ((Farrant 2000), (Tuba et al. 1993)). Together these changes bring about a tremendous shift in the overall biochemistry and biophysics of the drying cells and tissues. These changes induce anhydrobiosis: a state of chemical quiescence and mechanical stability that is compatible with the restoration of biological processes on rehydration ((Vertucci & Farrant 1995), (Oliver et al. 1998)).

The presence or absence of the ability to induce these changes remains one of the key challenges in desiccation studies. This capacity for tolerance may be gained and lost both

through the plant's life cycle. In seed formation, this is usually gained during embryogenesis and lost on germination; in some resurrection plants, such as *M. caffrorum*, it is seasonally regulated (Farrant et al. 2009), while it is largely constitutive in others reported to date, *X. humilis* falling into the latter category. This capability, while far from fully understood, may involve the activation of sensing and signalling pathways capable of activating the functional gene expression networks associated with the desiccation response itself. Such activation could involve epigenetic modifications of ABA responsive elements.

These adaptations have been extensively reviewed elsewhere (for example see Farrant et al. (2012)), however since the focus of this dissertation is on the development of techniques for high-throughput metabolomic analysis, some additional detail concerning DT-associated metabolites is presented below.

Metabolites

Desiccation tolerant plant tissues change their metabolite profiles in response to water loss. They accumulate metabolites that tend to confer protection to cells in the dehydrated state, and reduce the concentration of metabolites that would become toxic at high concentrations, or in the absence of functional cellular metabolism (Hoekstra et al. 2001). A number of hypotheses have been advanced to account for the roles of these metabolite changes in mitigating the effects of water loss upon plants cells and cellular components. These include antioxidant / photoprotective roles to reduce chemical damage(Kranner & Birtić 2005), and water replacement / compatible solute roles to maintain the structural integrity of membranes and proteins in the absence of the vital structural role usually played by ambient cellular water and its capacity for hydrogen bonding, hydration shell formation and other polar interactions ((Hoekstra et al. 2001), (Farrant 2000), (Oliver et al. 2000)).

Highlights of the existing literature on resurrection plant metabolites include:

Sugars

Both orthodox seeds and vegetative desiccation tolerant tissues display large changes in carbohydrate complement on drying. Numerous studies have recorded the accumulation of sucrose and raffinose family oligosaccharides in desiccation-tolerant plant tissues on drying (reviewed by Farrant et al. (2012) and Oliver et al. (1998)).

The accumulation of these sugars is believed to play a role in cytoplasmic vitrification at very low water contents, and consequently in the stabilisation of cellular components and the reduction in rates of deleterious chemical reactions ((Vertucci & Farrant 1995), (Walters et al. 2002), (Hoekstra 2005), (Berjak & Pammenter 2007)).

Vitrification beyond sugars

Despite the central role of simple sugars such as sucrose in the formation of cytoplasmic glasses on desiccation, it has been observed that the properties of these anhydrobiotic glasses are not fully accounted for by the vitrification of simple sugar solutions, and that other compounds must be involved in their formation (Buitink & Leprince 2004).

It has been suggested by Choi et al. (2011) that specific molar ratios of sugars and key organic acids may serve to form ionic liquids *in vivo* that sufficiently account for the stability of anhydrobiotic systems while explaining the high concentrations of organic acids in plant tissues beyond that required by the kinetics of core metabolism.

Other Compatible Solutes

A number of other compounds have been reported in response to water loss in plants, and are hypothesised to fulfil the role of compatible solutes. These include proline, glycine

betaine, pinitol and sugar alcohols such as sorbitol, inositol and mannitol ((Ingram & Bartels 1996), (Vicré et al. 2004), (Bartels & Hussain 2011), (Majee et al. 2012)).

The polyphenol 3,4,5 tri-O-galloylquinic acid, found in abundant quantities in the vacuoles of the resurrection plant *Myrothamnus flabellifolia*, has been proposed, *inter alia*, as a potential compatible solute in that plant's response to desiccation (Moore et al. 2005).

Antioxidant Systems

Numerous antioxidant systems have been reported in resurrection plants, which together serve to protect desiccated tissues from the considerable damage that would otherwise accrue from ROS and free radical damage. These mechanisms include macromolecular components, such as the enzymes ascorbate peroxidase and superoxide dismutase but most antioxidant capacity is mediated through small molecules. These metabolites include water-soluble compounds such as ascorbate and lipid-soluble compounds like tocopherols (Farrant et al. 2007). Some antioxidants that appear to have particular significance for resurrection plants include glutathione (Kranner & Birtić 2005), α-tocopherol (Kranner et al. 2002) and a wide variety of polyphenols ((Smirnoff 1993), (Moore et al. 2005)) as well as seed-specific antioxidants such as 1-cys-peroxiredoxin (Mowla et al. 2002).

Pigments

Resurrection plants employ two distinct strategies to defend against photooxidative damage, namely homoiochlorophylly and poikilochlorophylly. These correspond to two distinct profiles of light-absorbing compound on drying. In poikilochlorophyllous desiccation tolerant plants (like *Xerophyta* spp), the catabolism of chlorophyll accompanying the disassembly of the photosynthetic apparatus causes a dramatic rise in the relative abundances of carotenoids, and the yellowing of leaves (Tuba & Lichtenthaler 2011). *X.*

humilis accumulates zeaxanthin and lutein after the cessation of photosynthesis on drying, and emits isoprene (Beckett et al. 2012). In homoiochlorophyllous desiccation tolerant plants, the xanthophyll cycle appears to have a role in dissipating excess light energy (Bartels et al. 2011) but otherwise pigment complements are highly variable.

Xerophyta species as Systems Biology Models

The construction of comprehensive models of organism function require the comparison and integration of data from many different experiments to develop a useful mathematical model of such extraordinarily complex systems. Typically, such research is performed on well-characterised model organisms in which the work of many laboratories may (in principle) be directly compared and assimilated toward a comprehensive description of the organism's function. Popular model organisms for basic eukaryotic biology include the brassica *Arabidopsis thaliana,* as a plant model and the nematode *Caenorhabditis elegans,* as an animal model.

A number of properties contribute to the usefulness of such species as biological model organisms. Ankeny (2001) and Ankeny & Leonelli (2011) describe the requirements as:

- Short generation time with reliable, high-throughput propagation methods;
- Available genome;
- Reliable inbred strains;
- Easy generation of mutant / knockout libraries;
- An ecosystem of researchers working on various aspects of the organism's biology.

Like most wild plants, *Xerophyta* spp do not provide us with inbred strains. The propagation methods, while increasingly reliable, are not high-throughput, and no methods for the generation of targeted mutations / knockouts in this species have yet been published.

Nor, at the time of writing, has a genome sequence been published, although a sequencing project is currently underway (Farrant, J. and Hilhorst, H., personal communication 2013)

Fortunately, desiccation tolerant plants have attracted considerable interest overt from numerous research groups and now enjoy a diverse and active research community. While *Xerophyta* will never be *Arabidopsis*, it is emerging as a compelling organism to study poikilochlorophyllous desiccation tolerance. With efforts underway to sequence the genome of a *Xerophyta* species, improve the reliability and throughput of propagation methods, and develop targeted knockout / knockdown technologies, it is hoped that the lacunae in its role as a model will soon be filled.

Aside from the progress being made in developing *Xerophyta* as a model system, increasing attention has been paid in the literature to techniques for integrative biology in non-model organisms (for example Williams et al. (2011)). As functional genome annotations in established models continue to improve, the barriers to annotation of new species' genomes and proteomes diminish considerably, and comparisons with ever-better-known functional networks become practical, albeit still challenging.

Metabolomics

The -omics disciplines

Metabolomics is one of a series of biological disciplines aimed at determining the content and function of the various levels of organisation in living cells and systems.

Broadly speaking, each set of molecules that has a coherent physical, chemical or functional role may be analysed as a full set in a corresponding -omics discipline (Cushman & Oliver 2011).

These disciplines include (but are not limited to):

- **Genomics**: The genome is the inventory of a cell or organism (or organelle, in the case of mitochondria and plastids). This represents stable hereditary material, and the core informational resource for the development and environmental interaction of cells and organisms;
- **Transcriptomics**: The transcriptome is the inventory of mRNA transcripts in a cell or tissue. This set represents transcribed genes that are available for protein synthesis;
- **Proteomics**: The proteome is the full set of proteins currently present in a cell, tissue or organism. This incorporates the full range of protein functional, including structural, metabolic and signaling / control roles;
- **Metabolomics**: The metabolome is the full set of small (generally soluble) metabolites in a cell, tissue or organism. These small metabolites include both the core molecules essential to life (primary metabolites) and those that may provide competitive advantages, but are not essential to biological function (secondary

metabolites.) The diversity of compounds comprising the metabolome is further explored in "Exploring the Metabolome" in this chapter.

These well-understood sets of molecules correlate neatly with the Fundamental Dogma of molecular biology. In addition, any other set of chemical species that presents a coherent set -- whether through chemical similarity or biological function -- may be usefully considered using an -omics approach. Such sets include:

- **Lipidome**: The set of lipids (including glycerides, sphingosides and phospholipids) in an organism or tissue ((Han 2003), (Hermansson et al. 2005));
- **Volatilome**: The set of volatile chemicals produced by an organism under specified conditions (Bicchi & Maffei 2012);
- **Wallome**: The set of compounds that comprise a bacterial, plant or fungal cell wall. Although cell walls are chemically diverse, they form a key, coherent biological function and owe their properties to their chemical composition;
- **Interactome**: The set of protein-protein interactions in a cell and a vital component of systems network models (Arabidopsis Interactome Mapping Consortium et al. 2011);
- **Fluxome**: The set of kinetic fluxes through various pathways involving metabolites, signalling interactions and macromolecular synthesis and degradation (Wiechert et al. 2007).

These approaches, considering full sets of biomolecules that comprise whole functional 'layers' of an organism, have been made possible by advances on two fronts: Analytical instrumentation and computation. High-throughput analysis and laboratory automation allow the acquisition of chemical data with greater sensitivity and specificity and at higher rates than ever before. The increasing availability of computational capacity, and advances in algorithms and software for massively multivariate data, allow the constructive mining and analysis of volumes of data which would not have been previously possible.

None of these approaches represents any single, fundamentally new technology. DNA sequencing, protein sequencing, electrophoresis, chromatography, mass spectrometry and NMR have all been common techniques for decades. However, the shift from gene sequencing to genomics, from peptide sequencing to proteomics, from analytical biochemistry to metabolomics, has been enabled by a combination of incremental (but rapid) changes. Improvements in throughput and cost-effectiveness of analytical techniques allow faster acquisition of more data. Improved data processing allows the mining of genomic or chemometric data to generate new hypotheses based on unbiased inputs. The development of the internet allows these data to be shared through easily-accessible and searchable databases. Most importantly, the willingness of scientists to restore observation-driven science to a place of honour alongside hypothesis-driven experimentation creates the potential for this abundance of data to be a wellspring of novel insights (Zhang et al. 2010).

As both biochemical and computational analysis have advanced, -omics work is moving beyond the simple "parts list of an organism", and towards the ability to ask and answer nuanced questions about organisms' development, environmental interactions (including stress responses) and interactions with other organisms (for example Noble (2006) and Urano et al. (2010)).

Additionally, developments in the -omics disciplines show promise in bridging the divide between molecular biology and ecophysiology (Sanford et al. 2002). Ecological and physiological studies have already made tremendous use of metagenomics: the sequencing and enumeration of the total genomic content of an environment (for example a soil sample), allowing novel insights into the interactions among plants or animals and their microbial environments.

To place this Masters study in context, it is important to understand not only what metabolomics is in relation to other -omics disciplines, but also how it expresses a facet of

functional genomics, and how this contributes toward a comprehensive, Systems Biology understanding of plant desiccation tolerance.

Metabolomics as Functional Genomics

Rapid advances in DNA sequencing technology have driven down the cost and time required for both first-read and subsequent resequencing of whole genomes. The challenge of sequencing no longer requires enormous, global projects, but is within the reach of modestly equipped research teams worldwide.

The resulting explosion in DNA sequence data provides a tremendously valuable resource for biologists, however the actual biological meaning of sequence data frequently remains elusive, even the era of Genome-Wide Association Studies and other sophisticated analyses.

Genes do not exist in a vacuum, and their biological effect is mediated through a physiological context. For a gene to bring about a phenotypic effect, it must be expressed in a process orchestrating signaling, promotors and transcription factors, RNA catabolising or silencing factors, ribosomes, protein folding factors, protein-protein interactions and interactions with non-protein metabolites.

Determining the function of a gene (or a genome) therefore transcends the one-dimensional world of nucleotide sequences, and incorporates every aspect of cell biology, including the effect of gene products on protein-protein interaction networks, metabolite pools and metabolic kinetics (Schauer & Fernie 2006).

A full understanding of genome function requires not only a genome sequence, but also an understanding of what physiological changes (including changes to metabolite pools) are linked with the activation of particular genes, or sets of genes. Metabolomics, along with transcriptomics, proteomics and interactomics, are all important components in fully

understanding gene function within the complex environment of a living organism (Fiehn 2002).

In the 21st century, the cost and effectiveness of high-throughput analytical techniques (for example nucleic acid sequencing and mass spectrometry) have improved dramatically, while new computational techniques have been developed. These two threads of life science technology have fostered the emergence of Systems Biology as a formal discipline in which the relationship between genes and environment may be accurately modelled and computed.

In the context of desiccation tolerance biology, it has been suggested ((Illing et al. 2005), (Oliver et al. 2000)) that vegetative desiccation tolerance phenotypes are largely a product of the repurposing of existing capabilities that are more usually observed in the biology of simple, non-vascular plants and in the orthodox seeds of higher plants. Part of the goal of VDT research is to improve the water deficit tolerance of agriculture crops and pasture grasses. Achieving this goal requires more than genome sequences, and more than a physiological understanding of stress responses. The ability to engineer plant performance improvements calls for a thorough, nuanced understanding of the relationship between genes, the control of gene expression, and their function in the context of the (stressed) plant cell (Sanford et al. 2002).

This metabolomic study forms part of an ongoing attempt to contextualise and enrich our understanding of gene function in the water stress response of desiccation tolerant plants such as *X. humilis.*

Exploring the Metabolome

The great challenges of metabolomics lie in the chemical diversity of metabolites, and in the handling of their statistical and computational analysis.

Genomes and proteomes comprise biological macromolecules that are informationally and physically complex, but are composed chemically of heteropolymers of small sets of well-defined monomer units. This chemical uniformity allows nucleic acids and proteins to be analysed and sequenced using modular, scalable and well-defined technology.

Metabolomes, by contrast. have no such chemical uniformity. The set of soluble metabolites comprises an enormous range of organic compounds with a wide variety of physical characteristics and functional roles. In addition, their solubility in polar versus non-polar solvents varies, posing both qualitative and quantitative analytical challenges ((Kopka 2006), (Allwood & Goodacre 2010)) .

Table 1 illustrates a limited subset of the range of soluble compounds involved in plant metabolism, with a brief description of their biological role, chemical and physical properties (Examples drawn from Anderson & Beardall (1991) and Hanson (2003)) .

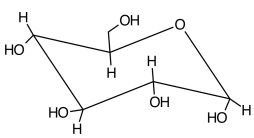
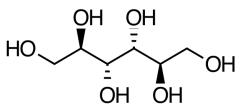
This list is by no means exhaustive, but illustrates the tremendous analytical challenge faced by metabolomists. The variety and chemical diversity of small-molecule metabolites in a plant cell makes it impossible to select a single solvent system in which they are all soluble, and certainly presents enormous challenges to the selection and optimisation of suitable analytical techniques.

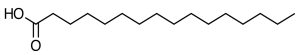
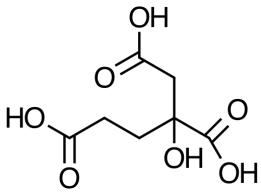
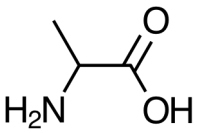
At this time there is no generally accepted "shotgun" analytical technique that successfully identifies and quantifies all plant metabolites. The selection of analytical approaches, even in untargeted studies, presents compromises among the type of metabolites

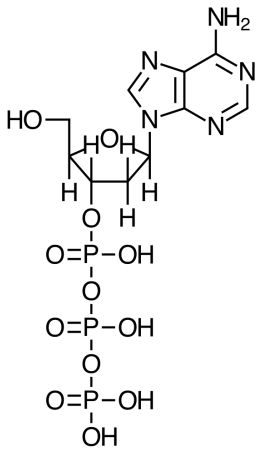
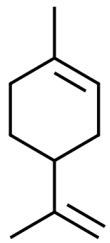
of interest, the capabilities of various instruments and, of course, their availability to the researcher.

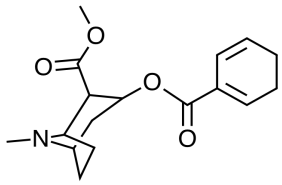
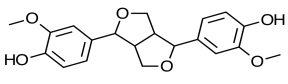
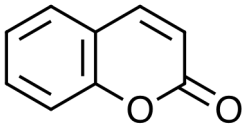
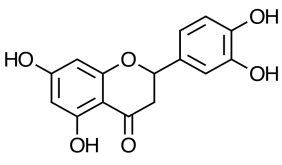
Nevertheless, a number of approaches have been identified that, individually or in combination, permit a reasonably broad view of a plant metabolome. This view is never perfectly comprehensive, but can be biologically useful, and retain the vital capacity to surprise, by detecting and quantifying unexpected and previously unlooked-for compounds ((Allwood et al. 2009), (t'Kindt et al. 2009), (Kanani et al. 2008), (Kopka et al. 2005)).

Table 1: A Selection of metabolite classes, with examples and properties

Class	Examples	Functions	Properties
Sugars	Glucose  α-D-glucopyranose Also: Fructose Sucrose Maltose Arabinose Various oligo- and polysaccharides	Energy metabolism, storage and storage Glass formation / water replacement / compatible solutes in anhydrobiosis Structural polymers	Polyhydroxy-aldehydes and -ketones. Mono- and oligomers are highly solube in aqueous solvents (e.g. Water, 50% methanol) Some (e.g. Glucose) have strong reducing capacity that can cause harm at elevated concentrations by cross-linking proteins
Sugar Alcohols	Mannitol  D-(-)-mannitol Also: Glycerol Erythritol Arabitol Xylitol Ribitol Sorbitol	Compatible solute / vitrification in anhydrobiosis	Sugar analogs, with carbonyl group reduced to hydroxyl Similar solubility and physical characteristics to sugars Lower chemical activity and available metabolic energy

Fatty Acids and Lipids	Palmitic acid 	Energy storage Structural role in membranes as phospholipids Intermediates in production of many secondary metabolites and volatiles	Not significantly soluble in water Soluble in low-polarity organic solvents Distinctive physical properties and biological role leads to lipids being analysed separately as *lipidome*.
Organic Acids	Citric Acid 	Primary energy metabolism (via TCA cycle)	Highly polar or ionic, water-soluble Varying number of carboxyl groups Low pH in solution
Amino Acids	Alanine 	Monomers of proteins Vital components of overall nitrogen metabolism Some (notably phenylalanine and tryptophan) are critical precursors of secondary metabolites Upregulation of some amino acids is observed in generalised stress responses	Water soluble Participate in amine chemistry (for example by cross-linking with aldehydes)

Nucleosides and Nucleotides	ATP  Also: Adenine Adenosine Adenosine monophosphate (AMP) Cyclic AMP Adenosine diphosphate Similar sets of derivatives of cytidine, guanidine, thymidine and uridine.	Primary energy metabolism (particularly ATP and ADP) Monomers of DNA and RNA Signalling (notably cAMP) Components of several nitrogen metabolism pathways	Soluble in water Phosphorylated nucleotides are negatively charged and will not cross phospholipid bilayer membranes
Terpenes	Limonene 	Secondary metabolites Components of many essential oils Plant growth regulation Anti-herbivory Several drugs	Frequently volatile Low polarity

Alkaloids	Cocaine 	Secondary metabolites Many drugs and poisons (cocaine, nicotine, morphine, mescaline, adrenaline)	Nitrogen containing Structurally diverse
Lignans	Pinoresinol 	Secondary metabolites Phytoestrogens Beneficial phytonutrients	
Coumarins	Coumarin 	Secondary metabolites Possible anti-herbivory defence Used as fragrance compound in perfumes and household products Moderately toxic, but some derivatives useful as drugs	Volatile
Flavonoids	Luteolin 	Plant pigments Antioxidant activity Numerous physiological roles	polyhydroxy polyphenolics Usefully detected by UV-Vis spectra

Experimental Approaches

Any experimental technique must meet certain keep criteria if it is to be usefully applied to problems in metabolomics. These criteria relate to the number and diversity of metabolite species, their range of concentrations and the need for both qualitative and quantitative output data.

The key criteria include:

- Resolution: The ability to distinguish many different compounds in a sample, such that they can be independently described and quantified. Similarly, the technique should be applicable to a wide range of small biological molecules;
- Sensitivity: The ability to detect relatively rare molecular species and reliably distinguish them from noise;
- Dynamic range: The ability to reliably quantify metabolites at a wide range of concentrations, from very rare to extremely abundant.

Although a wide variety of spectroscopic techniques have been used for the parallel detection of multiple metabolites, most metabolomic projects rely some combination of chromatographic separation with mass spectroscopy (MS) or nuclear magnetic resonance (NMR) spectroscopy. Detection technologies (MS or NMR) may be used in tandem with chromatographic separation, or directly on complex metabolite mixtures.

Chromatographic Methods

The use of chromatography prior to MS or NMR analysis serves to separate complex mixtures, presenting simpler analytes to detection technologies. Reducing the number of simultaneously detected metabolites helps to eliminate multiple overlapping signals, simplifying both the quantitation and identification of molecular species. In addition, the use of highly reproducible chromatographic systems adds an analytical dimension to the data

available about a compound, in the form of an absolute retention time or relative retention index.

Despite these advantages, the addition of a chromatographic step can increase the cost and complexity of metabolomic studies. They introduce an additional step that must be optimised, and can reduce the effective sensitivity of the detection step by spreading out each single over the peak elution time.

A number of chromatographic techniques, including paper chromatography, thin-layer chromatography (TLC), gas chromatography (GC) and liquid chromatography (LC) have been used to separate complex metabolite mixtures for analysis, with TLC, GC and LC remaining in common use ((Fox et al. 1998), (Kopka 2006), (t'Kindt et al. 2009), (Allwood & Goodacre 2010)).

The characteristics of each of these techniques are summarised in Table 2.

Table 2: Chromatographic modes and their uses

Phases (Mobile / Stationary)	Typical Analytes	Detection
TLC		
Liquid Solvent / Solid silica gel on glass or aluminium support	Mid-polar compounds, pigments, amino acids	UV fluorescence, MALDI MS, DART MS
GC		
Gas / silica column with specific surface activity	Volatile compounds, low polarity. Includes chemically derivatised primary metabolites.	EI MS
LC		
Liquid solvent / various stationary phases. Reverse phase separation with C-18 columns most common for metabolomics	Enormous variety of analytes. C-18 columns separate mid- to low-polarity compounds. Other options such as HILIC columns for polar metabolites	ESI / APCI MS

Mass Spectrometric Methods

Mass spectroscopy provides a sensitive and versatile detection tool for the high-throughput analysis of metabolite mixtures. Various ionisation methods, mass analyser designs and detectors compose a wide variety of instruments, suited to diverse analytical roles. Further versatility is provided in some instruments through the implementation of MS-MS systems in which ions (either molecular or primary fragments) are further fragmented, providing additional structural information for quantitative analysis.

The design of a MS instrument determines what kind of chromatographic method (if any) may be interfaced to the device, and the mass accuracy, precision, sensitivity and dynamic range of the mass spectra produced. ((Allwood & Goodacre 2010), (Kopka 2006))

A brief summary of common ion sources and mass analysers follows in Table 3, with notes on their utility in metabolomics:

Table 3: Mass Analysers in common MS instruments

Ion Source	Properties	Applications
Electron Impact (EI)	High-vacuum bombardment of analyte molecules by accelerated electrons. Extensive fragmentation of most analytes ('Hard' ionisation).	Highly reproducible; extensive metabolite databases exist. Not suitable for LC systems, limiting utility to gas-phase inputs. 'Hard' ionisation can render molecular ion undetectable, posing a challenge for some analyses.
Chemical Ionisation (CI)	Indirect ionisation of analyte by charge transfer from ionised intermediate molecule.	Can ionise compounds harder to analyse by EI, but poorer reproducibility and utility for most cases.
Electrospray Ionisation (ESI)	Desolvation of fine, charged droplets of liquid-phase analyte solutions. Limited fragmentation of many analytes ('Soft' ionisation).	Enables MS analysis of LC-separated mixtures. 'Soft' ionisation renders molecular ion detectable, but fragmentation patterns, are less reproducible than EI spectra. Susceptible to quenching, multiple charges, salt suppression. Consequently, few ESI spectrum databases exist.
Matrix-Assisted Laser Desorption Ionisation (MALDI)	Ionisation of analytes embedded in a thin layer of a special ionisation matrix. Laser illumination ionises the matrix, that in turn ionises the analyte via charge transfer.	MS analysis of analytes, including macromolecules, on a specially prepared surface. Analytes may be spotted onto stable base, or present in a biological section. Analysis of a matrix of sites on the samples allows the construction of an image of the distribution of compounds of interest through a sample section (MALDI Imaging)
Ambient Ionisation (DESI / DART)	CI variant: Ionisation of surface compounds of solid or liquid samples in ambient air, via stream of charged solvent droplets (DESI) or gas (DART)	Unique capability to analyse solid surfaces without disruptive sample preparation. Useful for forensic analysis. Can implement TLC-

		MS experiments, adding MS capability to TLC separations
Mass Analyser		
Quadrupole	Scans selected spectrum range by deflecting / focusing ion bean using magnetic quadrupole. High speed, low cost, limited mass accuracy.	Routine analysis of almost any ion stream. Most useful for smaller molecules.
Triple quadrupole (QQQ)	Two quadrupole analysers separated by a reaction cell in which secondary fragmentation of primary ions occur. Fragment ions are analysed by the second quadrupole, generating MS^2 spectrum of selected parent ion.	MS^2 analysis enhances sensitivity and specificity in the analysis of unknown metabolites or discrimination between very similar compounds (for example fatty acids)
Time of Flight (TOF)	Mass analysis through timing the flight of ions through a long flight tube. Ions are given equal kinetic energy through charge-based acceleration. Since $E_k = ½mv^2$, mass may be accurately determined if the time of flight is precisely known.	Suitable for analysis of large ions, including peptides and nucleic acids. High mass accuracy in modern TOF instruments enables use of mass defects to determine molecular formulas and fragment ion compositions.
Quadrupole Time of Flight (QTOF)	MS-MS method in which a primary ion, selected by a quadrupole analyser, is fragmented in a collision cell, with the daughter ions analysed by TOF-MS.	MS^2 analysis, with the high mass accuracy of a TOF instrument.
Ion Trap	Traps ion populations in magnetic field traps of various types (Linear / Orbitrap / FT-ICR), accumulating large populations of analyte ion over time. Can be used in innovative ways, including indefinite MS^n sub-fragmentation.	Analysis of rare ion species with concentrations below the threshold of detection for other methods. Sophisticated MS^n experiments for challenging analytes.

Nuclear Magnetic Resonance Spectrometric Methods

Nuclear Magnetic Resonance (NMR) spectroscopy provides a suite of analytical methods that reveal rich data about the composition and structure of organic molecules, including metabolites and biological macromolecules (Kim et al. 2011).

Given the richness of chemical information available through NMR studies, it is unsurprising that these techniques have been applied to problems in metabolomics, metabonomics, natural products chemistry and pharmacognosy.

NMR based metabolomic experiments are usually performed on crude or solvent-partitioned metabolite extracts, since few laboratories have the means to couple HPLC platforms to NMR spectrometers. Nevertheless, where such equipment does exist, it provides a particularly compelling metabolomics platform, particularly since the nondestructive nature of NMR analysis allows downstream MS analysis of HPLC eluent. Such LC-NMR-MS platforms represent the current pinnacle of metabolomic instrumentation.

In practice, the NMR instruments available in many laboratories present challenges which limit their use for high-throughput metabolomics. The sensitivity of NMR instruments is considerably lower than that of MS instruments for most metabolites. This relatively poor sensitivity becomes even more of a problem once an analyst wishes to collect spectra from nuclei other than the proton. ^{13}C, for example, while extremely chemically informative, is much less abundant than ^{12}C, and has a far smaller resonant energy than ^{1}H, making the detection of rare molecular species in ^{13}C spectra almost impossible.

Additionally, the analysis of complex metabolite mixtures is difficult given the sheer number of overlaid signals, particularly in a 1-dimension proton NMR spectrum (Schripsema 2010).

Nevertheless, the convenience and information density of NMR makes it an attractive technique for metabolomics. Problems with both sensitivity and signal complexity are reduced through the use of higher-field instruments, longer acquisition times, and the selection of appropriate NMR experiments.

Generally, NMR metabolomics experiments employ simple 1D proton spectra for numerical analysis, and the statistical determination of peaks that very significantly among sample groups. These peaks of interest are characterised using a range of 2D NMR experiments.

A selection of NMR experiments and their utility is listed in Table 4.

Table 4: Common NMR experiments and their metabolomic applications

Experiment	Properties	Applications
^{1}H (1-D proton spectrum)	Reveals chemical environment / deshielding of protons, along with coupling constants that provide structural information. Signal intensity is linearly related to number of protons with effectively unlimited dynamic range.	Primary spectrum for metabolomics work. Spectral information statistically binned and analysed to reveal peaks of interest.
^{13}C (1D carbon spectrum)	Reveals chemical environment / deshielding of ^{13}C nuclei. No coupling owing to rarity of isotope.	Not generally useful for primary metabolomic work since many species too rare for detection. Extremely useful for purified / fractionated compounds of sufficient abundance.
J-Resolved (2D plot of chemical shift vs coupling constant (J))	Reveals coupled protons in analyte molecules.	Useful in analysis of highly convoluted, overlapping signals where coupling may be difficult to determine without this instrumental assistance.
COSY (2D Correlated Spectroscopy)	Reveals relationship between protons on adjacent or nearby protons within many organic molecules. Choice of pulse program parameters determines the proximity that is detected.	Information about adjacent or nearby protons useful in determining the structures or identities of proton peaks of statistical interest.
TOCSY (2D Total Correlation Spectroscopy)	A COSY-like pulse sequence that reveals the relationships between protons on all correlated carbons in a molecule.	Useful in distinguishing the proton signals from each molecular species, with the caveat that certain molecular features, such as ethers and esters, decorrelate the carbon skeletons on either side.
HSQC (2D Heteronuclear Single Quantum Coherence)	Reveals the relationship between directly bonded ^{1}H and ^{13}C nuclei in an	Correlation of proton with carbon shifts, obtained at the sensitivity of proton

	organic molecule.	NMR, provides additional data about the chemical environment in each region of a molecule, and the relationship among various molecular features.
HMBC (2D Heteronuclear Multiple Bond Coherence)	Reveals coupling between a proton and relatively distant carbon nuclei.	Provides information about the skeleton of an analyte species.

Other Spectrometric Methods

Mass Spectrometry and NMR, in conjunction with various chromatographic systems, are the most popular techniques used in high-throughput metabolomic research. They are not, however, the only ones. A variety of methods capable of simultaneous detection of multiple analytes have been used either in conjunction with or as an alternative to MS / NMR.

For the sake of completeness, a selection of these techniques is listed below:

- **UV-Visible spectroscopy**: Detects conjugated double bond systems in organic compounds. Frequently used in HPLC systems via a Diode Array Detector (DAD). UV/Vis spectra can be used as supplementary data in parallel with MS data, or alone for the determination of plant pigments and other natural products
- **Infrared (IR) spectroscopy**: Detects vibrational and rotational modes in molecular species. Since hydroxyl groups (and water molecules) absorb intensely and broadly in the infrared spectrum, IR spectroscopy is limited to nonaqueous samples. This limits its utility in the analysis of water-soluble metabolomes, but it remains useful in the analysis of lipids and other nonaqueous metabolites, as well as finding extensive application in the chemometrics of the pharmaceutical and petrochemical industries.

Statistical Challenges in Metabolomics

Untargeted comparative metabolomic studies (such as this one) aim to identify those metabolites which differ between two populations or experimental treatments in as unbiased a way as possible. When these populations are of similar organisms in different physiological conditions (such as hydrated and dehydrated *X. humilis*), differences in metabolite pools may provide insights into the organism's mechanisms of environmental adaptation.

Providing such an unbiased view into metabolite pools requires not only sensitive and broad-spectrum analytical techniques (such as GC-MS, LC-MS and NMR), but also methods of numerical analysis that allow us to extract meaningful information from massively multivariate data. For example, in this study the aligned data matrix for leaf extract LC-MS data contained 7168 variables over 21 samples.

Such data sets present a challenge to traditional multivariate analysis for the following reasons (Varmuza & Filzmoser 2009):

- Their extremely high dimensionality over limited sample numbers;
- The mixed level of variable independence, and high level of correlation among some variables. For example, metabolites fructose and glucose each have two GC peaks in our chromatography protocol, meaning that one metabolite gives rise to two, perfectly correlated variables;
- The extreme range of metabolite concentrations creates challenges to scaling, and the weighting of variable significance. A picomole of an obscure signaling molecule is as interesting as a nanomole of an abundant saccharide, but instrument noise is not interesting, and its contribution to the analysis should be minimised.

While these challenges for data of this degree of dimensionality remain an active area of research among specialist statisticians, a number of statistical methods are available to the

biologist that are relatively robust to these challenging datasets, and reasonably simple to work with.

Unsupervised Techniques

In unsupervised techniques, the algorithms have no *a priori* knowledge of which samples belong to which populations. They use only the data values themselves to examine the relationships among samples.

PCA -- Principal Component Analysis

Principal Component Analysis ((Varmuza & Filzmoser 2009), (Wehrens 2012), (Ramsay et al. 2012)) is a technique for data reduction and visualisation of multivariate data sets. The technique involves calculation of a covariance matrix for the data set, and the computation of a set of eigenvectors that correspond to a set of orthogonal, latent variables (Principal Components) that succinctly describe the variability within the data set and identify those variables that most contribute to it. By plotting samples' Principal Component scores it is possible to visualise whether sample sets of interest cluster within the multidimensional data space, and so to determine which principal components (and hence which original variables) account best for the differences among sample populations. In most experiments, the overwhelming majority of the data set variability is accounted for by the first few principal components, despite there being as many components as there are variables in the full data set.

PCA has been shown to be particularly useful for megavariate datasets with high levels of co-variance, such as those generated in metabolomic experiments.

Sometimes the experimental conditions of interest (such as desiccation stress in this study) do not contribute the majority of sample variation. In these cases, treatments of

interest may not cluster usefully in PCA plots, and no meaningful information about the variables responding to the experimental treatment will be provided. In these cases, supervised techniques such as PLS Discriminant Analysis will be more useful.

HCA -- Hierarchical Cluster Analysis

Some experiments compare traits across numerous sample populations. In these instances, it can be useful to determine the relative degree of difference or similarity among populations, and to produce a hierarchical tree illustrating clusters of similarity among the populations (Varmuza & Filzmoser 2009).

Since experiments involving only two populations cannot generate meaningful dendrograms, HCA is not used in this study and is mentioned here only for completeness.

Supervised Techniques

Other statistical algorithms use explicit knowledge of sample of sample types, or of the predicted variable, as input. These algorithms attempt to fit an algebraic model using on the measured variables such as to successfully predict the population category / dependent variable (in our case, the water content of a dehydrating resurrection plant) based on the independent variables (relative concentrations of various metabolites).

While Linear Regression and Linear Discriminant Analysis are perhaps the most commonly known techniques of supervised multivariate regression, their value is limited in the case of metabolomics data. These techniques assume perfect independence of the independent variables, while metabolomic variables tend to be highly correlated, both because of the relatedness of compounds in common pathways, and also because individual compounds can give rise to multiple peaks in both chromatographic separation and NMR

experiments. This lack of variable independence is reported to result in unstable linear regressions with diminished predictive value (Varmuza & Filzmoser 2009).

A number of statistical techniques have been developed to overcome this challenge, and have found considerable acceptance as chemometric tools. The most commonly used of these are described below.

PLS-regression - Partial Least Squares Regression

PLS regression, like PCA, employs a set of orthogonal latent variables, but unlike Principal Components, the PLS components are chosen not to encapsulate the maximum total variability, but maximise the extent to which the variability along the axis of the latent variable correlates with the dependent variable -- in other words, the extent to which the latent variable is predictive of, for example, water content. As for PCA, the regression coefficients for each individual constituent of the latent variables selected correspond to the degree to which an original independent variable (metabolite) contributes to the biological change of interest ((Varmuza & Filzmoser 2009), (Wehrens 2012)).

PLS-DA - Partial Least Squares Discriminant Analysis

Many studies have dependent variables that are categorical rather than continuous. For example, comparisons between wild-type and knockout strains of a model organism, or between virus-infected and uninfected tissues, or as in the case of this study, between fully hydrated and desiccated resurrection plants. In these cases, the goal of the analysis is to produce a model that successfully classifies samples into into one of the dependent-variable categories based on the independent variables. The resulting model allows us to determine

which independent variables are most significantly associated with the biological character of the populations.

Such classification problems are a special case of regression analysis and can make use of the same algorithms. In the case of PLS regression, the method is given the name of PLS Discriminant Analysis (PLS-DA) when used as a classification technique rather than as a regression to a continuous dependent variable, but the practical use and application of the technique is similar ((Jansen et al. 2010), (Varmuza & Filzmoser 2009)).

OPLS - Orthogonal Projection onto Latent Structures

The area of megavariate statistical analysis has seen considerable, active research by statisticians, largely in response to the complex data sets being generated in the life sciences. One of the alternative methods developed to handle metabolomics data is OPLS, or Orthogonal Projection onto Latent Structures. OPLS seeks to improve predictive power and reduce the overfitting that can plague PLS regression models. Unfortunately, the algorithm is patented and is only currently available in few software packages. While this technique is promising, Tapp et al (2009) suggest that it does not significantly improve on PLS in most practical analyses (Tapp & Kemsley 2009).

Assessing Model Quality

Fitting a function to a data set does not guarantee that the resulting model will be successfully predictive for new samples. In order to have confidence in the biological conclusions being drawn from a metabolomic model, the model must be validated and its predictive power confirmed.

In traditional multivariate statistics, in which the number of samples is considerably greater than the number of variables, it is possible to compare multiple independent subsets

of the data for model consistency and prediction error. In metabolomics experiments, however, the number of independent variables vastly exceeds any practically plausible number of samples analysed. This constrains the available approaches for model validation.

In such cases, the available validation techniques consist of leave-one-out (or leave-multiple-out) cross-validation, jack-knifing and permutation. These techniques are all variations on the theme of empirical goodness-of-fit tests. Leave-one-out (or leave-multiple-out) validation tests whether the model on a diminished training set (less one or several samples) still successfully predicts or classifies those samples, while permutation tests randomly reassigns samples among categories, to confirm the loss of predictive power from resulting models ((Jansen et al. 2009), (Varmuza & Filzmoser 2009)).

This Study

Choice of Approaches

The choice of experimental approaches for this study was governed by a combination of instrument availability, experimental goals (particularly for the untargeted component of the study) and the limited tissue mass available from each *X. humilis* plant.

At the start of the study, it was anticipated that NMR would be the primary analytical technique, since a 400MHz instrument was readily available, and the analytical versatility of the technique was attractive. Preliminary experiments were disappointing, however, for two reasons: First, the limited sensitivity of NMR, combined with the small size of a *X. humilis* sample, meant that spectrum signal strength was poor and several peaks did not achieve a good enough Signal to Noise Ratio to be reliably quantifiable. Second, the 400MHz instrument did not provide good peak separation in the δ3.0 to 4.5 ppm region of the proton spectrum, which is vital for the relative quantification of carbohydrates.

While issues with both sensitivity and resolution of NMR signals could be resolved through the use of a higher-field instrument, the relative accessibility of mass spectrometers made the use of MS techniques more attractive.

During the second year of this study, GC-EI-QQQ and LC-ESI-QTOF mass spectrometers were installed in our department. These instruments provided the flexibility for a number of possible experimental approaches to a plant metabolome. The availability of both gas and liquid chromatography platforms enabled a broad survey of plant metabolites with the possibility of both assessing the roles of known compounds in *X. humilis* desiccation

tolerance through a targeted analysis, and finding previously unknown compounds of statistical interest, through an untargeted investigation.

To achieve these ends, a targeted component of the study was designed, based on the GC-MS instrument, in which hydrated and desiccated leaves and roots of *X. humilis* were compared in terms of a number of known metabolites including sugars, sugar alcohols, amino acids and simple organic acids.

This targeted approach was augmented by an untargeted component, in which compounds found by spectral deconvolution were statistically compared between hydrated and desiccated populations, and unknown compounds of statistical interest were subjected to spectral analysis and database searches in order to assign putative identities.

A second, untargeted experiment was designed in which LC-MS was used to find and, where possible, identify secondary metabolites up- or down-regulated in response to desiccation stress. In this case, owing to the difficulty of interpreting ESI-TOF mass data, supplementary QTOF MS^2 spectra were generated at a range of collision energies to assist in compound identification.

To further aid in compound identification, a range of NMR spectra (1D proton, COSY, TOCSY and HSQC) were generated for pooled hydrated and dehydrated leaf and root extracts, and used to select the most probable among a shortlist of candidate compounds generated by MS analysis.

Untargeted mass spectrometric data sets were subjected to both supervised (PLS-DA) and unsupervised (PCA) multivariate methods in order to refine the peak list to those compounds most significantly associated with the difference between hydrated and desiccated states. Only the most significant peaks were considered as candidates for compound identification.

Towards a Systems Approach

A complex trait such as vegetative desiccation tolerance involves a coordinated response across every level of organisation of a plant in order to achieve a physiological state of tolerance (the capability of a hydrated plant to enter an anhydriobiotic state on drying), and to reach anhydrobiosis itself.

A biological understanding of desiccation tolerance, whether as basic biology or with a view to engineering more stress-tolerant organisms or ecosystems, must include quantitative and qualitative models of the networks of interaction underlying the trait, in the molecular, chemical and physiological contexts of a living plant. These networks include interactions involved in environmental sensing, signalling and gene expression, protein synthesis, protein interactions, including the nonspecific interactions of intrinsically disordered and chaperonin-like proteins, control of cellular organisation (for instance the transformation of chloroplasts to desiccoplasts in PDT plants), the regulation of biochemical pathways and the thermodynamic and biophysical effect of the resulting metabolite pools on the state of the physiological system.

The overall question, although asked of a biological system, is a question of thermodynamics and network theory: What is the difference between a biological network that, when perturbed by water loss, diverges towards disorder, and one that converges to an ordered stasis that, on reintroduction of the missing component, returns to its original functional state?

A model encapsulating the answer to this question would both reveal a great deal about the underlying biology of plant water relations, and have the predictive power to enable targeted bioengineering modifications aimed at making other organisms and systems more robust to water loss.

To construct such a model is a formidable undertaking, and requires a wide variety of inputs, including both comprehensive lists of network components (genes, proteins, metabolites etc.) and details of the nature, physical location and kinetics of the most relevant interactions. Such data requires both careful curation and rigorous statistical analysis to be useful.

This study aims to make a small contribution to such a future model by:

- Contributing to knowledge of changes to metabolite pools associated with desiccation in *X. humilis*;
- Optimising techniques for high-throughput analysis of metabolite extracts in this system;
- Exploring techniques for statistical analysis of high-throughput metabolite data and determination of peaks of interest;
- Beginning the process of peak identification for peaks of interest;
- Illustrating how such metabolite changes might be contextualised through the use of pathway databases such as KEGG.

This is intended to contribute toward an integrative approach in which such data may be meaningfully compared and computationally modelled along with those from other experiments.

Chapter Two: Materials and Methods

Plant Source and Maintenance

Xerophyta humilis plants were harvested from Pilansberg Nature Reserve, Northwest Province, South Africa, and maintained in trays in standard greenhouse conditions prior to experimental use as previously described ((Sherwin & Farrant 1996), (Dace et al. 1998)).

Plants were transferred to Percival growth chambers seven days prior to the start of each experiment, where they were maintained on a daily cycle comprising 12 hours light at 25 °C and 12 hours dark and 17 °C. The light intensity during the light phase was 227 μmol photons m^{-2} s^{-1}, delivered as a mixed spectrum comprising both incandescent and fluorescent sources. Trays were watered daily during this acclimation period.

After seven days of acclimation, plants were harvested for the "Hydrated" treatment at a water content deemed to be at full turgor (100% RWC). Thereafter, watering was stopped. For drying curve determination, leaves were harvested daily for water content determination. Samples in the "Dehydrated" treatment were harvested for metabolite analysis after leaf relative water content reached 5%. Ten plants were harvested per treatment, for each mode of analysis. All harvesting took place midway through the light phase of the daily cycle (i.e.

at simulated noon). Tissue (leaves or roots) was immediately frozen in liquid nitrogen, ground to a fine powder with a pestle and mortar and lyophilised. A portion (one or two leaves, or a similar mass of root tissue) of each sample was withheld from freezing and used for RWC determinations.

Changes in leaf relative water content during the full period of dehydration (from 100% to 5%) were followed by daily harvesting of randomly selected (n = 10) for RWC determination.

Water Content Determination

Relative Water Content of *Xerophyta humilis* samples was determined gravimetrically, using the standard method of our laboratory:

Tissue samples, were weighed on a Metler analytical balance, oven-dried at 80°C for 24 hours, and re-weighed. Absolute water contents were calculated as the mass of water lost as a fraction of the tissue dry weight. For expression as RWC, water content at full turgor was deemed to be the mean water content of the first set of tissue samples harvested at full turgor, immediately prior to cessation of watering. In the case of leaf samples (n=10), tissue samples consisted of whole leaves trimmed of their proximal, non-photosynthetic and distal necrotic tissue.

To assess the kinetics of desiccation in the growth chamber program used in this study, a drying curve was generated for a number of trays of *X. humilis* plants, each tray containing approximately 20 plants. It was observed that, owing to soil and root bed heterogeneity, drying rates exhibited a degree of variability both within and between trays, however all plants achieved a final desiccated water content of < 5% RWC within approximately 15 days of cessation of watering (Figure 4.1).

Metabolite Extraction

In an untargeted metabolomic study such as this, the goal is to survey as broad a range of cellular metabolites as possible, and to select the most relevant compounds through appropriate multivariate statistical techniques.

To achieve the goal of a broad, substantially unbiased survey of *X. humilis* metabolites, extraction protocols were developed to complement the analytical strengths of the primary chromatographic techniques. These protocols were not intended to be optimal for any specific compounds or conditions, but to provide a high-fidelity view of as broad a range of compounds as reasonably possible, and to provide data which could be meaningfully compared with that being generated in other laboratories.

Since gas chromatography(GC) is particularly well-suited to the separation (after derivatisation) of small carbohydrates and organic acids (i.e. most primary metabolites), extraction protocols for GC-based metabolomic experiments should select for these small, polar molecules while minimising both enzymic and non-enzymic modification to their nature or abundance.

C-18 Reverse Phase Liquid chromatography, by contrast, performs poorly at separating such highly polar compounds but is extremely capable at separating mixtures of mid- to low-polarity compounds, including terpenes, coumarins, flavonoids, alkaloids and many other plant secondary metabolites.

In each case, established protocols were adapted to the resources of our laboratory, the amounts of plant material available, and the configuration of our analytical instruments. Additionally, protocols were modified through the addition of a sonication step, since it has been suggested that this improves extraction efficiencies through enhanced swelling of the cellular matrix (Kim & Verpoorte 2010).

Both of the extraction protocols used require lyophilised, powdered plant tissue. In all cases, this was produced by grinding plant material (leaf or root) to a fine powder under liquid nitrogen, followed my freeze drying for 48 hours. Lyophilised samples were either used immediately or stored at -20°C prior to analysis.

Protocol A (High polarity, for GC-MS):

(Adapted from Roessner et al. (2000))

20mg lyophilised Xerophyta humilis tissue was suspended in 500µl analytical grade methanol, and heated at 70°C for 15 minutes. 500µl Milli-Q water was added, and the suspension sonicated for 15 minutes. 50µl of a 2mg/ml solution of ribitol was added to each extract as a GC internal standard. Samples were centrifuged at 13,000 x g for 15 minutes, and aliquots were decanted and dried *in vacuo* for derivatisation and analysis. Aliquot size was determined varied to correct for slight variations in the mass of lyophilised tissue extracted, with the nominal mass of 20mg corresponding to a decanted aliquot of 650µl.

Protocol B (Low polarity, for LC-MS and NMR):

(Adapted from t'Kindt et al. (2009))

20 mg lyophilised tissue was suspended in 640µl ice-cold analytical grade methanol, and incubated at 4°C for 15 minutes with constant mixing. 160µ ice-cold Milli-Q water added, and the samples were sonicated for 15 minutes. Tubes were centrifuged at 13,000 x g for 15 minutes, and 500µl aliquots were decanted for analysis.

GC-MS Experiments

The separation of analytes by gas chromatography requires that they be volatile, stable at the temperatures used for chromatography (up to 250°C for our method) and

sufficiently non-polar to achieve meaningful differential affinity with the column chemistry. To achieve this, samples were derivatised by alkylation and trimethylsilylation according to the method published by Roessner et al. (2000). Specifically, after extraction in 50% methanol, 650µl of each metabolite extract was decanted and dried *in vacuo*. The dried metabolite pellet was derivatised in 80µl of 20mg/ml methoxyamine hydrochloride in pyridine for 90 minutes at 30°C. This was followed by the addition of 80µl N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) and further derivatisation for 30 minutes at 37°C.

In the case of standard substances (see below), 5µl of a 10mg/ml aqueous solution of each standard compound was dried *in vacuo* and derivatised as above, except that 50µl of each derivatising agent was used.

All GC-MS hardware was supplied by Agilent. Experiments were performed on a model 7890A gas chromatograph equipped with a 7693 Autosampler, and interfaced to a 7000A Triple Quadrupole mass spectrometer.

The gas chromatography parameters were modified from those of Fiehn et al. (2000) with parameters empirically adapted for our instrument.

The GC injection port temperature was set to 250°C, the initial oven temperature to 70°C and the transfer line temperature to 200°C. The instrument was configured for a constant gas flow of 1.39 mL/min.

1µl of each sample was injected at a 10:1 split ratio. The oven temperature was maintained at 70°C for 30s, then ramped at 30°C/min to 150°C (reached 3.167 minutes after injection), and then at 5°C/min to 220°C (reached 32.167 minutes after injection). This temperature was held for 15 minutes. Analytes were separated on an Agilent HP-5ms column of 30m length and 250µm internal diameter. The HP5-ms is a low-polarity column

with a (5%-Phenyl)-methylpolysiloxane stationary phase suitable for high-resolution separation of TMS-derivatised compounds.

The mass spectrometer was configured for Electron Impact (EI) ionisation at -70 eV, with the mass analyser in single-quadrupole mode.

In addition to the *X. humilis* extracts, standard substances as listed in Table 5 were subjected to GC/MS analysis:

Table 5: Standard substances for peak identification

Amino Acids	**Sugars**
Valine	Arabinose
Serine	Fructose
Leucene	Fucose
Threonine	Galactose
Proline	Glucose
Glycine	Maltose
Alanine	Mannose
Methionine	Raffinose
Proline	Ribose
t-4-hydroxyproline	Sucrose
Phenylalanine	Trehalose
Glutamic Acid	Xylose
Asparigine	2-ketoglutarate
Glutamine	Fructose-6-phosphate
Histidine	
Lysine	**Sugar alcohols**
Tyrosine	D-arabitol
Tryptophan	D-Erythrose
Arginine	D-mannitol
Cysteine	D-sorbitol
Isoleucine	Galactitol
	Glycerol
Other Organic Acids	Iso-erythritol
Citric Acid	Maltitol
Fumaric Acid	Xylitol
Isocitric Acid	
Malic Acid	
Oxalic Acid	
Oxaloacetic Acid	
Succinic Acid	

LC-MS Experiments

For LC-MS experiments, a set of three quality control samples was generated by mixing equal volumes of each analytical sample, with the expectation that these samples should cluster near the origin of a PCA plot, since they are by definition average samples.

All LC-MS hardware was supplied by Agilent. HPLC separation of *X. humilis* extracts was performed on a 1290 Infinity HPLC system. Mass spectrometry was performed with a 6530 Accurate-Mass Q-TOF fitted with an Agilent JetStream Electrospray Ioniser (ESI).

Reverse Phase gradient chromatography was performed using an Agilent Zorbax Eclipse Plus C18 column of 2.1 x 50mm dimensions, and packed with 1.8µm media. A compatible guard column was fitted for all analyses.

The solvents used or the solvent gradient were: 97% water, 3% acetonitrile and 0.1% formic acid (Solvent A) and 90% acetonitrile, 10% water and 0.1% formic acid (Solvent B).

For chromatographic analysis, 5µl of each sample was injected into a 0.4ml/min solvent flow. A flow of 100% solvent A was held for 2 minutes, followed by an 18 minute gradient to 40% solvent B, and a 5 minute gradient to 100% solvent B. This flow was maintained for 5 minutes before being returned to 100% solvent A over a two minute gradient. The total gradient time was 32 minutes.

The column eluent was ionised in an Agilent JesStream ESI source, with a nozzle voltage of 1kV, desolvation gas flow of 8l/min at 300°C and sheath gas flow of 11l/min at 350°C.

For MS experiments, the spectrometer was used as a LC-TOF instrument, configured in positive ion mode. Accurate Mass spectra were collected in the mass range of range 200

to 1700 m/z, with reference masses of 121.0509 and 922.0098 used to maintain mass calibration. Spectra were recorded in profile mode.

For MS/MS experiments, the same TOF parameters were used, however the instrument was configured to automatically perform MS^2 scans by Q-TOF selection. Precursor ion selection was based on peak abundance, with the 5 most abundant ions over a threshold of 10,000 counts per second being subject to MS^2 scans. Candidate precursor ions were excluded from further scanning for 0.2 minutes after being scanned. MS/MS experiments were performed at collision energies of 5eV, 10eV, 15eV, 30eV and 50eV.

NMR Experiments

NMR experiments were used as a confirmatory technique to back up the identification of peaks in LC-MS experiments. Since ESI-TOF and ESI-Q-TOF spectra do not have the robust reproducibility of EI spectra, there are far fewer practical databases of LC-MS data than there are of GC-MS data. Despite the tremendous versatility, sensitivity and data-richness of LC-MS techniques, peak identification remains a challenge, so NMR provides useful additional lines of evidence to support peak identification.

In an ideal peak identification scenario, each LC peak would be subjected to both MS and NMR analysis. However such an approach can only meet the high-throughput, low-cost goals of an -omics experiment in laboratories equipped with integrated LC-NMR-MS workflows. At this time, such instrumentation is not available to our laboratory.

As an alternative approach, a variety of NMR spectra were collected from concentrated, pooled samples of hydrated and dehydrated *X. humilis* leaf and root tissue. Primary peak identification was performed on TOF and Q-TOF mass spectra. The presence

or absence of corresponding NMR signals at appropriate ratios was used to confirm putative compound identities.

To prepare metabolite samples for NMR, lyophilised, powdered leaf or root tissue was extracted in an 80% methanol / 20% water solvent system using the method described in "Metabolite Extraction". Supernatants were dried *in vacuo* overnight and redissolved in d4-methanol.

400 µl of each sample was subjected to NMR analysis, generating 1D ^{1}H, and 2D COSY, TOCSY and HSQC spectra. Parameters for ^{1}H spectra were as described in (Pauli & Jaki 2005) as listed in Table 6. TOCSY experiments were performed using an 80ms mixing time.

NMR was performed on a Bruker Avance 400MHz instrument, and spectra were analysed using MestReNova 8.0.1 software.

Table 6: Experimental NMR parameters

NMR Parameter	Value
Acquisition Time	2-4s
Relaxation delay	3-10s
Pulse width	15-45º
Time domain	64k
Spectral width	Sample ±3ppm
Transmitter offset	Centre of spectral width
Number of scans	128-1024

Chapter Three: Numerical Analysis

Data Pre-Processing

There is a substantial gap between the raw output from a mass spectrometer (ion counts over m/z and time) and the goal of chemometric statistics (comparison of compound quantities among biological samples). A number of data pretreatment workflows are possible to transform ion counts into relative compound abundances, but they must involve the assignment of chromatographic / spectral peaks to compounds (catering for possible coelution of multiple analyte species), the transformation of ion counts to compound quantities (in arbitrary but linear units), the normalisation of quantitative values to reflect relative abundance in source tissue, and statistical pretreatment suitable for the quantitative methods to be used.

In this study, slightly different pretreatment strategies were employed for GC-MS and LC-MS data, owing to the differences between the source data, and the use of an internal standard in GC-MS samples.

GC-MS Pre-Processing

GC-MS data consisted of gas chromatograms comprising EI-Quadrupole mass spectra of TMS-derivatised metabolites.

The limited mass accuracy, consistent ionisation pattern and relatively compact data size (compared with LC-MS output) lent itself to compound assignment by chromatogram deconvolution using the instrument vendor's proprietary software solution (MassHunter B.05, from Agilent). Chromatogram deconvolution was performed on all sample data files using a signal to noise ratio threshold of 2.00, an ion peak extraction window of -0.3 to +0.7 AMU, a sharpness threshold of 25%. Mass peaks were filtered by height, requiring a minimum height of 500 counts, and a relative height of at least 2% of the largest peak. Compound peaks were filtered by area, requiring an absolute area of at least 5000 counts, and a relative area of at least 1% of the largest peak.

Compound retention times, base peak m/z and peak areas were exported to Microsoft Excel, where compound alignment across samples was performed manually. A single matrix of peak area per compound per sample was exported for further statistical analysis in Matlab and R.

LC-MS Pre-Processing

LC-MS data consisted of HPLC chromatograms comprising ESI-TOF spectra of mid-polarity metabolites.

The high mass accuracy of this data, and its recording in profile mode led to very large data files being generated. These data volumes, coupled with the challenges of LC-ESI-TOF chromatographic and spectral reproducibility (see “Experimental Approaches”)

presented practical challenges with the use of Agilent Masshunter for chromatogram deconvolution.

To generate a matrix suitable for statistical analysis using PCA and PLS-DA, sample data files were exported from MassHunter in mzData format, and imported to MZmine 2.9.1.

The mzMine workflow comprised data import, baseline correction, mass detection, chromatogram building and deconvolution, deisotoping, peak alignment, gap filling, and export for statistical analysis.

Centering and Scaling

All compound scores were normalised to internal standards (in the case of GC values) and to the recorded mass of plant tissue represented per sample.

Data were mean-centered and scaled using Matlab R2013a. Scaling was performed using the Pareto algorithm as implemented in the libPLS library. The Pareto algorithm, shown below, was chosen to scale the majority of significant peaks while minimising the contribution of baseline noise to statistical calculations (van den Berg et al. 2006).

$$\widetilde{x}_{ij} = \frac{x_{ij} - \overline{x}_i}{\sqrt{s_i}}$$

The effect of this choice on the interpretability of Principal Component Analysis, in comparison with not scaling or autoscaling, is discussed and illustrated in Appendix A.

Statistical Treatment

Targeted Study

The statistical treatment of the compounds specifically targeted in the GC-MS analysis was relatively straightforward compared with that for the untargeted analysis.

Standard compounds that were successfully detected by GC-MS, with a discernible chromatographic peak and characteristic mass spectrum, were selected for targeting in the plant tissue samples, while those that were not successfully detected were omitted from the targeted study.

The selected standard compounds were sought in the plant sample data, matched by retention time and mass spectrum, and subjected to standard univariate statistical scrutiny. Their mean and standard deviation per population, and t-test significance are reported in the section “Targeted Compounds” in chapter 4.

Untargeted Study

One of the challenges of untargeted metabolomic analysis is the enormous number of variables generated by sensitive analytical techniques. In this study, deconvolution of GC-MS chromatograms or leaf tissue yielded a list of approximately 270 peaks. The MZMine

workflow generated a list of over 7000 peaks. Many of these software-detected peaks were doubtless instrument noise or represented compounds with no bearing on the biological question of vegetative desiccation tolerance. In addition, many of these peaks are intimately covariant (for instance, both fructose and glucose produce two peaks in this gas chromatography system, invariably in the same proportion).

The first statistical task then is to choose a subset of variables for further analysis that is enriched for variables of interest, and eliminates many irrelevancies.

To this end, Principal Component Analysis was performed as specified below. Since PCA produced clear clustering of hydrated and desiccated tissue based primarily on the first principal component, the use of supervised techniques such as PLS-DA was unnecessary, and the variables most heavily weighted in the first and second principal components were selected for closer scrutiny.

Principal Component Analysis

PCA was performed in Matlab using the libPLS package, and in R using native functions and the ChemometricsWithR package. Principal Components were generated by Singular Value Decomposition, and scores and loadings were plotted for principal components describing 95% of sample variability.

Variables with loading coefficient more than one standard deviation above or below the mean were selected for further investigation.

PLS Discriminant Analysis

Partial Least Squares Discriminant Analysis (PLS-DA) was performed on each dataset using the mixOmics package for R, version 5.0-1. Variable selection was performed using the splsda function, with models being constrained to the 50 most relevant variables per

component for the GC-MS leaf and LC-MS root datasets, 20 variables per component for the small GC-MS root data set and 200 variables per component for the very large LC-MS data set. Quality Control samples were omitted from LC datasets since they do not represent a true biological category of interest, and their role in assuring the data quality was fulfilled in the PCA score plots. Since all PLS-DA models produced clear, successful clustering of all samples in the first PLS component / latent variable, variables of interest were chosen for further, univariate analysis if they displayed a correlation greater than 0.5 on the first component axis.

Significance Tests and Abundance Ratios

Variables identified as being of interest by PCA or PLS-DA were subjected to Student's t-tests for significance, with those with $p < 0.05$ assessed for the quantitative change between hydrated and desiccated plants.

Key peak metadata, Inter-population mean ratios (I.e. n-fold change), and p-values are reported in section Appendix B.

Compound Identification

While some compounds of interest were identified by comparison with standard compounds, this approach was useful only for a subset of the GC-MS peaks. Several GC peaks, and all LC peaks, identified in the untargeted analysis, remained unidentified.

While it would be impossible, using current techniques, to identify all of these peaks using the data and time available, attempts were made to identify compounds of major interest using a combination of spectrum databases, QTOF MS^2 spectra and NMR spectra from pooled samples.

Comparison with Standard Compounds

The identification of compounds in GC-MS experiments by physical standards was straightforward. Since the standards were subject to chromatography in a modern instrument, on the same column, using the same parameters, and in the same batch runs as the biological samples, retention times were highly consistent across injections.

Using the retention times of the standard compounds in single-compound runs, peaks containing matching spectra were sought in the compound lists generated by chromatogram deconvolution in Agilent MassHunter. Integrated compound peak areas were used as numerical scored of compound prevalence.

Assigning compounds identities *from* known standards *to* unknown peaks avoided identification difficulties for the targeted compounds.

EI Spectrum Database Searches

Untargeted compounds scored as being of interest in the GC-MS samples were subject to database search in order to assign putative identities.

Compound spectra were extracted from chromatograms during the deconvolution process in MassHunter. Spectra were exported and the fractional-mass spectra generated by the mass spectrometer were rounded and consolidated using Microsoft Excel and a custom Perl program (see Appendix F).

The consolidated unit mass spectrum was searched against the Golm Metabolite Database (Hummel et al. 2010) at (http://gmd.mpimp-golm.mpg.de/).

Prospective identities for peaks without laboratory standards are reported in Appendix B.

ESI-MS Analysis: Accurate Mass and Molecular Formula

The LC-MS experiments yielded a substantial list of compound peaks of statistical interest. The identification of these compounds presented a bigger practical challenge than those yielded by GC-MS, since the ESI spectra yielded by this LC-QTOF instrument are less reproducible than EI spectra, are more subject to ion suppression and do not readily yield consistent fragmentation patterns.

Notwithstanding these weaknesses of the technique, the TOF mass analyser in the Agilent 6530 instrument reliably achieves high mass-accuracy m/z measurements. These accurate ion masses permit the inference of molecular or fragment formulas based on the mass defects of isotopes that may reasonably comprise the analyte molecule.

Such calculations (automated by MZmine software) generate a short list of possible molecular formulas for a molecular ion of specified accurate mass. Formula lists may be

trimmed by chemical plausibility, since the generating algorithm does not filter for such concerns as carbon valency and the nitrogen rule.

Compound Shortlist Generation

Candidate molecular formulas or accurate mass data were used to search the KEGG compound database (Kanehisa & Goto 2000), at (http://www.kegg.jp/). Results that were both chemically and biologically plausible were assigned as putative peak identities.

This approach is, of course, restricted to shortlisting known and well-characterised compounds. The characterisation of novel compounds would require an extensive process of chromatographic fractionation and multiple lines of evidence, such as UV-Visible spectra, mass spectra and NMR spectra. Such characterisations fall outside the scope of this project.

Q-TOF MS^2 Assessment

Where multiple plausible compounds were returned from an accurate mass or molecular formula search, the proposes compounds were compared against MS^2 fragmentation patterns for the corresponding LC peaks. Where possible, database searches were employed using METLIN or Chemspider to find similar MS^2 fragmentation patterns. Where such database entries were not available, peak MS^2 spectra were inspected for chemically plausible fragmentation schemes, and compounds discarded from the shortlist if they could not have feasibly generated the observed MS^2 spectra.

NMR Verification

Compounds that were sufficiently abundant to generate strong NMR signals, and which had NMR spectra readily available (for instance via Chemspider) were verified by NMR and are indicated as NMR-verified in the tables of results. In these cases, previously

published NMR signals for each compound were compared against spectra generated in this study. To be deemed verified, the characteristic peaks of a compound had to be visible in the spectra for *X. humilis* extracts, at a similar quantitative ratio between hydrated and desiccated samples to that derived from mass spectrometry.

Pathway Inference

As a first step towards a fuller model of metabolic interaction in *X. humilis* desiccation tolerance, a selection of the metabolites identified in this study were used to identify key pathways that may be regulated in the plant's response to water deficit.

A table of the relative abundances of the most confidently identified metabolites (those verified by laboratory standards) was generated for each measured sample. Pathway inference was performed using the R Pathway Activity Profiling (PAPi) module (Aggio et al. 2010).

Pathway inference results were manually filtered to exclude pathways not found in plants, since no automated mechanism was found to filter PAPi results based on kingdom-level classifications.

Chapter Four: Results and Discussion

Drying Kinetics

Under the experimental conditions used in this study, plants had reached a desiccated state within approximately two weeks of cessation of watering. The drying kinetics are illustrated in the Figure 4.1.

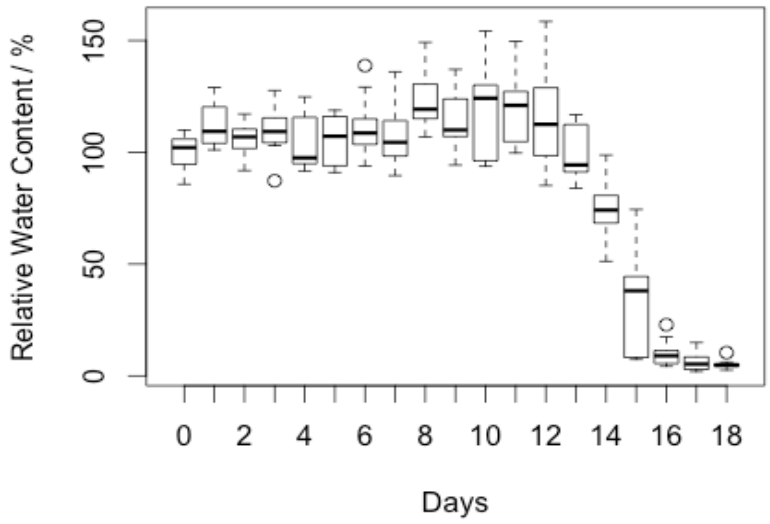


Figure 1. Relative Water Content of *X. humilis* leaves after cessation of watering on Day 0

While the primary purpose of this measurement was to ensure that a suitable and predictable timeframe to harvest desiccated tissue was achieved, the drying kinetics show some noteworthy features.

The most obvious feature of the plot above is the precipitous decline in Relative Water Content (RWC) after a long period of full hydration. This is consistent with suggestions that the plants employ a water conservation strategy until passing a threshold beyond which a strategy of rapid water loss is employed (Farrant et al. 2012). The second notable feature is the high degree of variability of leaf RWC during the period of rapid water loss between 12 and 15 days after cessation of watering, indicating a relatively high degree of biological variation typical of natural systems. This limits the utility of this experimental system (as-harvested *X. humilis* communities in plant trays) for the investigation of intermediate water contents, since the drying rate within and among such trays is so variable. A more reproducible and predictable plant culture system would be necessary to investigate the sequence of metabolome, proteome or transcriptome changes during dehydration. In the current study, tissue from only fully hydrated (100% RWC) and fully desiccated (5% RWC) plants was analysed for metabolic content.

Principal Component Analysis

In the untargeted portion of this study, the goal was to determine the compounds that are most strongly associated with desiccation in *X. humilis* and, where feasible, to identify these compounds. In massively multivariate datasets such as these, the major challenge in data analysis lies in reducing the overwhelming number of variables to a number that can be more easily analysed and visualised, with less scope for the inevitable false positives that arise from a probabilistic approach to significance. To this end, PCA analysis was performed on preprocessed mass spectral data from *X. humilis* root and leaf extracts, from both GC-MS and LC-MS datasets.

For each data set, the contribution of the principal components to total variability was plotted, and this contribution was used to choose the number of principal components to use in plotting the actual datasets. Standard PCA biplots (Gabriel 1971) were generated but where loading vectors obscure samples, these are represented in separate plots below. In order to further reduce the visual congestion of the plots, only the most prominent loading vectors are labeled. Loading vectors correspond to individual compounds, but are labeled anonymously, with labels referring to peak retention times, or IDs within an mzMine dataset.

Leaf Analysis by GC-Quadrupole

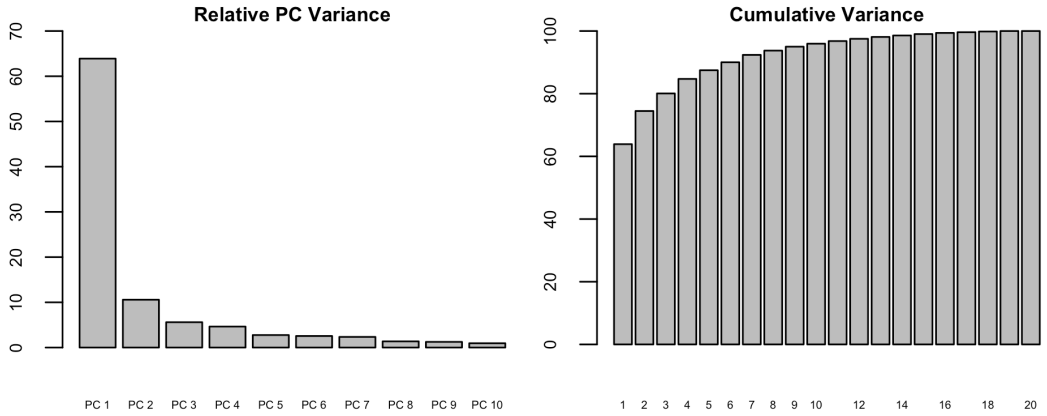


Figure 2. Relative and cumulative variance per Principal Component for GC-MS peaks of *X. humilis* leaf extract.

The variance within GC-MS derived data for leaf tissue is successfully described within the first few principal components (Figure 2). While the first nine components describe 95.01% of the variation within the sample population, only the first two principal components (Figure 3a) revealed relevant clustering of the sample populations. Nevertheless, principal components 3 and 4, which together with the first two PCs, describe 84.72% of the variability, are also plotted in Figure 3b for illustration and discussion.

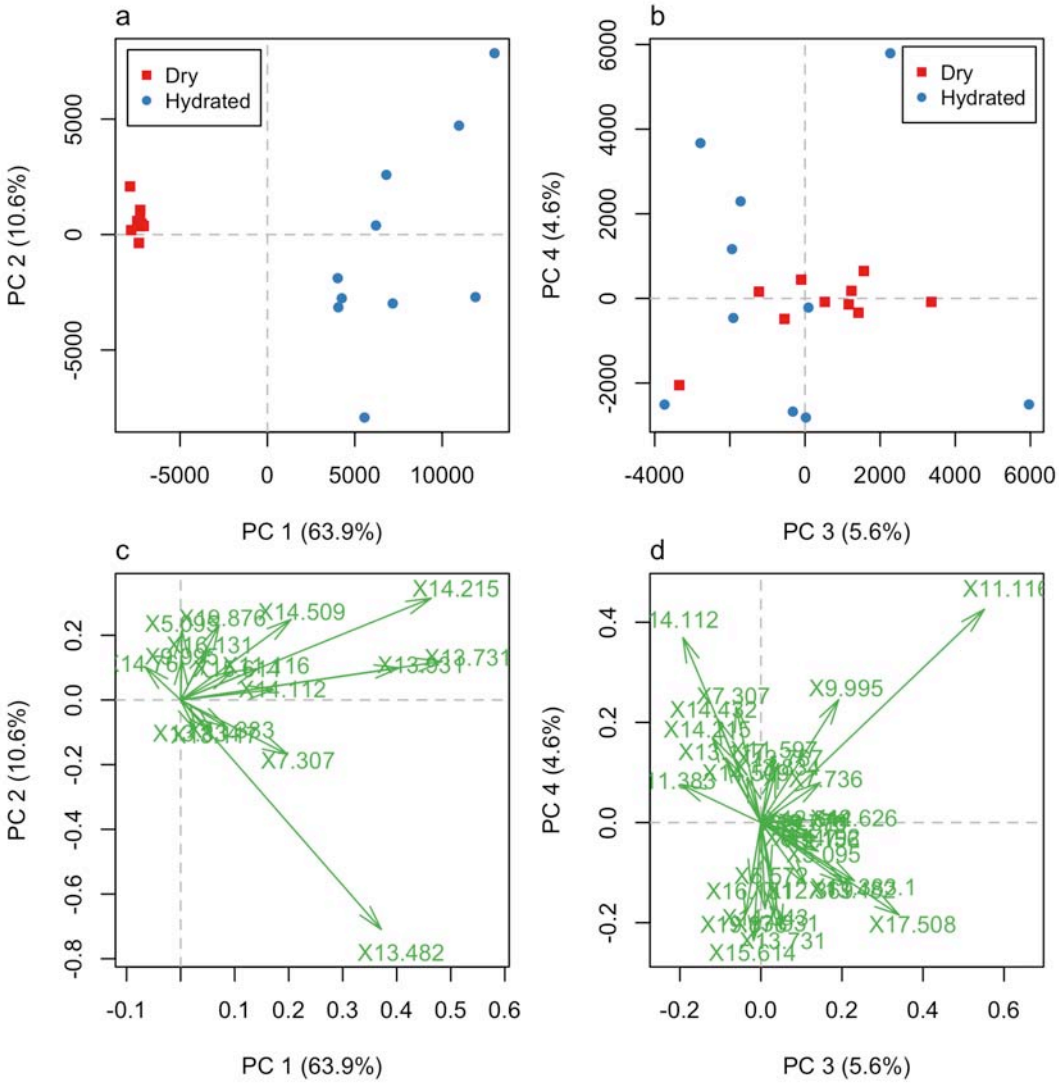


Figure 3: Score and loading plots for GC-MS leaf extract data. ***(a)*** **Scores of Principal Components 1 and 2,** ***(b)*** **Scores of Principal Components 3 and 4,** ***(c)*** **Loadings of Principal Components 1 and 2, and** ***(d)*** **Loadings of Principal Components 3 and 4.**

It is clear from these plots that the majority of the informative variation is described in the first component. None of others shown displays any helpfully predictive degree of clustering of plant treatments, with the possible exception of the relative paucity of hydrated samples from the first quadrant of the second plot, suggesting that a variable vector positively or negatively correlated with this quadrant may be informative.

According to the variable loadings in Figure 3c, the first principal component is positively associated with compounds at retention times 7.307, 11.116, 11.383, 13.117, 13.482, 13.731, 13.931, 14.112, 14.215, 14.509, 15.614, 16.131 and 19.876 minutes. Their association with a component direction so strongly associated with fully hydrated metabolism suggests that these compounds are likely to be reducing sugars or other abundant primary metabolites that would be deleterious on drying and are thus down-regulated in response to water loss. In addition, the compound at retention time 14.76 minutes has a negative coefficient on the first principal component, implying that it is more prevalent in the desiccated plants that cluster at a negative value on the first principal component.

The plot of the third and fourth principal components in Figure 3d, should be interpreted with some caution, however some features do stand out, namely the loading vectors for the compounds at retention time of 9.95 and 11.11 minutes. These compounds may be associated with desiccation, given the rarity of hydrated samples in the first quadrant of the score plot.

Root Analysis by GC-Quadrupole

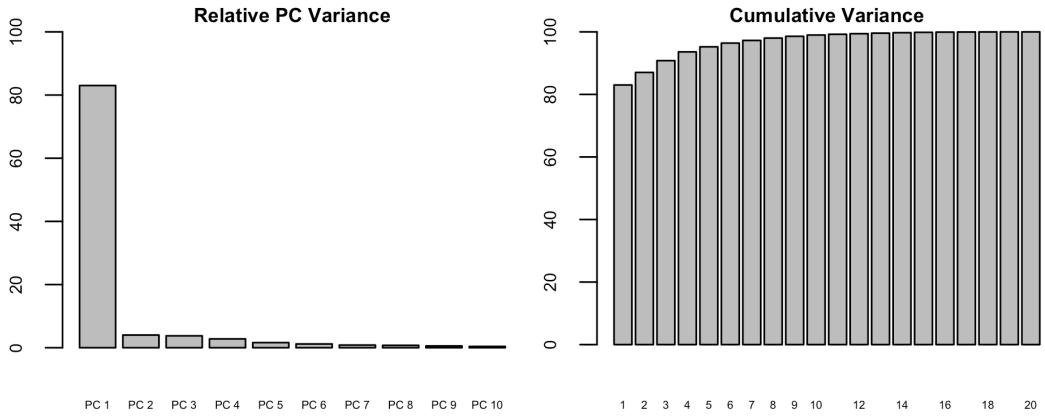


Figure 4: Relative and cumulative variance per Principal Component for GC-MS peaks of *X. humilis* root extract,

The first five principal components describe 95.22% of the variance in *X. humilis* root extracts analysed by GC-MS (Figure 4).

Despite the overwhelming dominance of the first principal component in describing sample variability (Figure 5a), the third principal component also displays some degree of clustering of sample groups, and is plotted in Figure 5b and 5c, in combination with the fourth and the first principal component respectively.

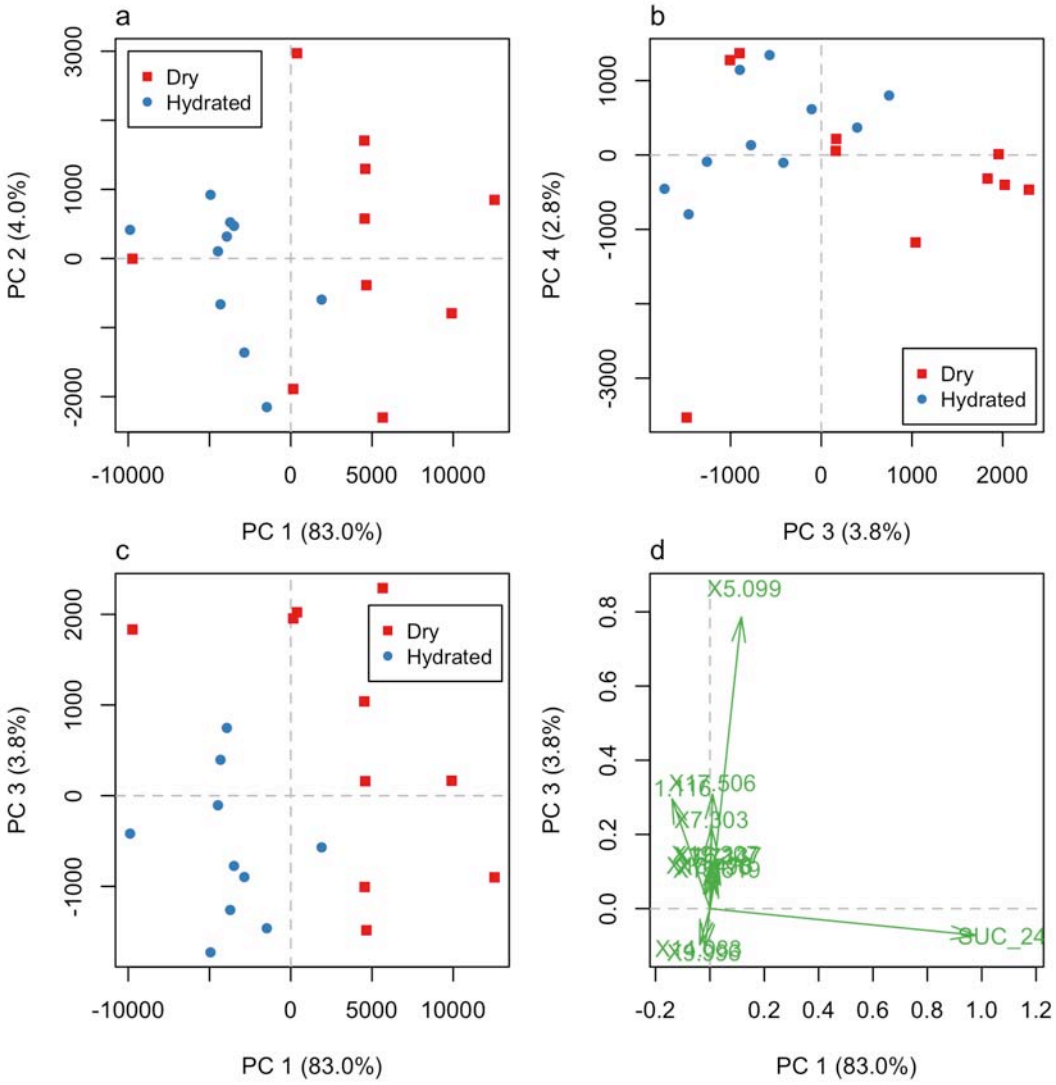


Figure 5: Score and loading plots for GC-MS root extract data. ***(a)*** **Scores of Principal Components 1 and 2,** ***(b)*** **Scores of Principal Components 3 and 4,** ***(c)*** **Scores of Principal Components 1 and 3, and** ***(d)*** **Loadings of Principal Components 1 and 3.**

Since the first and third principal components, in combination, appear effective in discriminating between hydrated and desiccated states, their combined loading vectors are plotted in Figure 5d.

In this analysis, the first principal component, which is positively associated with desiccation in the score plot, is completely dominated by the sucrose peak that elutes at 24 minutes. The third principal component is less strongly correlated with desiccation, but does show some correlation with desiccation when interpreted in conjunction with the first principal component (which is effectively the sucrose signal). This suggests that compounds eluting at 5.099, 11.116, 14.755, 17.506 and 19.207 minutes may be worth further scrutiny.

Leaf Analysis by LC-TOF

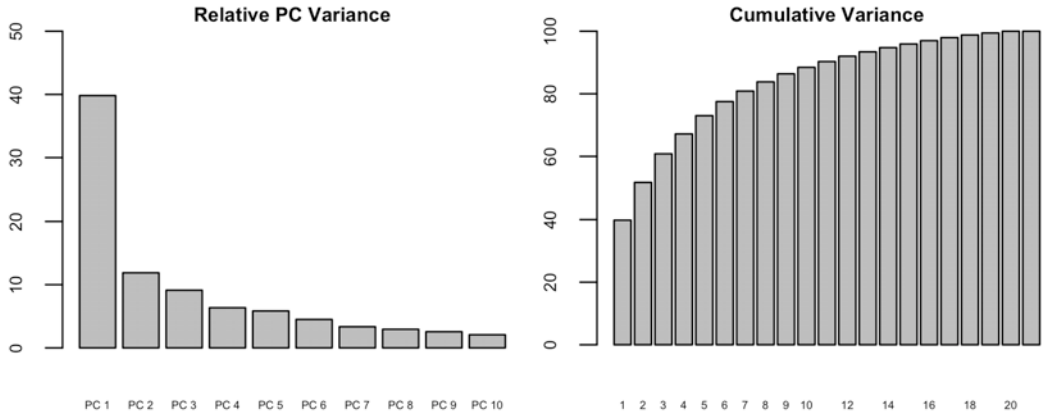


Figure 6: Relative and cumulative variance per Principal Component for LC-MS peaks of *X. humilis* leaf extract

LC-MS data for *Xerophyta* leaf extracts remains variable along numerous axes in the SVD-transformed data space, with the 15 principal components required to describe 95% of the sample variability, the first two together describing only 51.70%.

In practice, many of the components are not usefully informative about compounds related to the *X. humilis* desiccation response. Score plots are displayed for informative components (Figure 7a) , with an example of an uninformative combination included for illustrative purposes (Figure 7b).

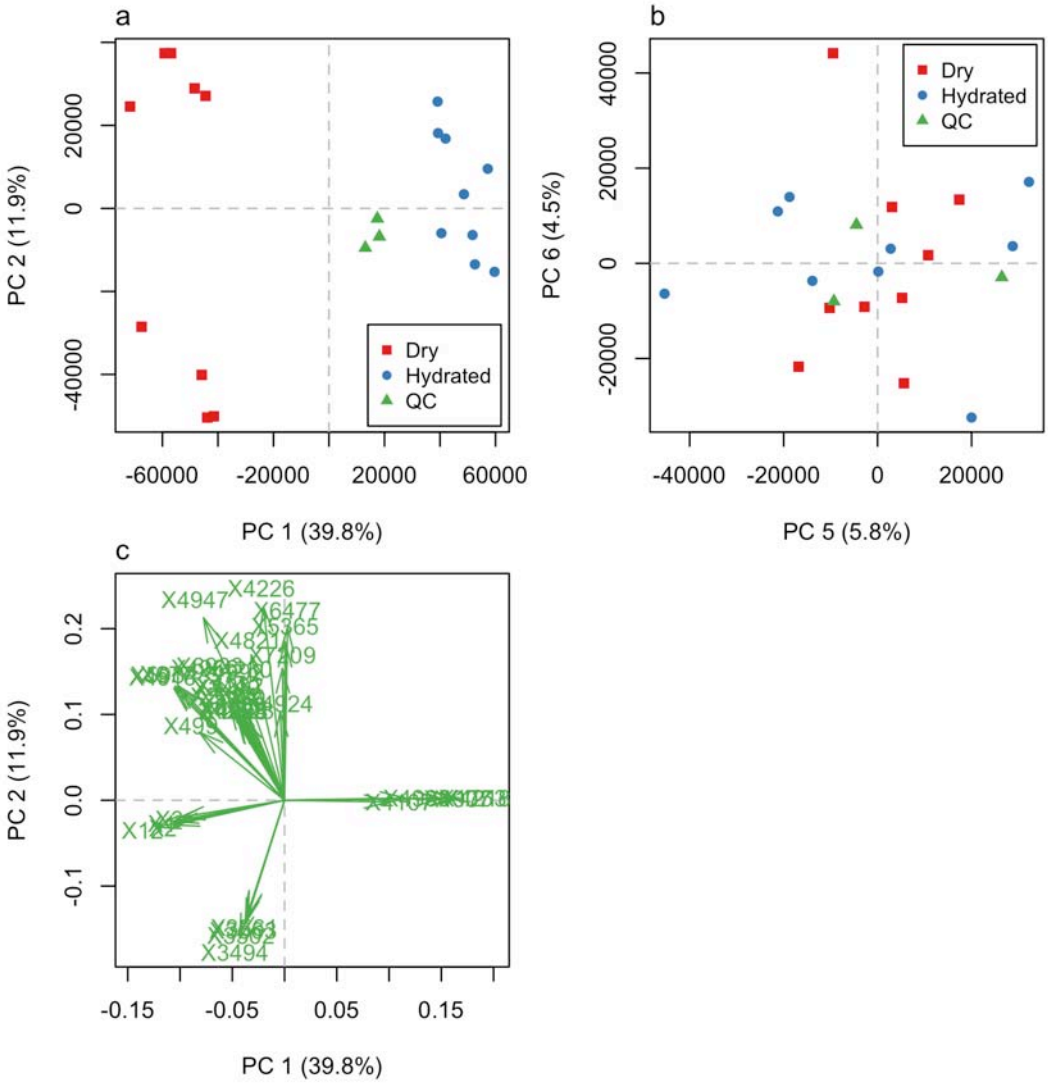


Figure 7: Score and loading plots for LC-MS leaf extract data. *(a)* Scores of Principal Components 1 and 2, *(b)* Scores of Principal Components 5 and 6, and *(c)* Loadings of Principal Components 1 and 2.

In Figure 7a, all hydrated plants cluster closely positive values for the first principal component. Quality control samples cluster close to the origin of the first plot, as expected, while desiccated plants are not well clustered on the second principal component, but are universally negative on the first. Thus PC1 is likely to strongly represent compounds that are clearly different between hydrated and desiccated plants, while PC2 is likely to represent compounds that are vital to normal cellular function but are highly variable in anhydrobiotic

tissues. An interesting feature of this plot is the division of dehydrated samples into two subpopulations, suggesting at least two possible clusters of secondary metabolites associated with viable anhydrobiotic states. It has been observed in the laboratory that viable desiccated *X. humilis* leaves may be either dark or light in colour, and it would be instructive to determine whether these visible traits are associated with the two chemical states observed here.

Figure 7b illustrates an uninformative pair of principal components. It shows no strong clustering of either plant population on either axis. In addition, the wide scattering of the QC samples, and their extreme values in comparison with the individual samples of which they are a chemical average, suggests that these principal components largely represent instrument noise and other sources of random variation within the data.

Since only the first pair of principal components appears to provide useful information about the *X. humilis* desiccation response, their loading plot is shown in Figure 7c.

The loadings on the positive horizontal axis, associated with the clearest clustering, suggest a positive association between compounds 1753, 4218, 4908, 4907 and 4107, and the hydrated, "normal" state of *X. humilis* leaves, and a negative association between the hydrated state and compounds 1, 2, 3 and 12. A number of compounds on the negative diagonal axis are also likely to play a role in desiccation, particularly to the extent to which they are associated with the first principal component. These include compounds 499, 677, 5934, 4916, 4947, 6933, 4966, 4821, 4992, 1232, 2419, 4230, 2884 and 5170. A number of compounds are positively (4226, 6477, 5365, 7209) or negatively (3494, 3502, 3563, 3561) associated with the second principal component. In this case, the compounds with positive coefficients are likely to be those necessary for normal metabolism but neutral with respect to anhydrobiosis, while those with a negative coefficient are likely to be those largely absent in

normal metabolism but sometimes elevated in desiccated leaves. Such compounds could conceivably be may be as breakdown products from drying-related damage, the oxidised form of antioxidants or other compounds indirectly associated with desiccation tolerance.

Root Analysis by LC-TOF

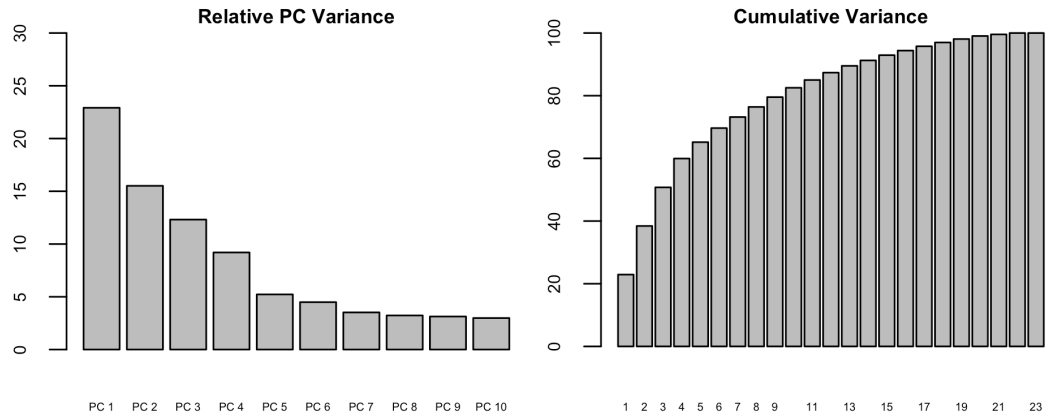


Figure 8: Relative and cumulative variance per Principal Component for LC-MS peaks of *X. humilis* root extract.

In the case of LC-MS data for *X. humilis* root compounds, numerous principal components are required to fully describe the variability of the data set. The first and second principal components together describe only 38.43% of the data variability, and seventeen principal components are required to achieve the targeted 95% coverage.

In practice, the later principal components are not informative, however score plots for the first six are shown in Figures 9.

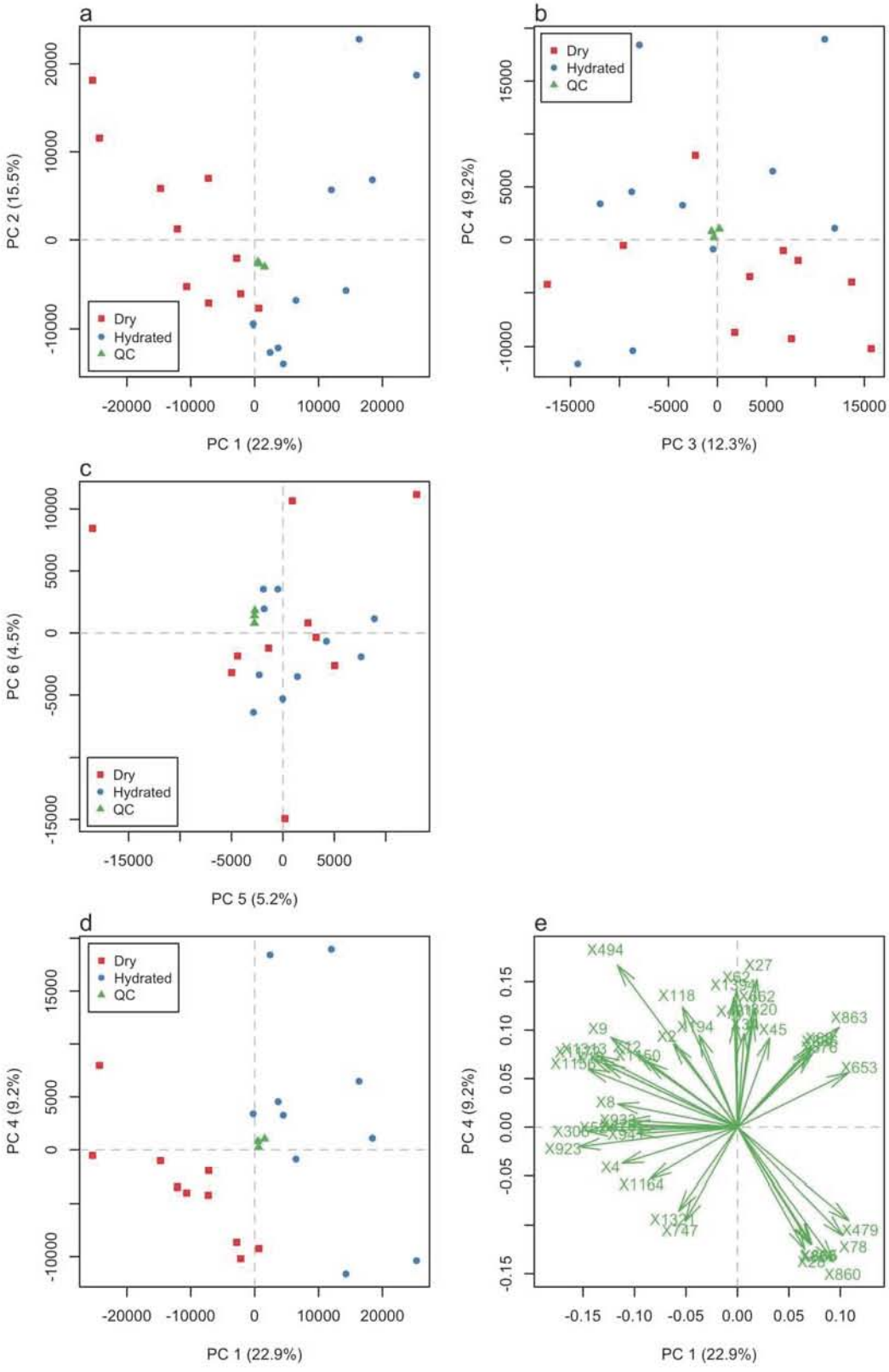


Figure 9: Score and loading plots for LC-MS root extract data. ***(a)*** **Scores of Principal Components 1 and 2,** ***(b)*** **Scores of Principal Components 3 and 4,** ***(c)*** **Scores of Principal Components 5 and 6,** ***(d)*** **Scores of Principal Components 1 and 4, and** ***(e)*** **Loadings of Principal Components 1 and 4.**

Since samples cluster sucessfully on both (and only) the first and fourth principal components, they are plotted together in Figure 9d. This plot shows a clear segregation of hydrated and desiccated root samples, with a tight clustering of QC samples inspiring confidence that these components are chemically meaningful. The loading plots for this combination of components is shown in Figure 9e.

Based on these scores, compounds positively associated with the first principal component are expected to be more strongly associated with normal, hydrated metabolism in *X. humilis* roots but those negative for both the first and fourth principal components to be associated with the dry state in roots.

Based on the visible clustering, a positive association with desiccation may be hypothesised for compounds 747, 1321, 1164, 4, 923, 52, 1156, 8, 1172, 306, 10, 933, 36, 941 and 476. A negative association with desiccation, and a positive association with normal metabolism may be hypothesised for compounds 27, 662, 1320, 45, 863, 653, 184, 475, 876, 856, 129, 69, 479, 78, 860, 865, 866, 28 and 195.

Partial Least Squares Discriminant Analysis

For each data set, individual sample plots and variable correlation plots (functionally corresponding the PCA score and loading plots in the preceding section) are shown in Figures 4.26 to 4.33 and interpreted below. In all cases, the majority of clustering is observed in the first component of the model, and only plots for the first two components are shown. It is noted that PLS-DA succeeded in highlighting a number of compounds that were not detected using PCA, but show significant changes on drying.

Leaf Analysis by GC-MS

Like PCA, PLS-DA satisfactorily clusters data from all four datasets in this study. The sample and variable correlation plots for metabolites in *X. humilis* leaves subjected to GC-MS analysis are shown in Figure 10. This regression used a maximum of 50 variables per component.

Based on the distribution of samples with respect to the PLS components, a number of compound peaks can be identified as being strongly correlated with either the desiccated or the hydrated state in *X. humilis* leaves. Compounds associated with hydrated and desiccated metabolic stated are listed in tables 17 and 18 in Appendix B.

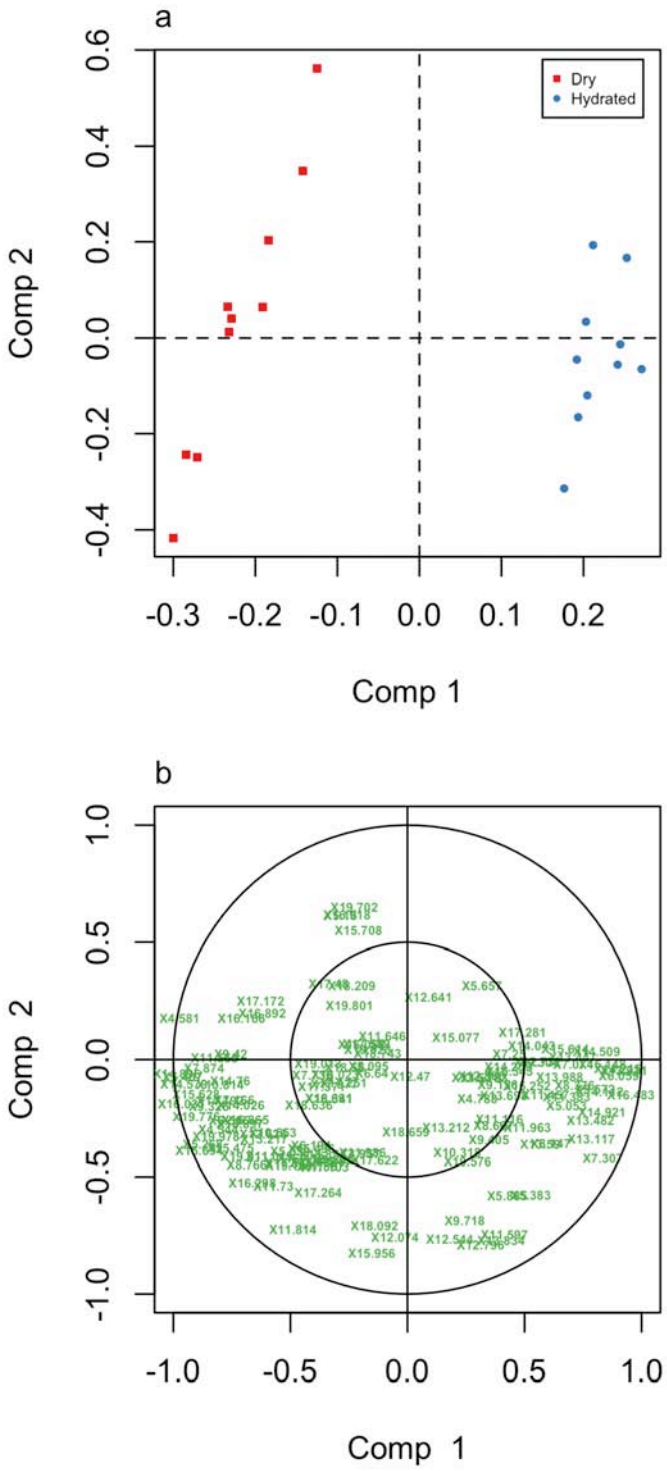


Figure 10: ***(a)*** **Sample scores, and** ***(b)*** **variable correlations of the first two PLS regression components for GC-MS peaks of** ***X. humilis*** **leaf extracts**

Root Analysis by GC-MS

Although root samples yielded relatively few peaks when analysed by GC-MS, treatments were successfully categorised by PLS-DA. The sample and variable correlation plots for metabolites in *X. humilis* roots subjected to GC-MS analysis are shown in Figure 11. This regression used a maximum of 20 variables per component.

GC-MS compound peaks strongly associated with the desiccated or hydrated states in *X. humilis* roots are listed in Tables 19 and 20 in Appendix B.

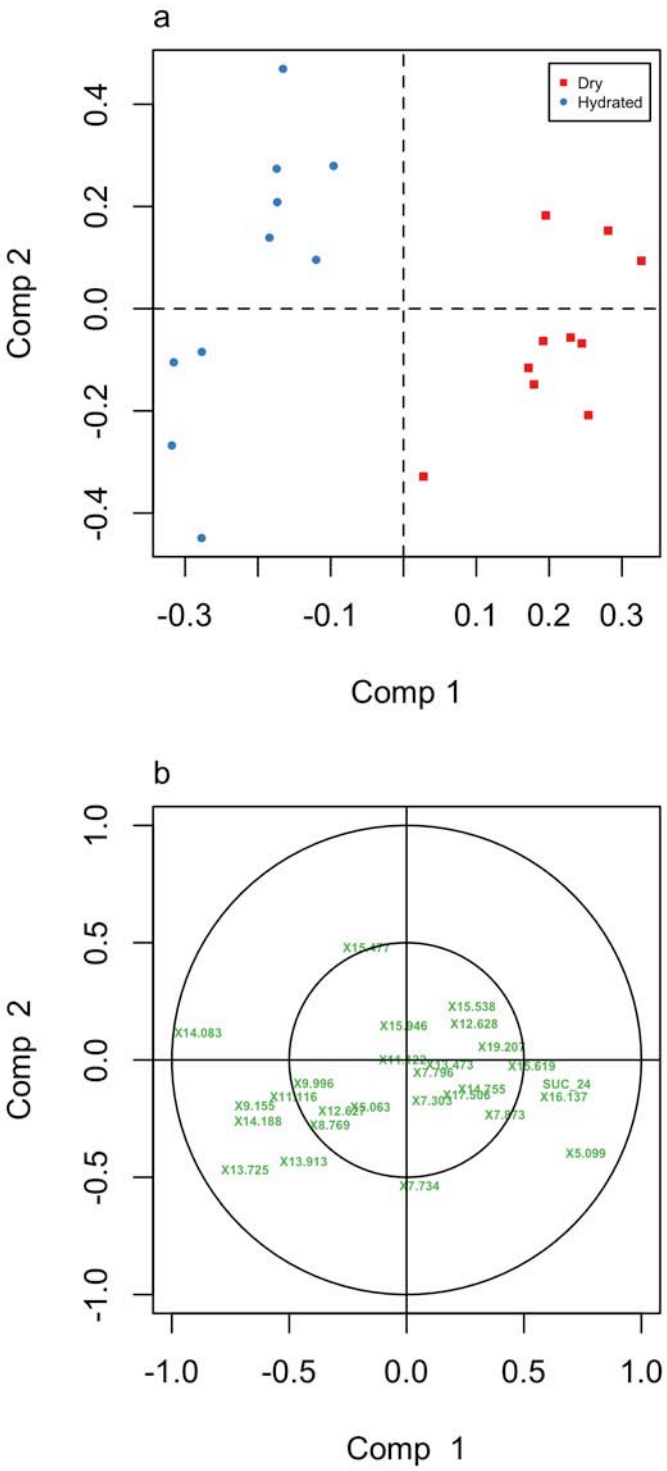


Figure 11: ***(a)*** **Sample scores, and** ***(b)*** **variable correlations of the first two PLS regression components for GC-MS peaks of** ***X. humilis*** **root extracts**

Leaf Analysis by LC-MS

LC-MS analysis of leaf extracts yielded a great many peaks. These were successfully categorised with sample and variable correlation plots shown in Figs 4.30 and 4.31 respectively. This regression used a maximum of 200 variables per component.

LC-MS compound peaks strongly associated with the desiccated or hydrated states in *X. humilis* leaves are listed in Tables 21 and 22 in Appendix B.

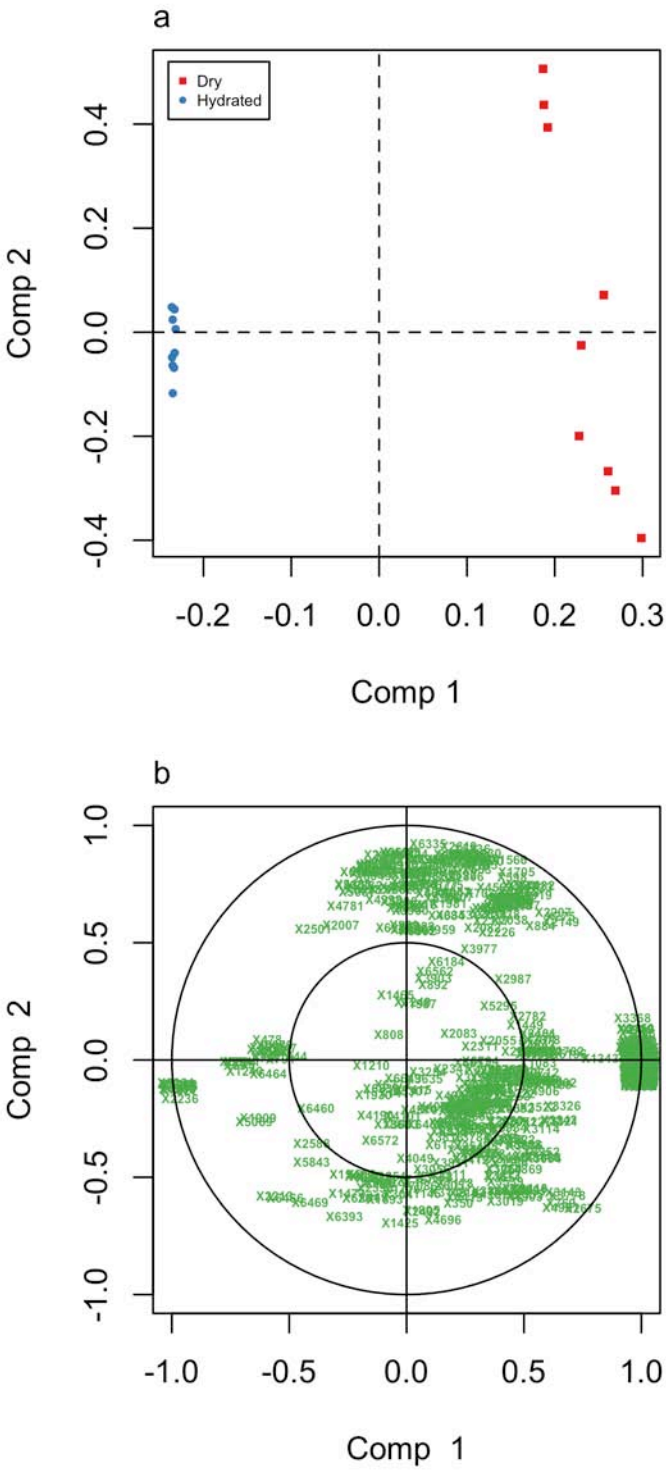


Figure 12: ***(a)*** **Sample scores, and** ***(b)*** **variable correlations of the first two PLS regression components for LC-MS peaks of** ***X. humilis*** **leaf extracts**

Root Analysis by LC-MS

The sample and variable correlation plots for metabolites in *X. humilis* roots subjected to LC-MS analysis are shown in Figure 13. This regression used a maximum of 50 variables per component.

LC-MS compound peaks strongly associated with the desiccated or hydrated states in *X. humilis* roots are listed in Tables 23 and 24 in Appendix B.

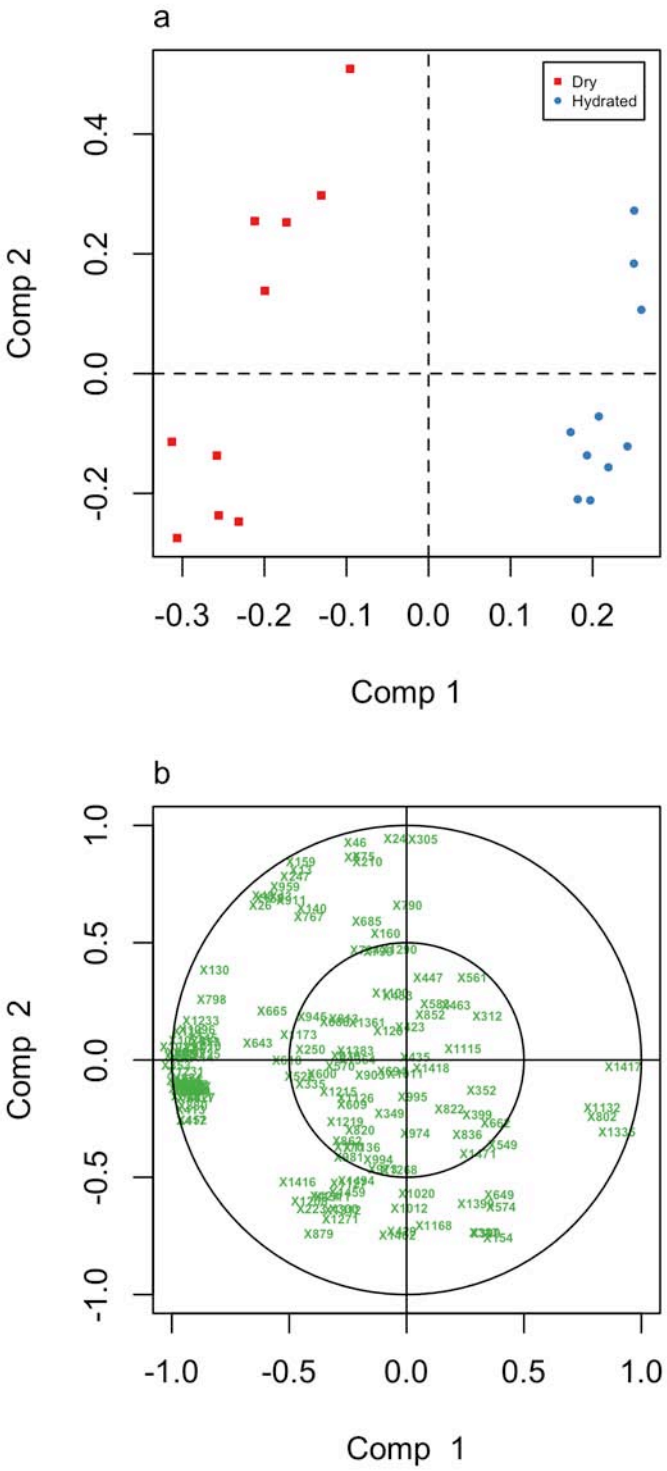


Figure 13: ***(a)*** **Sample scores, and** ***(b)*** **variable correlations of the first two PLS regression components for LC-MS peaks of** ***X. humilis*** **root extracts**

Targeted Compounds

Of the targeted compounds listed in section "GC-MS Experiments" in chapter 2, those which met the criteria of being successfully detected as a standard compound by GC-MS, and detected in more than half of either the hydrated or dehydrated *X. humilis* leaf or root samples, are given in Tables 7 to 13. The abundance ratio is supplied only in the case of compounds for which there is a statistically significant ($p < 0.05$ in a 2-tailed Student's t-test) difference in population means between hydrated and desiccated samples.

It must be noted that the quantification of sucrose presented a special problem given its extremely high prevalence in both hydrated and desiccated tissues of both leaves and roots. As a result of its overloading of the gas chromatography column its broad peak was not successfully deconvolved by MassHunter, and was quantified by manual peak specification within the software. The amount of sucrose present in the tissues also overloaded the ion detector, meaning that the tables below probably underestimate the magnitude of sucrose accumulation on drying.

Nevertheless, the measured change was unequivocal, and in agreement with numerous previous studies of both seed and vegetative plant desiccation tolerance (Reviewed in (Farrant et al. 2012)).

Compounds in Leaves

Leaves are chemically complex organs. Their photosynthetic capability, coupled with the fluctuations in light temperature and herbivory to which they are exposed, necessitates a complex, adaptable metabolic environment. This complexity resulted in the detection of approximately 200 distinct peaks in the GC profile, of which 23 represented compounds targeted by successfully resolved standards. Of these 23 compounds, 16 showed significant quantitative changes on drying, as shown in Tables 7 to 13.

Sugars

Table 7: Targeted sugars up-regulated on desiccation

Compound	p-value	Ratio
Arabinose	2.80×10^{-2}	1.30
Sucrose	5.03×10^{-4}	2.08[1]

Table 8: Targeted sugars down-regulated on desiccation

Compound	p-value	Ratio
Glucose	7.58×10^{-5}	20.54
Fructose	3.59×10^{-4}	229.49
Fructose 6 Phosphate	1.50×10^{-2}	6.97

1. Column and ion detector overloading may have led to this figure being underestimated.

Amino Acids

Table 9: Targeted amino acids up-regulated on desiccation

Compound	p-value	Ratio
Aspartic Acid	5.10×10^{-3}	2.35
Glycine	*[2]	
Phenylalanine	1.08×10^{-2}	3.50
Threonine	1.16×10^{-3}	2.03
Tryptophan	2.69×10^{-3}	4.73
Tyrosine	1.52×10^{-3}	15.28

Other Organic Acids

Table 10: Targeted organic acids up-regulated on desiccation

Compound	p-value	Ratio
Citric Acid	1.92×10^{-2}	1.80

Table 11: Targeted organic acids down-regulated on desiccation

Compound	p-value	Ratio
Malic Acid	4.08×10^{-5}	1.92
Succinic Acid	1.24×10^{-2}	1.39

Sugar Alcohols

Table 12: Targeted sugar alcohols up-regulated on desiccation

Compound	p-value	Ratio
Arabitol	7.57×10^{-4}	1.80
Mannitol	9.17×10^{-4}	2.00
Xylitol	4.86×10^{-2}	4.59

Table 13: Targeted sugar alcohols down-regulated on desiccation

Compound	p-value	Ratio
Galactinol	2.67×10^{-2}	9.01

2. Free glycine was detected only in desiccated samples, and was below the detection threshold in all hydrated samples.

Compounds in Roots

The metabolite profile of *X. humilis* roots was markedly different from that of leaves. Given their lower absolute water content and lack of photosynthetic activity, it is unsurprising that their metabolite profile is dramatically simper than that of leaves.

In the root samples, 31 compounds were detected by GC-MS, of which 8 corresponded to targeted standards. Of these eight targeted compounds, two were found only in hydrated roots, and only three others displayed a statistically significant change in abundance on drying (Table 14).

Table 14: Targeted compounds significantly regulated in roots

Compound	p-value	Ratio
Glycerol	3.15×10^{-5}	2.18-fold up-regulation on drying
Sucrose	5.21×10^{-3}	2.14 fold up-regulation on drying
Fructose	-	Only in hydrated samples
Glucose	-	Only in hydrated samples
Glutamic Acid	4.90×10^{-3}	2.35-fold down-regulation on drying

The sensitivity of root experiments may be constrained by the relative paucity of growing, metabolically active cells in root tissue. Analysis of samples enriched in cells from root tips and elongation zones may yield a more complex metabolite profile.

Untargeted Compounds

Peaks identified as being of interest by PCA or PLS-DA were tested for significance using Student's t-test. Those compounds for which the null hypothesis (that there is no real difference between the population means of signal intensity between hydrated and desiccated tissues) was rejected at the 95% confidence level, are listed in Appendix B.

Some notable summary statistics are presented in Table 15, with highlights illustrated in Figure 14. Figure 14a illustrates the relative success of PLS-DA and PCA at identifying peaks of interest. For these datasets, PLS-DA significantly outperformed PCA, particularly in the case of the LC-MS leaf analyses. In the case of the latter experiment, the large number of variables (7186 in this study) may limit the utility of matrix transformation like PCA in favour of a model regression like PLS. By contrast, PCA performed relatively well in simpler data sets, such as those generated for root extracts, with their smaller number of compound peaks. This illustrates the importance applying a variety of analytical tools to datasets.

Overall, more peaks were identified as being up-regulated than down-regulated on drying (Figure 14b). While this may to some extent be an artefact of the variable filtering in PLS-DA (see “PLS Discriminant Analysis” in chapter 3), it probably also represents an accumulation of numerous metabolites involved in supporting the anhydrobiotic state.

The relative abundance of compounds detected in leaves compared with roots was previously discussed (see "Targeted Compounds" in this chapter), and the untargeted analysis of the dataset reinforces this observation (Figure 14c). Leaves are more chemically complex than roots, and an equivalent mass of root tissue contains less metabolically active cytoplasm.

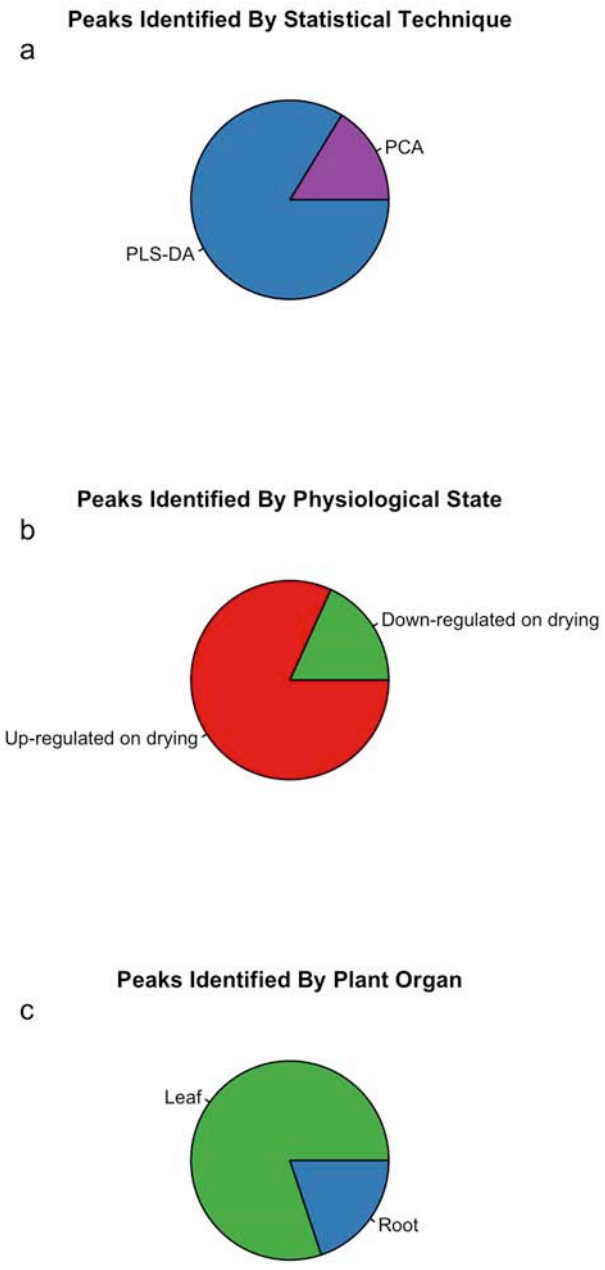


Figure 14: Untargeted compound peaks identified by *(a)* statistical technique, *(b)* quantitative response to drying, and *(c)* plant organ.

Table 15: Summary of peaks found per experiment

Experiment	Tissue	Condition	Peaks	PCA	PLS_DA	Both	p < 0.05
GC-MS	Leaf	Hydrated	246	13	29	10	28
GC-MS	Leaf	Desiccated	246	3	38	1	35
GC-MS	Root	Hydrated	28	5	4	0	4
GC-MS	Root	Desiccated	28	2	4	2	2
LC-MS	Leaf	Hydrated	7186	5	32	2	34
LC-MS	Leaf	Desiccated	7186	18	248	4	242
LC-MS	Root	Hydrated	1497	19	4	0	11
LC-MS	Root	Desiccated	1497	15	54	1	67

The compounds identified in these datasets both agree with existing literature, and present a large number of compounds either previously unreported in association with plant desiccation tolerance, or undescribed.

Comparing these data with those reported for the bryophyte *Selaginella lepidophylla* (Yobi et al. 2013) reveals both similarities and differences between the metabolomic water stress responses of the angiosperm *X. humilis* with the bryophyte *S. lepidophylla*.

The detailed compound lists are presented in Tables 17 to 24 in Appendix B. Only peaks with a p-value less than 0.05 are presented. Putative peak identities were assigned using the method described in section “Compound Identification” in chapter 3.

It is noteworthy that most of the peaks successfully identified in the GC-MS datasets were identified on the basis of locally run standards, although eight were putatively identified on the basis of spectrum searches of the Golm Metabolite Database. Database searches of LC-MS peaks provided some putative assignments, however accurate mass alone cannot distinguish among isomers, resulting in many ambiguous assignments. In addition, considerable subjective judgement was exercised in rejecting some assignments, while preferring others. This represents a potential source of error in experiments of this kind. The

use of NMR to resolve these ambiguities was constrained by the lack of broadly available annotation libraries of NMR data for secondary metabolites. These observations underscore the importance of developing extensive local standards libraries in a metabolomics laboratory.

Compound Pathway Inference

The KEGG pathway analysis described in “Pathway Inference” in chapter 3 yielded a list of metabolic pathways that were inferred to be differentially active between the hydrated and desiccated states in *X. humilis*. Inspection of these pathways allows the identification of possible metabolic control points that may be involved in the plant's response to water loss. Such control points may be hypothesised to show altered activity, either through protein modification or changed expression levels, in response to water stress.

This inference process is not, however, a replacement for a fluxome study, and the interpretation of the pathway list must be cautious. In the context of the overall metabolic network of the plant cell, the inference of specific pathway activity from the relative abundances of a small subset of metabolites can only generate tentative hypotheses, which should be subject to additional future verification. This point is underscored by the inclusion of many non-plant pathways, including "Long-term depression", "Type II diabetes mellitus" and "Cocaine addiction" among the inferred pathways. A list of 41 such non-plant pathways were removed from the list presented. In addition, an anhydrobiotic system is somewhat special since none of the inferred pathways is truly active in the quiescent state of a desiccated resurrection plant. Nevertheless, the pathways associated with the dry state may have shown enhanced activity during the drying process.

Figure 15 illustrates the proposed relative activity levels among the listed pathways. Note that three pathways (Purine metabolism, Lysine degradation and Thiamine metabolism) are omitted from the plot but since their absence from the inferred activities of the hydrated state led to anomalies in representing relative activities. The full set of inferred plant pathways is listed in Table 16.

A number of features in this data set stand out as interesting. These include the very high absolute activity levels of the ABC transporter pathway, particularly on drying. This accords with previous work suggesting large-scale transport of compatible solutes into vacuoles (Farrant 2000), as well as changes to the metabolite compositions of cell walls (Moore et al. 2008).

Carbohydrate mechanism features strongly among the differentially active pathways, with very large differentials seen in sucrose metabolism. This is unsurprising given the established role of sucrose as an osmoprotectact in desiccation-tolerant plants, as reviewed in (Farrant et al. 2012).

A number of pathways in the list appear counterintuitive, such as the low activity of tryptophan metabolism in drying plants compared with hydrated plants. However this appears to refer to a catabolic pathway in KEGG, and a reduction in tryptophan catabolism is consistent with the elevated levels of tryptophan observed in dry *X. humilis* leaves.

One interesting observation is in the frequency of secondary metabolite pathways in this table, including a number of alkaloid and terpenoid biosynthetic pathways. Although this PAPi query was based only upon standard-verified primary metabolites, their levels appear to have been predictive of the differential abundance of several secondary metabolites detected in the LC-MS experiments.

Although the output of this pathway inference algorithm is to be treated with circumspection, this list of proposed differential pathway activities can form a useful starting

point for identifying control points involved in switching the metabolic network in *X. humilis* to an anhydrobiotic state. They represent a set of testable hypotheses concerning the regulation of metabolism under water stress in *X. humilis*. The experiments required to investigate such hypotheses are beyond the scope of this project, however they could form part of a follow-up project that expanded upon these observations.

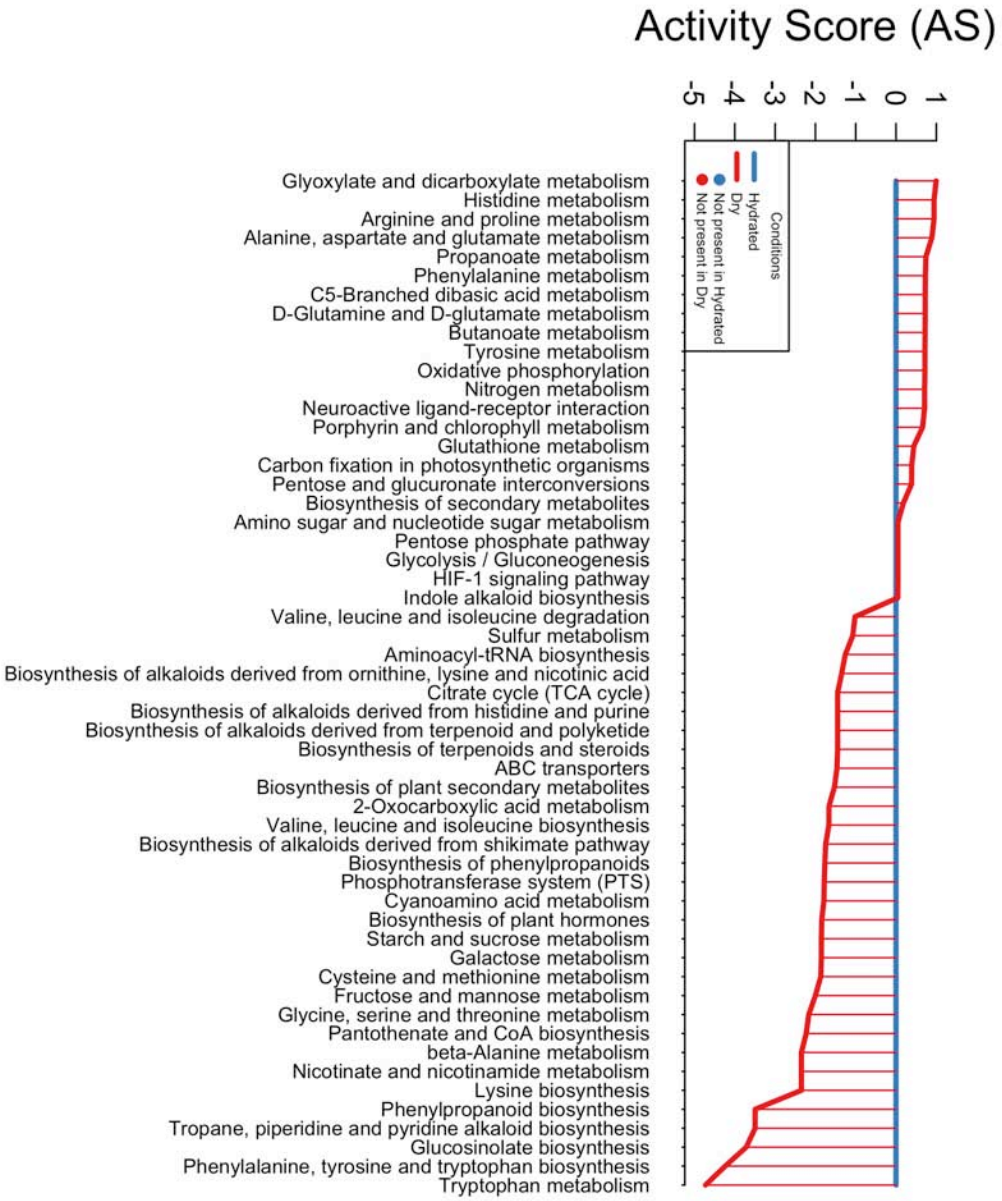


Figure 15: Relative activity scores of inferred pathways

Table 16: Pathway activities inferred from metabolite data.

Pathway Name	Activity: Dry	Activity: Hydrated	p-value
2-Oxocarboxylic acid metabolism	3341103	2009233	4.46×10^{-3}
beta-Alanine metabolism	12296631	5237953	2.62×10^{-3}
Biosynthesis of alkaloids derived from histidine and purine	2067418	1427827	3.29×10^{-2}
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	3198465	2371871	3.73×10^{-2}
Biosynthesis of alkaloids derived from shikimate pathway	2727979	1563326	4.16×10^{-3}
Biosynthesis of alkaloids derived from terpenoid and polyketide	2067418	1427827	3.29×10^{-2}
Biosynthesis of phenylpropanoids	2684317	1520418	4.06×10^{-3}
Biosynthesis of plant hormones	3080983	1672488	1.62×10^{-3}
ABC transporters	2358277637	1609408568	4.68×10^{-2}
Biosynthesis of plant secondary metabolites	4074732	2667036	6.00×10^{-3}
Biosynthesis of secondary metabolites	5194827	28235821	5.09×10^{-5}
Biosynthesis of terpenoids and steroids	2067418	1427827	3.29×10^{-2}
Butanoate metabolism	17724108	24545648	1.55×10^{-2}
Carbon fixation in photosynthetic organisms	6675883	17088877	3.43×10^{-2}
Carbon fixation pathways in prokaryotes	47550603	32840026	3.29×10^{-2}
Citrate cycle (TCA cycle)	20674175	14278272	3.29×10^{-2}
Cyanoamino acid metabolism	7978890	4459290	1.70×10^{-3}
Cysteine and methionine metabolism	14813733	7933964	5.51×10^{-3}
Fructose and mannose metabolism	91265938	45609562	1.93×10^{-4}
Galactose metabolism	5264946979	2842642785	1.72×10^{-2}

Figure 4.37: Relative activity scores of inferred pathways

Glucosinolate biosynthesis	16836380	4529361	4.61×10^{-3}
Glutathione metabolism	12770126	28901280	7.02×10^{-3}
Glycine, serine and threonine metabolism	12712432	5873727	1.79×10^{-3}
Glycolysis / Gluconeogenesis	38581026	792632329	3.20×10^{-5}
HIF-1 signaling pathway	18668238	383531772	3.20×10^{-5}
Indole alkaloid biosynthesis	39607560	878619990	9.20×10^{-4}

Lysine biosynthesis	14279958	6082784	2.62×10^{-3}
Lysine degradation	5676641	NA	NA
Amino sugar and nucleotide sugar metabolism	120721275	2480172126	3.20×10^{-5}
Nicotinate and nicotinamide metabolism	18643279	7941412	2.62×10^{-3}
Oxidative phosphorylation	5357970	7467558	4.51×10^{-3}
Pantothenate and CoA biosynthesis	6164577	2765962	2.26×10^{-3}
Pentose and glucuronate interconversions	10423147	27039495	3.37×10^{-2}
Pentose phosphate pathway	41070124	843769898	3.20×10^{-5}
Phenylalanine, tyrosine and tryptophan biosynthesis	9260277	2184786	4.35×10^{-3}
Phenylpropanoid biosynthesis	10067728	2879626	2.00×10^{-3}
Phosphotransferase system (PTS)	3778573905	2122757721	1.09×10^{-2}
Porphyrin and chlorophyll metabolism	34181571	51883671	4.33×10^{-2}
Propanoate metabolism	6813621	9173344	6.03×10^{-3}
Purine metabolism	11111723	NA	NA
Starch and sucrose metabolism	6549080389	3535970294	1.72×10^{-2}
Thiamine metabolism	3140270	NA	NA
Tropane, piperidine and pyridine alkaloid biosynthesis	12447373	3560265	2.00×10^{-3}
Tryptophan metabolism	38564453	8156893	1.00×10^{-2}
Tyrosine metabolism	25450356	35470902	4.51×10^{-3}
Valine, leucine and isoleucine biosynthesis	2283015	1370584	8.36×10^{-4}

Metabolite Changes in the Context of Plant Desiccation

The set of desiccation-associated metabolites identified in *X. humilis* by this study includes carbohydrates, sugar alcohols, amino acids, other organic acids including TCA cycle intermediates, nucleotides and numerous secondary metabolites, some of which have been previously described in association with desiccation tolerance, but others are novel in this context.

Sugars

In both root and leaf tissues, large reserves of sucrose accumulated on drying, with the monosaccharides glucose and fructose, abundantly present in the hydrated state, being reduced to trace levels on desiccation. This observation is compatible with many other published observations of desiccation-tolerant vascular plants ((Whittaker et al. 2001), (Peters et al. 2007)). By contrast, non-vascular desiccation tolerant plants such as the moss *S. lepidophylla* maintain constitutively high levels of sucrose, presumably because the drying rate they experience does not permit sufficient accumulation on drying (Yobi et al. 2013). The reduction in abundance of glucose and fructose, both reducing sugars, presumably protects against the nucleophilic reactions that could occur between their aldehyde groups and the amino groups present in proteins and amino acids in the molecular crowding of low water contents. A modest (1.3-fold) increase in the abundance of arabinose on drying is

consistent with other species (Farrant et al. 2009), and with the reported accumulation of arabinans in the cell walls of some resurrection plants (Moore et al. 2006) and (Moore et al. 2008), including *X. humilis* (Moore et al. 2012).

The untargeted data also show a reduction in the abundance of 2-ketoglutarate on drying (See Table 17 in Appendix B). This may be a reflection of diminished flux through the TCA cycle as the cellular matrix transitions to an anhydrobiotic state.

Sugar alcohols

X. humilis up-regulates arabitol, mannitol and xylitol on drying, while galactinol is more abundant in the hydrated state. The up-regulation of arabitol and mannitol have been reported in several other resurrection plants ((Oliver et al. 2011), (Hoekstra et al. 2001)) and (Vicré et al. 2004). The decrease in galactinol abundance is consistent with that observed in *M. caffrorum* (Farrant et al. 2009) but presents an interesting contrast with transcriptomic results. *X. humulis* up-regulates the mRNA transcript for one galactinol synthetase gene while down-regulating another (Collett et al. 2004). One can speculate that the up-regulated transcript may be stored for translation on rehydration (Dace et al. 1998), or else that the pool of galactinol is depleted by the biosynthesis of other raffinose-family oligosaccharides. This question could potentially be resolved through the acquisition of metabolite flux data, or further experiments on the biophysical state of quiescent mRNAs, for instance via immunoprecipitation and/or ultracentrifugation.

Amino Acids

Several amino acids are up-regulated on drying in *X. humilis*, including glycine, aspartate, phenylalanine, threonine, tryptophan and tyrosine. In addition, the untargeted LC-MS data suggests downregulation of valine, ornithine and lysine on drying. While this set of

amino acids is markedly different from those reported to be up-regulated in *S. lepidophylla* (Yobi et al. 2013), the up-regulated group is strongly enriched with cyclic and aromatic (phenylalanine, tryptophan and tyrosine) and hydroxylated (tyrosine and threonine) amino acids. The aromatic amino acids have also been observed to accumulate in response to other physiological stresses, such as temperature stress ((Guy et al. 2008), (Kaplan et al. 2004)); where they appear to be involved in the up-regulation of downstream phenylpropanoid secondary metabolites, several of which are also observed to be up-regulated in dry *X. humilis*.

Simple Organic Acids

The composition of the set of simple organic acids in *X. humilis* changes markedly on drying, with a 1.8-fold up-regulation in citrate, and down-regulation of malate and succinate being down-regulated 1.9- and 1.4-fold respectively. In addition, the untargeted peak analysis has identified the up-regulation of malonate and down-regulation of maleate and erythronic acid drying. This change in the organic acid profile may be related to the accumulation of protective ionic liquids (Choi et al. 2011) in components of the cell, however this intriguing hypothesis has yet to be demonstrated in desiccation tolerant plants, and the change in organic acid composition may be a by-product of the sequence in which primary metabolic enzyme activities attenuate during drying.

Nucleotides

While nucleotides were not specifically targeted with standards, the untargeted LC-MS data shows significant increases in abundance of flavin mononucleotide (FMN), adenine, adenosine, AMP, guanine and guanosine on drying. Of this set, FMN may play a role as an

antioxidant and may serve to protect chromatin and other cellular components from light damage.

Secondary metabolites

A variety of secondary metabolites are statistically associated with desiccation in *X. humilis*, including phenylpropanoids, alkaloids and terpenoids. Down-regulated peaks include magnesium mesoporphyrin, which is presumably associated with the catabolism of chlorophyll a in poikilochlorophyllous plants (Tuba & Lichtenthaler 2011). This peak may in fact be a degraded chlorophyll peak, but this can only be confirmed with additional chemical standards.

Up-regulated peaks include betaine (220-fold up-regulation) which has been reported in numerous desiccation-tolerant organisms (Hoekstra et al. 2001) and (Ingram & Bartels 1996); and other environmental stress responses (Bohnert et al. 1995). Several pigments are up-regulated in leaves on drying, including the carotenoids ionone and 15,15'-dihydroxy-beta-carotene, and a flavonoid ambiguously identified as either xanthohumol or cristacarpin.

Consistent with the up-regulation of aromatic amino acids, a number of phenylpropanoids are up-regulated on drying, including a peak ambiguously identified as either neochlorogenic acid (caffeoyl quinic acid) or scopolin (scopoletin glucoside) up-regulated in roots. A number of similarly but ambiguously identified peaks are up-regulated in leaves, as is quinic acid itself. This group of compounds is closely related to the a polyphenol found to be critically involved in the stabilisation of the cellular matrix -- and vacuoles in particular -- in the resurrection plant *Myrothamnus flabellifolia* (Moore et al. 2005). This suggests a potentially vital role for these compounds that would benefit from further investigation.

A variety of up-regulated alkaloids in leaves, including elaeocarpidine, cochlearine and emetamine, among several others, may be serve to provide additional defense against herbivory during periods of environmental stress, but may also have additional protective mechanisms within the plant itself.

Some compounds associated with the plant antioxidant system were found to be up-regulated in leaves on drying, including vanillic acid and oxidised glutathione. This observation is consistent with many studies (reviewed in (Farrant et al. 2012), (I. Kranner, & S. Birtić, 2005) and (I. Kranner et al. 2006)) that confirm the importance of antioxidants in water stress responses in general, and desiccation tolerance in particular.

Conclusion and Future Prospects

This study set out to achieve a number of objectives. These included the development and refinement of metabolomic techniques combining GC-MS, LC-MS and NMR analyses, the application of appropriate statistical methodologies to complex metabolomic data sets, identification of differentially abundant compounds, the assignment of chemical identities where possible, and the interpretation of these data in the light of the broader literature, and with a view to the generation of testable hypotheses.

These objectives have been successfully achieved, making this the first high-throughput metabolomic study in this laboratory, and the first reported for an angiosperm resurrection plant, the only other such study having been achieved on the fern-ally, *S. lepidophylla* (Yobi et al. 2013). The protocols developed are anticipated to be of value in future projects on *X. humilis* or other resurrection plant species.

The metabolite extraction and derivatisation protocols have proven effective, with good consistency among samples within each treatment group. Derivatisation for gas chromatography was effectively reproducible, although it may be helpful in future projects to derivatise under an inert atmosphere (Yobi et al. 2013).

Chromatographic methods achieved good separation, although additional refinement may be needed to improve the quantification of sucrose, since its late elution and abundance

contributed to poor quantification in GC-MS peaks. This could be corrected in future by running a separate, smaller injection of each sample on a faster temperature gradient specifically to improve the resolution of this peak.

The pre-processing of spectral data was effective and consistent, and the parallel use of PCA and PLS-DA as numerical techniques proved effective. PLS-DA proved particularly useful for extracting peaks of interest to this study, particularly from the LC-MS data sets, however the particular utility of a supervised technique in this instance does not negate the advantages of unsupervised methods in general. There remains considerable scope to explore additional statistical learning techniques for megavariate analysis, with various forms of cluster analysis likely to prove useful in larger studies assessing more species or treatments.

The greatest challenge for metabolomics projects in general lies in the chemical identification of peaks. To this end, the use of analytical standards remains the most effective approach, with database searches useful but not at the same level of confidence. This is likely to remain the case until considerably broader annotation of plant metabolomes is achieved. For this reason, future studies in this laboratory would benefit from the use of more standards, or the construction of an in-house library of MS and MS^2spectra for EI and ESI ionisation modes. The initiation and ongoing development of such a library should be made a priority. Nevertheless, 57 peaks were assigned single putative identities, and an additional 17 were assigned ambiguous identities on the basis of accurate mass searches. Several of these compounds had not been previously documented in resurrection plants and warrant further investigation in future.

Despite the usefulness of libraries, the sheer diversity of plant secondary metabolites means that some identifications are still likely to require traditional, "low-throughput" semi-preparative fractionation and investigation through multiple lines of evidence (melting points, UV-Vis, IR and NMR spectra). Such investigations lie beyond the scope of this project but

should be considered for compounds that appear to be of special interest (such as LC-MS peak 4787, which is up-regulated over 1000-fold on drying with a p-value of 6.58×10^{-8}, yet remains unidentified).

The use of statistical tools to propose testable hypotheses has been explored through the use of PAPi to infer the relative activities of various metabolic pathways from the relative metabolite abundances in hydrated and desiccated plants. The inferred pathways represent a set of hypotheses that could be tested directly through stable- or radioisotope labelled fluxomic experiments, or indirectly by correlation with transcriptome or proteome data. The explicit determination of pathway activation in response to water stress, and of the control points involved in such pathways, would represent a major additional step toward both the construction of a systems model of desiccation tolerance, and the possible application of resurrection plant biology in the development of stress-resistant agricultural species.

Numerous biological questions arise from the observations and inferences reported. These include the temporal sequencing of the *X. humilis* water stress response, and the biological role of a number of the metabolites identified. The overall biochemical kinetics of the response, the interaction of multiple pathways and their control may be clarified through the use of labeled flux experiments, and the pathways' control potentially elucidated through statistical modeling of gene expression, protein expression, protein modification and protein-protein interaction studies. These experiments may also resolve questions concerning which metabolite changes are intrinsically adaptive, and which are side effects of other pathway activities. The specific, physical roles of many metabolite species have not been fully understood, and biophysical studies on the thermodynamics of the cellular environment, and on chemically simulated biological mixtures, may help to resolve questions about the physical role of metabolites in protecting cellular and macromolecular stability. Examples of such studies could include investigating the possible structural protective effect of ionic

liquids comprising sugars and organic acids, using Differential Scanning Calorimetry, Infrared Spectroscopy and NMR.

In conclusion, this study has successfully pioneered the use in this laboratory of metabolomic techniques on resurrection plants. The data produce both substantially agree with other published data on resurrection plant metabolites, and advance understanding of desiccation tolerance through revealing the differential regulation of metabolites not previously known to be associated with desiccation tolerance. Finally, a set of testable hypotheses has been proposed that could form the basis of future projects.

Appendices

Appendix A: Selection Among Scaling Algorithms

In the chapter on Numerical Analysis, it was noted that for metabolomic applications, Pareto scaling is often preferred to either no scaling or autoscaling. This approach was validated for these data sets, with PCA being performed on the GC-MS data set for leaf extracts using each of the three scaling functions mentioned. Biplots plots generated for these three approaches were instructive and worth discussing. These are shown below. In each case, red circles represent desiccated plants, green circles hydrated plants, and blue vectors variable loadings. Loadings are not labeled in these examples since peak identities are irrelevant to the point illustrated in this appendix.

No Scaling

In cases where no scaling is applied, more abundant peaks are preferentially loaded in the principal components, effectively leaving less abundant compounds statistically undetected (Figure 16).

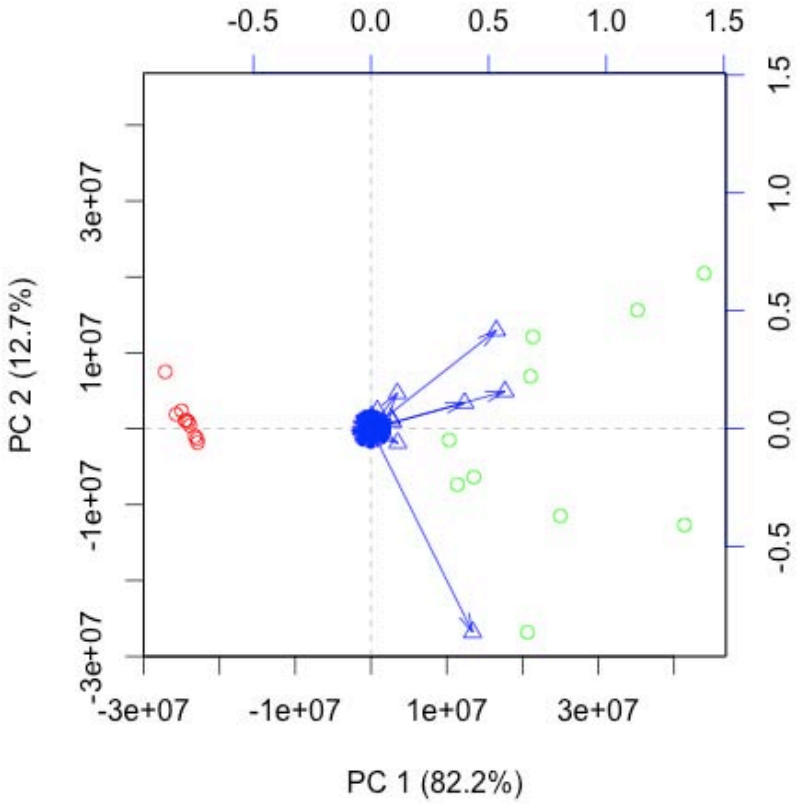


Figure 16: PCA with no data scaling

Autoscaling

In an attempt to make less abundant peaks more prominent, it is tempting to apply autoscaling, in which the values for each variable (i.e. chromatographic peak areas) are divided by the standard deviation of that variable. The effect of this is to equalise the variability of each peak.

Although this is an effortlessly standard function in all modern numeric software packages, it is not always suitable for chemometrics, since all information about the relative abundances of signals is lost, and relatively uninformative signals may be amplified more than is desired. A PCA biplot plot of an autoscaled version of the previous dataset is shown

in Figure 17. It will be clear to the reader that this does not provide any practical utility to the metabolomist.

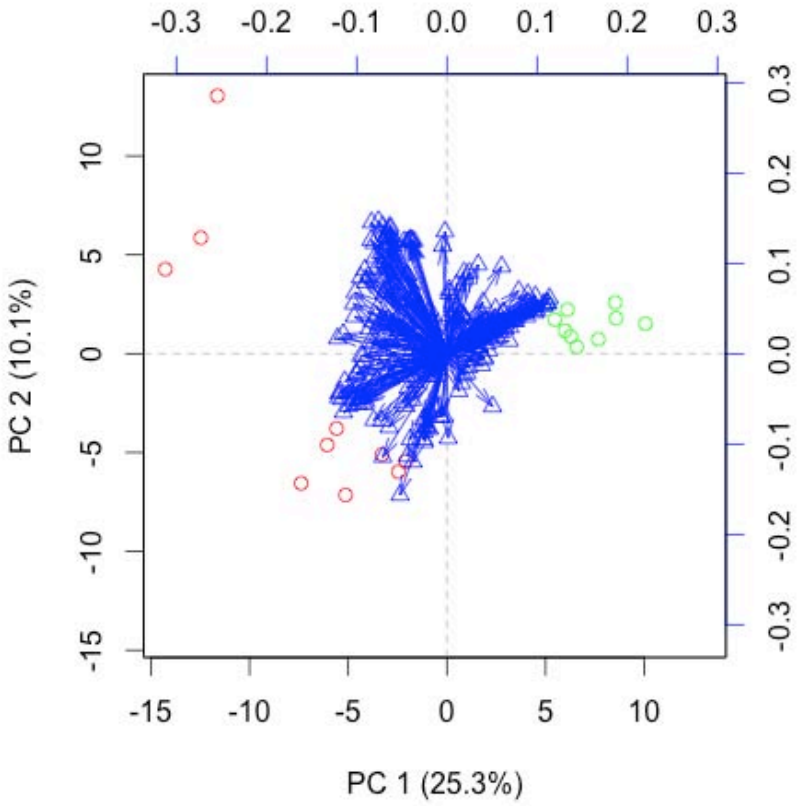


Figure 17: PCA with autoscaled data

Pareto scaling

If chemometric analysis of metabolomic datasets is to be useful in identifying relevant biomarkers other than the few most highly abundant compounds, a different approach to scaling is required. Such an approach needs to allow the contribution of smaller peaks to increase, while continuing to neglect baseline noise. One such method, commonly known as Pareto scaling, is to divide the values for each variable by the square root of the standard deviation of that variable. This tends to bring the magnitudes of each column closer together,

while leaving the contributions of very-small-valued variables (baseline noise) miniscule. The overall effect is that of a mild scaling, of "teasing out" the contribution of smaller peaks from the noise on the chromatographic baseline. The result of this technique is shown in Figure 18.

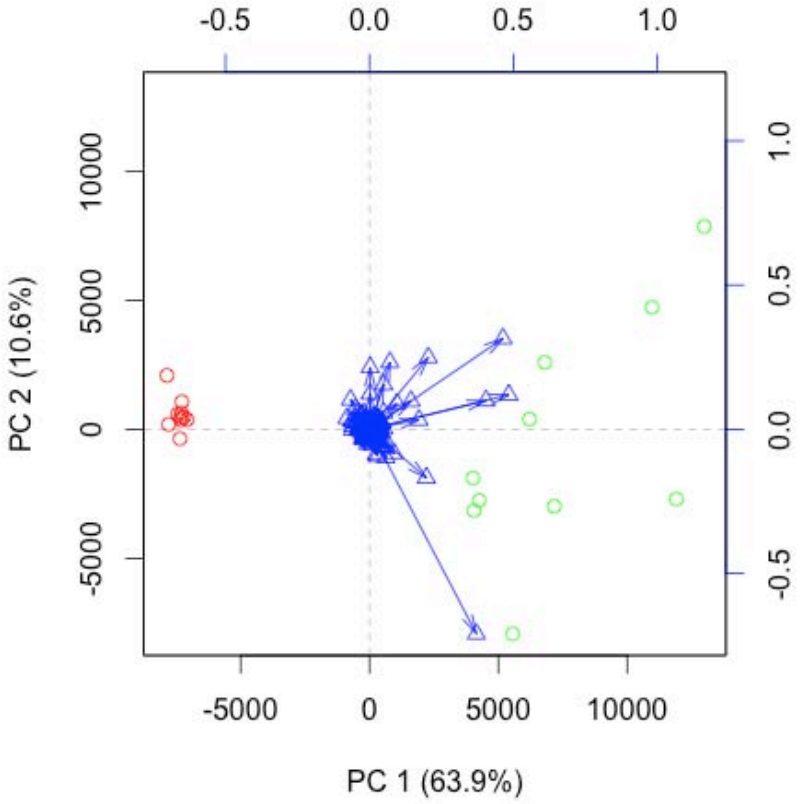


Figure 18: PCA with Pareto-scaled data

Appendix B: Untargeted Peak Lists

Tables 17 to 24 below list statistically significant peaks found in the untargeted portion of the study, with putative compound identities where possible. Identity confidence is indicated by colour. Peak identities in green represent compounds verified by standards analysed on the same instruments, by MS2 or by corresponding NMR peaks. Peak identities in blue represent compounds identified only by database searches. Peaks identities in orange have been verified by NMR and/or MS^2 spectra.

Untargeted GC-MS Peaks

GC-MS Peaks in Leaves

Table 17: Peaks predominantly observed in the hydrated state in leaves

RT / min	Found by	p-value	Ratio	Identity
5.053	PLS-DA	2.58×10^{-3}	1.67	
5.383	PLS-DA	4.51×10^{-3}	1.39	
5.572	PLS-DA	3.55×10^{-4}	13.76	
6.059	PLS-DA	1.80×10^{-5}	1260.51(*)[3]	Maleic Acid
7.074	PLS-DA	1.56×10^{-3}	5.89	
7.307	Both	4.81×10^{-7}	1.92	Malic Acid
8.147	PLS-DA	2.51×10^{-3}	1.78	Erythronic Acid
8.376	PLS-DA	1.84×10^{-3}	2.67	
9.304	PLS-DA	2.96×10^{-2}	265.57(*)	
11.383	Both	1.66×10^{-3}	185.59	
11.457	PLS-DA	1.20×10^{-2}	5.75	Vanillic Acid
11.963	PLS-DA	1.70×10^{-2}	430.61(*)	

3. *Ratio somewhat arbitrary since compound not detected in dry state.

RT / min	Found by	p-value	Ratio	Identity
12.732	PLS-DA	2.91×10^{-2}	1531.71(*)	
13.117	Both	1.05×10^{-4}	7357.68(*)	2-Ketoglutarate
13.447	PLS-DA	3.19×10^{-3}	56.22(*)	
13.482	Both	1.32×10^{-4}	3.90	Quinic Acid (^{1}H δ=2.048)
13.59	PLS-DA	7.97×10^{-3}	430.28(*)	
13.731	Both	6.82×10^{-6}	327.72	Fructose
13.931	Both	6.36×10^{-6}	188849.30(*)	Fructose
13.988	PLS-DA	5.73×10^{-3}	1513.93(*)	
14.043	PLS-DA	4.59×10^{-2}	3824.34(*)	Glucose
14.112	Both	1.17×10^{-4}	42.85	
14.215	Both	2.22×10^{-5}	255687.80(*)	Glucose
14.509	Both	4.72×10^{-4}	22.47	Glucose (alt. deriv.)
14.921	PLS-DA	1.09×10^{-4}	20.11	
15.148	PLS-DA	1.09×10^{-4}	4.44	
15.614	Both	6.84×10^{-3}	12820.76(*)	Glucose (alt. deriv.)
16.483	PLS-DA	1.48×10^{-7}	3196.62(*)	

Table 18: Peaks predominantly observed in the desiccated state in leaves

RT / min	Found by	p-value	Ratio	Identity
4.544	PLS-DA	2.41×10^{-4}	1.50	Malonic Acid
4.581	PLS-DA	1.29×10^{-11}	587.34(*)	
5.268	PLS-DA	1.79×10^{-4}	2.03	Threonine
5.407	PLS-DA	2.28×10^{-6}	1207.80(*)	Glycine
6.505	PLS-DA	1.57×10^{-3}	2.61	Aspartic Acid
7.166	PLS-DA	6.18×10^{-4}	2.75	Aspartic Acid (alt. deriv.)
7.874	PLS-DA	1.36×10^{-6}	5.95	Glycine (alt. deriv.)
8.766	PLS-DA	8.72×10^{-3}	1.60	
9.326	PLS-DA	2.36×10^{-4}	36.63	
9.42	PLS-DA	2.51×10^{-4}	2.72	
10.653	PLS-DA	2.80×10^{-2}	1.30	Arabinose
10.814	PLS-DA	7.57×10^{-4}	1923.45	
11.019	PLS-DA	2.50×10^{-2}	2.47	
11.555	PLS-DA	2.79×10^{-4}	5665.03	

RT / min	Found by	p-value	Ratio	Identity
11.73	PLS-DA	4.17×10^{-2}	1.26	
12.626	PLS-DA	7.36×10^{-3}	1.80	Citric Acid (^{1}H δ=2.54)
13.217	PLS-DA	8.83×10^{-3}	3.19	
14.026	PLS-DA	4.84×10^{-3}	1505.90	
14.192	PLS-DA	1.01×10^{-3}	38.27	
14.579	PLS-DA	3.07×10^{-6}	25.37	
14.76	Both	1.93×10^{-4}	2.00	Mannitol
14.896	PLS-DA	2.26×10^{-7}	3571.82	
15.475	PLS-DA	1.77×10^{-3}	4.48	
15.628	PLS-DA	1.22×10^{-5}	10.16	
15.694	PLS-DA	3.85×10^{-5}	3.53	
16.037	PLS-DA	2.76×10^{-7}	12.75	
16.166	PLS-DA	1.42×10^{-3}	634.70	
16.298	PLS-DA	1.34×10^{-2}	1.95	
16.865	PLS-DA	5.51×10^{-3}	77.50	
16.892	PLS-DA	6.40×10^{-3}	11.29	
17.172	PLS-DA	3.61×10^{-3}	428.18	
18.897	PLS-DA	2.01×10^{-3}	5.54	
19.011	PLS-DA	8.44×10^{-3}	7.66	Tryptophan
19.776	PLS-DA	7.91×10^{-5}	6103.65	
19.978	PLS-DA	1.96×10^{-4}	8.45	

GC-MS Peaks in Roots

Table 19: Peaks predominantly observed in the hydrated state in roots

RT / min	Found By	p-value	Ratio	Identity
13.725	PLS-DA	1.97×10^{-2}	930.72	Fructose
14.083	PLS-DA	3.95×10^{-5}	2519.87	Glucose
14.188	PLS-DA	1.82×10^{-2}	1061.90	Glucose
9.155	PLS-DA	4.90×10^{-3}	2.35	Glutamic Acid

Table 20: Peaks predominantly observed in the desiccated state in roots

RT / min	Found By	p-value	Ratio	Identity
5.099	Both	3.15×10^{-5}	2.18	Glycerol
15.619	PLS-DA	3.38×10^{-2}	1498.23	
16.137	PLS-DA	2.02×10^{-3}	3.06	Hexadecanoic Acid
24	Both	5.21×10^{-3}	2.14	Sucrose

Untargeted LC-MS Peaks

LC-MS Peaks in Leaves

Table 21: Ions predominantly observed in the hydrated state in leaves

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
2236	PLS-DA	0.76	118.0858	2.31×10^{-10}	2.99	Valine
2213	PLS-DA	0.89	146.0829	6.12×10^{-3}	1.42	4-Acetamidobutanoate OR 6-Amino-2-oxohexanoate OR (S)-5-Amino-3-oxohexanoic acid OR L-2-Aminoadipate 6-semialdehyde OR 4-Guanidinobutanoate
6304	PLS-DA	8.29	223.2173	3.05×10^{-3}	2.02	
6464	PLS-DA	8.29	133.0869	1.88×10^{-2}	1.60	Ornithine
3511	PLS-DA	8.29	241.1826	3.28×10^{-3}	2.04	
79	PLS-DA	8.29	223.1534	1.58×10^{-3}	2.21	
89	PLS-DA	8.29	240.1801	3.84×10^{-3}	2.07	
1240	PLS-DA	8.29	240.3234	2.86×10^{-3}	2.22	
477	PLS-DA	9.37	289.1609	1.39×10^{-2}	2.20	
835	PLS-DA	9.37	103.0750	1.97×10^{-2}	2.08	Valeric Acid OR Isovaleric Acid OR 2-Methylbutyric acid

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
6679	PLS-DA	9.38	286.2120	2.78×10^{-2}	1.90	
5947	PLS-DA	9.38	285.2085	3.13×10^{-2}	1.94	
1106	PLS-DA	9.38	147.1027	1.41×10^{-2}	2.08	Lysine
192	PLS-DA	9.38	267.1791	2.12×10^{-2}	2.07	
62	PLS-DA	9.38	284.2060	2.61×10^{-2}	2.00	
6875	PLS-DA	9.38	268.1809	1.69×10^{-2}	2.07	
797	PLS-DA	9.39	305.1362	1.32×10^{-2}	2.22	Angustibalin (Terpenoid) OR Arteglasin A (Terpenoid) OR Linifolin A (Terpenoid) OR Ceceline OR Nopaline
791	PLS-DA	9.39	284.3612	3.11×10^{-2}	2.06	
1909	PLS-DA	16.05	1323.6573	2.16×10^{-3}	1.26	
5345	PLS-DA	19.92	181.1218	1.18×10^{-10}	5.51	
478	PLS-DA	23.81	415.2114	8.97×10^{-3}	1.14	Magnoshinin (Lignan) OR Deacetylvindoline (Alkaloid)
6728	PLS-DA	24.77	1471.0864	1.05×10^{-7}	32.48	
5069	PLS-DA	25.88	375.2624	6.40×10^{-3}	2.04	
6466	PLS-DA	26.62	976.5675	1.19×10^{-2}	1.43	
5311	PLS-DA	27.65	589.2458	3.24×10^{-8}	42.05	Magnesium mesoporphyrin
4107	PCA	27.77	621.2710	3.07×10^{-7}	592.62	
5264	PLS-DA	28.63	828.4778	4.36×10^{-9}	65.57	Tubulysin D
4218	Both	29.98	607.2936	2.85 x	38158.92	- Cepharanthine - (Alkaloid)

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
				10^{-8}		
4907	Both	30.06	608.2961	3.49×10^{-8}	38939.68	
1753	PCA	30.08	607.2903	1.61×10^{-6}	214.51	Clerodendrin A (Terpenoid)
4927	PLS-DA	30.16	608.4200	1.32×10^{-7}	237.44	
5078	PLS-DA	30.27	1252.5343	7.12×10^{-8}	57.75	
5267	PLS-DA	30.29	646.2499	2.64×10^{-8}	41.22	
4908	PCA	31.18	792.5609	7.85×10^{-4}	121.25	1,2-dipalmitoyl-sn-glycero-3-thiophosphocholine (Lipid)

Table 22: Ions predominantly observed in the desiccated state in leaves

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
197	PLS-DA	0.61	455.1151	2.44×10^{-9}	19.59	
2864	PLS-DA	0.61	456.1184	5.65×10^{-9}	14.18	
2919	PLS-DA	0.62	457.1155	7.32×10^{-9}	13.11	Flavin mononucleotide
275	PLS-DA	0.74	472.2012	2.71×10^{-3}	5.06	
2149	PLS-DA	0.74	473.2037	2.27×10^{-3}	3.96	5-formiminotetrahydrofolate
884	PLS-DA	0.86	634.2534	1.13×10^{-2}	6.07	
423	PLS-DA	0.87	348.0677	3.60×10^{-8}	10.44	AMP
654	PLS-DA	0.99	113.0267	8.82×10^{-12}	7.81	2-Furoate
1786	PLS-DA	1.17	137.0672	1.00×10^{-8}	8.45	
248	PLS-DA	1.18	165.0521	2.15×10^{-8}	7.84	
1419	PLS-DA	1.18	119.0430	7.43×10^{-10}	6.15	Methylmalonate
174	PLS-DA	1.18	182.0791	2.55×10^{-8}	6.86	Tyrosine
212	PLS-DA	1.19	136.0647	2.49×10^{-9}	14.09	Adenine
1907	PLS-DA	1.30	270.1057	1.36×10^{-9}	20.15	
2441	PLS-DA	1.30	269.1051	8.10×10^{-9}	20.47	
660	PLS-DA	1.30	268.2540	1.02×10^{-8}	41.68	
54	PLS-DA	1.31	268.1036	4.81×10^{-10}	27.40	Adenosine
3148	PLS-DA	1.31	268.3363	8.36×10^{-8}	21.44	
3402	PLS-DA	1.31	268.4106	2.33×10^{-9}	18.94	
3065	PLS-DA	1.31	268.1725	1.66×10^{-11}	15.60	Elaeocarpidine (alkaloid)
1005	PLS-DA	1.60	152.0569	5.23×10^{-12}	6.06	Guanine
1543	PLS-DA	1.60	284.0990	1.15×10^{-11}	5.24	Guanosine
2781	PLS-DA	1.91	201.1204	1.79×10^{-10}	107.09	
317	PLS-DA	1.92	118.0863	1.73×10^{-8}	219.86	Betaine (^{1}H δ=3.25)
535	PLS-DA	1.94	230.1023	2.14×10^{-8}	148.94	
624	PLS-DA	1.94	184.0963	1.33×10^{-8}	144.28	
1403	PLS-DA	1.95	247.5114	5.25×10^{-8}	616.38	
207	PLS-DA	1.95	247.2738	5.00×10^{-8}	1613.85	
2643	PLS-DA	1.95	247.6922	3.34×10^{-8}	614.11	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
2691	PLS-DA	1.96	247.7619	2.92×10^{-8}	630.73	
3243	PLS-DA	1.96	247.7948	1.87×10^{-9}	617.04	
426	PLS-DA	1.96	247.2035	7.27×10^{-10}	253.77	
124	PLS-DA	1.96	248.1310	3.62×10^{-8}	628.36	Fagaramide
12	PLS-DA	1.96	247.1278	3.81×10^{-13}	376.24	2,4-Bis(acetamido)-2,4,6-trideoxy-beta-L-altropyranose
2046	PLS-DA	1.96	249.1334	2.36×10^{-8}	429.89	6-Hydroxymelatonin
1883	PLS-DA	1.97	247.7249	4.38×10^{-8}	559.86	
6858	PLS-DA	1.97	247.3234	3.99×10^{-8}	885.08	
911	PLS-DA	1.97	247.4270	9.74×10^{-8}	616.72	
1778	PLS-DA	1.98	247.6444	1.41×10^{-7}	525.87	
1525	PLS-DA	1.98	247.5761	5.45×10^{-8}	552.72	
2994	PLS-DA	1.98	247.7439	2.46×10^{-8}	770.16	
4193	PLS-DA	1.98	247.8492	2.68×10^{-8}	518.28	
3378	PLS-DA	1.98	247.8748	2.09×10^{-8}	400.62	
3676	PLS-DA	1.98	247.6640	3.65×10^{-8}	600.78	
2761	PLS-DA	1.99	247.8225	3.35×10^{-8}	366.86	
858	PLS-DA	2.00	247.3590	9.77×10^{-8}	663.79	
2800	PLS-DA	2.03	229.1139	5.26×10^{-9}	9.31	Lindenenone
3437	PLS-DA	2.18	262.0735	3.86×10^{-11}	53.20	
972	PLS-DA	2.18	279.2528	2.19×10^{-8}	588.75	
2497	PLS-DA	2.22	279.1708	5.91×10^{-8}	359.26	
69	PLS-DA	2.22	279.0996	2.56×10^{-9}	780.86	Tuliposide A
877	PLS-DA	2.23	150.0593	2.62×10^{-9}	77.81	Methionine
1722	PLS-DA	3.50	182.0804	3.83×10^{-9}	176.31	Tyrosine
116	PLS-DA	3.55	311.1229	7.61×10^{-9}	1600.21	7-Hydroxy-3-(4-methoxyphenyl)-4-propylcoumarin OR Porphyrin
1449	PLS-DA	4.74	687.1548	2.25×10^{-2}	1.43	
479	PLS-DA	4.99	244.1168	6.50×10^{-8}	89.21	Biotin amide
3186	PLS-DA	4.99	262.4536	1.50×10^{-8}	387.24	
502	PLS-DA	4.99	263.1457	4.07×10^{-8}	56.71	Physovenine

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
4110	PLS-DA	5.00	261.7541	4.00×10^{-8}	539.48	
3103	PLS-DA	5.02	261.7975	2.24×10^{-8}	505.05	
2645	PLS-DA	5.03	261.9795	2.50×10^{-8}	295.43	
1	PLS-DA	5.04	261.1435	2.50×10^{-13}	295.89	Lacinilene C 7-methyl ether OR 1-(4-Hydroxyphenyl)-1-decene-3,5-dione OR Dihydroshikonofuran
3091	PLS-DA	5.04	261.7376	7.28×10^{-8}	359.38	
1830	PLS-DA	5.05	261.9493	1.03×10^{-7}	386.18	
4120	PLS-DA	5.05	261.8098	3.75×10^{-8}	368.13	
896	PLS-DA	5.05	261.5362	1.19×10^{-7}	575.17	
2274	PLS-DA	5.05	261.8582	8.98×10^{-8}	566.77	
2449	PLS-DA	5.05	261.2215	2.80×10^{-11}	231.79	
380	PLS-DA	5.05	261.3436	1.07×10^{-7}	667.78	
1216	PLS-DA	5.05	261.6748	8.63×10^{-8}	346.30	
2745	PLS-DA	5.05	243.1308	1.44×10^{-9}	82.55	
4787	PLS-DA	5.06	261.2937	6.58×10^{-8}	1233.07	
82	PLS-DA	5.06	262.1463	6.24×10^{-8}	460.12	Cochlearine (Tropane alkaloid)
3751	PLS-DA	5.06	262.0028	5.45×10^{-8}	300.82	
2327	PLS-DA	5.06	261.9258	3.82×10^{-8}	397.10	
4133	PLS-DA	5.06	261.8842	2.42×10^{-8}	676.58	
597	PLS-DA	5.07	198.1109	5.27×10^{-8}	433.49	
2623	PLS-DA	5.07	261.9138	2.59×10^{-8}	392.11	
1901	PLS-DA	5.07	262.0181	6.18×10^{-8}	255.91	
4009	PLS-DA	5.08	261.9911	1.18×10^{-8}	214.33	
4157	PLS-DA	5.10	262.0470	2.14×10^{-8}	104.67	Phosphotyrosine
1297	PLS-DA	5.12	739.3126	4.19×10^{-8}	364.20	
236	PLS-DA	5.19	132.0985	4.35×10^{-8}	18.44	Leucine OR Isoleucine OR beta-Alaninebetaine OR N,N-Diethylglycine
241	PLS-DA	5.44	132.0983	6.65×10^{-8}	28.65	Same compound as 236 above

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
3339	PLS-DA	5.48	262.5744	4.54×10^{-8}	236.25	
2	PLS-DA	5.48	261.1429	2.47×10^{-11}	281.96	Lacinilene C 7-methyl ether OR 1-(4-Hydroxyphenyl)-1-decene-3,5-dione OR Dihydroshikonofuran
372	PLS-DA	5.49	261.2220	1.25×10^{-9}	169.70	
6902	PLS-DA	5.49	261.7456	8.87×10^{-8}	308.60	
1500	PLS-DA	5.50	779.2709	4.09×10^{-9}	165.48	
4812	PLS-DA	5.50	261.6921	1.40×10^{-7}	289.91	
4858	PLS-DA	5.52	262.0134	5.82×10^{-8}	190.97	
3368	PLS-DA	5.71	461.1409	1.87×10^{-9}	4.71	
523	PLS-DA	5.71	460.1380	6.89×10^{-11}	6.45	
404	PLS-DA	6.42	642.1737	9.23×10^{-8}	125.92	
138	PLS-DA	6.43	641.1702	6.02×10^{-8}	145.95	
3778	PLS-DA	6.48	296.7733	1.50×10^{-2}	53.46	
1667	PLS-DA	6.52	166.2050	1.12×10^{-7}	128.50	
2505	PLS-DA	6.54	296.3799	2.91×10^{-8}	90.44	
2707	PLS-DA	6.56	297.2909	1.03×10^{-8}	49.53	
44	PLS-DA	6.56	296.1327	1.11×10^{-9}	60.16	N6,N6-Dimethyladenosine
279	PLS-DA	6.56	297.1347	2.93×10^{-9}	40.50	2-Methylaminoadenosine
3891	PLS-DA	6.56	296.1799	7.15×10^{-9}	27.80	Serratanidine
2293	PLS-DA	6.56	296.6902	1.95×10^{-8}	76.20	
2148	PLS-DA	6.57	296.4958	2.70×10^{-8}	115.03	
1720	PLS-DA	6.57	296.6605	2.16×10^{-8}	87.93	
1315	PLS-DA	6.57	296.5327	6.69×10^{-9}	88.38	
903	PLS-DA	6.57	296.3617	3.55×10^{-9}	79.22	
3673	PLS-DA	6.57	296.6001	1.51×10^{-9}	73.63	
3702	PLS-DA	6.57	296.6761	2.10×10^{-9}	84.83	
1223	PLS-DA	6.57	296.4562	2.08×10^{-8}	84.18	
3155	PLS-DA	6.57	296.5849	4.81×10^{-10}	75.26	
3974	PLS-DA	6.58	296.6171	6.49×10^{-9}	92.66	
2450	PLS-DA	6.58	296.2908	6.46×10^{-9}	103.56	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
2132	PLS-DA	6.58	296.4195	8.86×10^{-9}	89.19	
2391	PLS-DA	6.59	186.0890	8.78×10^{-10}	123.19	
2697	PLS-DA	6.59	296.7489	1.30×10^{-10}	103.46	
3269	PLS-DA	6.60	296.8181	5.98×10^{-9}	168.09	
3	PLS-DA	6.62	295.1283	1.15×10^{-13}	223.59	
4124	PLS-DA	6.63	296.0288	4.84×10^{-9}	195.18	
123	PLS-DA	6.64	120.0801	2.30×10^{-9}	249.51	
6993	PLS-DA	6.64	296.0012	1.81×10^{-9}	285.57	
2858	PLS-DA	6.64	295.7565	2.97×10^{-9}	288.50	
110	PLS-DA	6.64	166.0857	2.64×10^{-9}	207.10	Phenylalanine
886	PLS-DA	6.64	295.7436	4.42×10^{-9}	414.84	
1230	PLS-DA	6.64	295.9556	5.68×10^{-9}	329.57	
729	PLS-DA	6.64	103.0537	1.77×10^{-9}	58.62	N-Formiminoglycine OR Cycloserine
2446	PLS-DA	6.64	295.4499	4.74×10^{-9}	456.05	
2888	PLS-DA	6.65	295.9714	2.56×10^{-9}	208.59	
259	PLS-DA	6.65	295.3399	2.05×10^{-9}	506.58	
2507	PLS-DA	6.65	295.7775	1.06×10^{-9}	301.19	
825	PLS-DA	6.65	295.6892	7.77×10^{-10}	217.50	
1643	PLS-DA	6.65	131.0486	6.17×10^{-8}	8.99	Ureidoacrylate
591	PLS-DA	6.65	295.5439	6.02×10^{-9}	311.25	
2531	PLS-DA	6.65	295.9087	4.28×10^{-9}	217.21	
105	PLS-DA	6.65	295.2861	4.19×10^{-9}	842.93	
3095	PLS-DA	6.65	295.9412	3.15×10^{-9}	254.48	
2517	PLS-DA	6.65	295.8520	1.25×10^{-9}	293.29	
2338	PLS-DA	6.65	249.1229	2.35×10^{-9}	39.00	
2114	PLS-DA	6.65	295.8818	2.28×10^{-9}	249.56	
600	PLS-DA	6.65	232.0963	2.09×10^{-9}	203.93	
1035	PLS-DA	6.66	295.8179	8.54×10^{-9}	319.64	
703	PLS-DA	6.66	295.6150	3.57×10^{-9}	378.56	
3944	PLS-DA	6.66	296.0152	2.99×10^{-9}	221.18	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
366	PLS-DA	6.66	295.2102	2.98×10^{-11}	134.76	
3926	PLS-DA	6.67	121.0806	8.40×10^{-11}	18.36	Benzamidine
2909	PLS-DA	6.67	296.0449	1.02×10^{-8}	270.68	
1789	PLS-DA	6.68	107.0487	1.68×10^{-11}	6.32	Benzaldehyde
2508	PLS-DA	6.68	812.2800	1.34×10^{-7}	75.82	
1822	PLS-DA	6.68	120.1829	4.73×10^{-8}	148.99	
474	PLS-DA	6.68	278.1015	3.47×10^{-9}	60.66	
2853	PLS-DA	6.68	167.0840	1.47×10^{-8}	10.68	
1698	PLS-DA	6.69	1004.3215	9.25×10^{-9}	209.94	trans-Hexadec-2-enoyl-CoA
991	PLS-DA	6.70	1003.3174	5.61×10^{-9}	196.13	
421	PLS-DA	7.15	655.1867	1.21×10^{-7}	159.03	2-(S-Glutathionyl)acetyl glutathione
3281	PLS-DA	8.01	126.0906	2.74×10^{-10}	5.05	
3686	PLS-DA	9.18	432.2778	4.60×10^{-2}	1.14	
2207	PLS-DA	10.19	463.1547	4.69×10^{-3}	2.62	
3143	PLS-DA	10.46	484.2401	1.68×10^{-2}	7.37	
2963	PLS-DA	10.91	225.0738	4.57×10^{-9}	7.10	Sinapic acid OR Benzylmalic acid
2313	PLS-DA	11.21	227.3035	5.80×10^{-9}	41.31	
132	PLS-DA	11.22	227.1655	1.19×10^{-8}	29.79	Cyclo(L-leucyl-L-leucyl)
748	PLS-DA	11.29	209.1521	4.77×10^{-9}	8.88	
2712	PLS-DA	11.36	569.1998	1.09×10^{-2}	1.52	
3522	PLS-DA	13.87	541.2095	2.91×10^{-2}	1.16	
1361	PLS-DA	14.97	211.1679	5.82×10^{-9}	26.03	
2989	PLS-DA	15.00	193.1571	3.68×10^{-12}	8.14	α- or β-Ionone (carotenoid)
3980	PLS-DA	15.32	486.2077	3.95×10^{-9}	22.91	
3223	PLS-DA	16.63	497.2098	5.53×10^{-8}	23.08	
2675	PLS-DA	16.67	954.4108	5.38×10^{-3}	6.09	
2012	PLS-DA	16.69	355.1523	1.24×10^{-8}	9.16	Cristacarpin (Flavonoid) OR Xanthohumol (Flavonoid)
1307	PLS-DA	16.70	373.1633	4.44×10^{-9}	14.06	Arctigenin (Lignan) OR Fargesone A (Lignan)

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
						OR Kadsurin A (Lignan)
4877	PLS-DA	16.71	356.1553	6.80×10^{-9}	10.89	Buntanine (Alkaloid) OR S-adenosylmethioninamine
1604	PLS-DA	16.71	151.0749	1.50×10^{-8}	7.55	4-Coumaryl alcohol OR Phenylpropanoate OR 4-Hydroxydihydrocinnamaldehyde
3025	PLS-DA	16.92	994.4251	4.03×10^{-8}	37.53	
144	PLS-DA	16.92	993.4230	5.26×10^{-8}	41.04	
3050	PLS-DA	16.92	995.4280	3.22×10^{-8}	28.97	
3695	PLS-DA	16.92	996.4322	9.93×10^{-9}	18.64	
1312	PLS-DA	16.93	497.7134	2.28×10^{-8}	25.71	
828	PLS-DA	16.93	497.2121	1.03×10^{-9}	24.79	
4802	PLS-DA	17.45	977.4168	3.55×10^{-8}	10.21	
833	PLS-DA	17.46	488.2066	1.04×10^{-13}	8.15	
3150	PLS-DA	17.46	488.7085	5.12×10^{-12}	9.41	
1723	PLS-DA	19.91	333.1317	1.39×10^{-2}	22.02	
7482	PLS-DA	22.96	302.2718	1.23×10^{-2}	39.66	
3919	PLS-DA	23.04	286.3001	9.91×10^{-3}	1.10	
121	PLS-DA	23.32	476.2765	1.39×10^{-9}	18.02	Tubulosine (Alkaloid) OR Alangimarckine (Alkaloid)
1986	PLS-DA	23.33	477.2789	1.38×10^{-9}	13.87	Emetamine (Alkaloid)
555	PLS-DA	23.43	577.2763	5.46×10^{-8}	40.90	Tiliacorine (Alkaloid)
2736	PLS-DA	23.52	520.2641	8.74×10^{-12}	21.41	
527	PLS-DA	23.57	540.3053	4.07×10^{-9}	16.51	
1712	PLS-DA	24.07	522.2823	3.67×10^{-9}	11.17	
1271	PLS-DA	24.07	579.2919	9.00×10^{-9}	17.54	
1668	PLS-DA	24.19	338.3399	1.23×10^{-2}	3.61	
300	PLS-DA	24.47	480.3075	1.29×10^{-10}	12.34	Delcorine (Terpenoid alkaloid)
3821	PLS-DA	25.35	683.4279	5.55×10^{-3}	1.75	
251	PLS-DA	25.42	482.3240	1.73×10^{-9}	10.30	
4033	PLS-DA	26.54	387.1350	4.37×10^{-2}	9.40	1-O-Sinapoyl beta-D-glucoside

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
						OR 4-O-beta-D-Glucosyl-sinapate OR Isosamidin (Coumarin) OR Peucenidin (Coumarin) OR Pteryxin (Coumarin) OR Samidin (Coumarin)
769	PLS-DA	26.73	674.4616	1.42×10^{-2}	1.30	
5196	PLS-DA	27.06	546.3580	4.84×10^{-3}	3.99	
2276	PLS-DA	27.16	965.6084	4.19×10^{-2}	3.38	
632	PLS-DA	27.71	1278.7194	1.01×10^{-7}	8.43	
3211	PLS-DA	28.08	731.4883	1.54×10^{-8}	11.31	(3R,2'S)-Myxol 2'-alpha-L-fucoside (Carotenoid biosynthetic intermediate)
3212	PLS-DA	28.65	1117.0772	4.97×10^{-9}	7.79	
3401	PLS-DA	29.25	1280.7371	5.72×10^{-9}	6.36	
5934	Both	29.57	778.5385	3.66×10^{-2}	5.16	
5923	PLS-DA	29.61	778.7868	3.15×10^{-2}	6.28	
677	Both	29.73	778.5359	4.19×10^{-2}	4.51	
2250	PLS-DA	29.74	817.5013	3.32×10^{-2}	5.43	
2401	PLS-DA	29.78	780.8844	3.09×10^{-2}	5.72	
4906	PLS-DA	29.79	900.4469	3.53×10^{-2}	8.85	
4916	Both	29.80	778.5369	3.81×10^{-2}	4.91	
4843	PLS-DA	29.80	779.1307	3.32×10^{-2}	5.46	
2078	PLS-DA	29.82	816.4989	2.61×10^{-2}	8.19	
6984	PLS-DA	29.83	778.8792	3.35×10^{-2}	5.43	
5063	PLS-DA	29.90	800.5219	3.36×10^{-2}	7.88	
6824	PLS-DA	29.90	802.5359	2.59×10^{-2}	5.79	
5868	PLS-DA	29.95	801.5261	2.94×10^{-2}	9.65	
2762	PLS-DA	30.08	571.4409	2.53×10^{-3}	2.12	15,15'-Dihydroxy-beta-carotene
3446	PLS-DA	30.08	568.4264	1.72×10^{-2}	2.43	
1189	PLS-DA	30.08	568.7158	7.89×10^{-3}	2.47	
602	PLS-DA	30.08	568.6444	4.45×10^{-3}	2.96	
3377	PLS-DA	30.24	709.5032	1.54×10^{-8}	14.78	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
3326	PLS-DA	30.32	1557.0709	1.35×10^{-2}	10.40	
3114	PLS-DA	30.37	817.5007	3.99×10^{-2}	4.08	
3342	PLS-DA	30.39	1556.0682	1.78×10^{-2}	9.51	
6946	PLS-DA	30.43	778.7849	4.38×10^{-2}	5.29	
6978	PLS-DA	30.45	779.7906	4.54×10^{-2}	4.95	
977	PLS-DA	30.45	779.5417	4.48×10^{-2}	5.10	
499	Both	30.47	778.5377	4.28×10^{-2}	5.43	
1622	PLS-DA	30.49	1139.6131	1.28×10^{-8}	12.78	
644	PLS-DA	30.52	1119.6858	5.12×10^{-9}	11.93	
434	PLS-DA	30.54	1118.6831	1.07×10^{-8}	12.16	
3054	PLS-DA	30.71	778.7870	4.21×10^{-2}	7.50	
3088	PLS-DA	30.76	779.7935	4.22×10^{-2}	7.51	
3110	PLS-DA	30.76	778.8811	4.41×10^{-2}	6.27	
3252	PLS-DA	30.78	779.1302	4.45×10^{-2}	5.40	
1083	PLS-DA	31.07	1124.6424	2.18×10^{-2}	26.79	
1343	PLS-DA	31.08	960.8464	1.21×10^{-4}	11.78	
838	PLS-DA	31.13	1118.6828	6.66×10^{-11}	14.36	
731	PLS-DA	31.17	1123.6379	4.70×10^{-12}	27.40	
4901	PLS-DA	31.28	1229.8167	2.29×10^{-2}	55.65	
2782	PLS-DA	31.63	500.3127	3.26×10^{-2}	2.50	(Unknown sterol)

LC-MS Peaks in Roots

Table 23: Ions predominantly observed in the hydrated state in roots

Peak ID	Found by	RT / min	p-value	m/z	Ratio	Identity
1132	PLS-DA	2.06	7.33×10^{-6}	166.0859	4.19	Phenylalanine
802	PLS-DA	2.06	1.62×10^{-5}	120.0803	4.21	
1417	PLS-DA	5.56	2.46×10^{-8}	298.095	4.16	
1320	PCA	5.71	2.96×10^{-2}	339.0534	2.36	
45	PCA	14.96	1.99×10^{-3}	1179.5800	1.69	
876	PCA	19.75	2.73×10^{-2}	917.4943	32.45	
856	PCA	19.91	1.24×10^{-2}	938.4739	10.15	
863	PCA	19.91	2.92×10^{-2}	916.4922	33.49	
184	PCA	19.95	1.78×10^{-2}	506.3255	4.51	
653	PCA	20.20	3.95×10^{-2}	506.3253	2.17	
1335	PLS-DA	20.47	1.76×10^{-6}	602.3092	115.80	

Table 24: Ions predominantly observed in the desiccated state in roots

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
573	PLS-DA	1.14	136.0663	1.07×10^{-5}	1.99	
1096	PLS-DA	1.24	269.1060	1.13×10^{-6}	4.13	
145	PLS-DA	1.24	268.1033	3.02×10^{-5}	4.91	Adenosine
114	PLS-DA	1.88	247.1284	4.58×10^{-7}	3.95	2,4-Bis(acetamido)-2,4,6-trideoxy-beta-L-altropyranose OR 2,4-Diacetamido-2,4,6-trideoxy-D-mannopyranose
1033	PLS-DA	1.99	280.1033	1.10×10^{-7}	11.08	
94	PLS-DA	2.08	279.0998	2.83×10^{-6}	8.83	
66	PLS-DA	4.82	261.1441	6.44×10^{-8}	5.65	
1078	PLS-DA	5.23	262.1460	2.28×10^{-7}	2.75	
550	PLS-DA	5.27	262.1460	1.31×10^{-6}	2.71	
65	PLS-DA	5.27	261.1439	9.56×10^{-7}	3.29	
353	PLS-DA	6.44	296.1312	9.41×10^{-8}	5.96	
44	PLS-DA	6.44	295.1280	6.39×10^{-8}	8.09	
389	PLS-DA	7.54	866.1129	6.64×10^{-6}	9.56	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
438	PLS-DA	7.66	865.8613	4.91×10^{-6}	9.68	
460	PLS-DA	7.68	866.3611	5.20×10^{-5}	9.91	
747	PCA	7.72	355.1026	5.03×10^{-3}	2.38	Caffeoyl quinic acid (Phenylpropanoid) OR Scopolin (Phenylpropanoid)
824	PLS-DA	7.72	866.6124	3.52×10^{-6}	12.15	
43	PLS-DA	7.94	340.2589	4.04×10^{-3}	1.13	
641	PLS-DA	8.43	878.1143	1.63×10^{-5}	17.77	
268	PLS-DA	8.56	878.6160	3.67×10^{-5}	13.98	
542	PLS-DA	8.61	878.8670	2.63×10^{-5}	16.88	
414	PLS-DA	8.68	879.1170	2.53×10^{-6}	13.21	
285	PLS-DA	8.88	878.3656	1.17×10^{-5}	12.19	
26	PLS-DA	9.86	453.3430	1.09×10^{-3}	1.16	
1035	PLS-DA	10.06	421.1161	1.27×10^{-6}	3.60	
21	PLS-DA	12.00	1142.5739	1.41×10^{-5}	3.60	
167	PLS-DA	12.02	582.7784	2.01×10^{-6}	2.41	
244	PLS-DA	12.04	581.3024	2.86×10^{-6}	2.86	
919	PLS-DA	12.04	580.2990	3.43×10^{-6}	2.83	
1431	PLS-DA	12.04	1144.5838	2.21×10^{-5}	3.12	
1321	PCA	12.70	981.5235	2.68×10^{-2}	1.49	
1164	PCA	19.92	982.531	2.67×10^{-2}	2.06	
618	PLS-DA	13.23	594.3706	1.72×10^{-2}	2.00	
1156	PCA	14.72	656.4132	2.27×10^{-3}	2.01	
10	PCA	14.91	656.4153	1.52×10^{-3}	1.73	
1172	PCA	14.93	656.4132	2.25×10^{-3}	1.89	
1129	PLS-DA	15.95	805.4910	1.43×10^{-7}	1.87	
1231	PLS-DA	15.95	807.4965	4.21×10^{-8}	1.77	
161	PLS-DA	15.95	804.7422	4.70×10^{-7}	2.08	
953	PLS-DA	15.95	806.4936	2.06×10^{-7}	1.83	
41	PLS-DA	15.95	480.3830	4.82×10^{-7}	1.89	
417	PLS-DA	15.95	804.9268	1.48×10^{-6}	1.71	
827	PLS-DA	15.95	804.8984	4.93×10^{-6}	1.74	

Peak ID	Found by	RT / min	m/z	p-value	Ratio	Identity
4	Both	15.95	804.4893	2.94×10^{-7}	1.91	
124	PLS-DA	15.95	642.4362	4.56×10^{-7}	1.89	
413	PLS-DA	15.95	805.8427	6.82×10^{-7}	1.68	
452	PLS-DA	15.95	805.0146	1.64×10^{-6}	1.72	
880	PLS-DA	15.95	805.7468	2.21×10^{-6}	1.85	
1387	PLS-DA	15.95	805.0899	9.88×10^{-7}	1.74	
1233	PLS-DA	16.11	445.2940	4.29×10^{-6}	5.48	
1225	PLS-DA	16.94	892.4945	3.30×10^{-6}	1.91	
14	PLS-DA	16.94	890.4897	3.43×10^{-6}	1.90	
810	PLS-DA	16.95	890.7511	5.69×10^{-6}	1.83	
798	PLS-DA	18.39	528.3650	7.43×10^{-7}	1.93	
923	PCA	19.08	492.3449	1.01×10^{-2}	7.66	
665	PLS-DA	19.10	504.3094	1.29×10^{-2}	1.74	
941	PCA	19.24	493.3484	1.16×10^{-2}	3.65	
130	PLS-DA	19.29	158.1536	5.82×10^{-6}	1.56	
933	PCA	19.62	493.3485	1.63×10^{-2}	4.31	
306	PCA	19.64	492.3443	1.20×10^{-2}	3.82	
643	PLS-DA	19.83	524.3368	1.31×10^{-2}	2.10	
52	PCA	19.92	492.3463	5.26×10^{-3}	2.49	
476	PCA	20.75	494.3615	1.51×10^{-2}	2.63	
959	PLS-DA	20.77	215.0919	3.32×10^{-3}	1.08	
48	PLS-DA	20.77	158.0268	8.38×10^{-4}	1.09	
158	PLS-DA	20.78	141.0002	1.70×10^{-3}	1.08	
36	PCA	20.89	492.3466	2.18×10^{-2}	2.62	

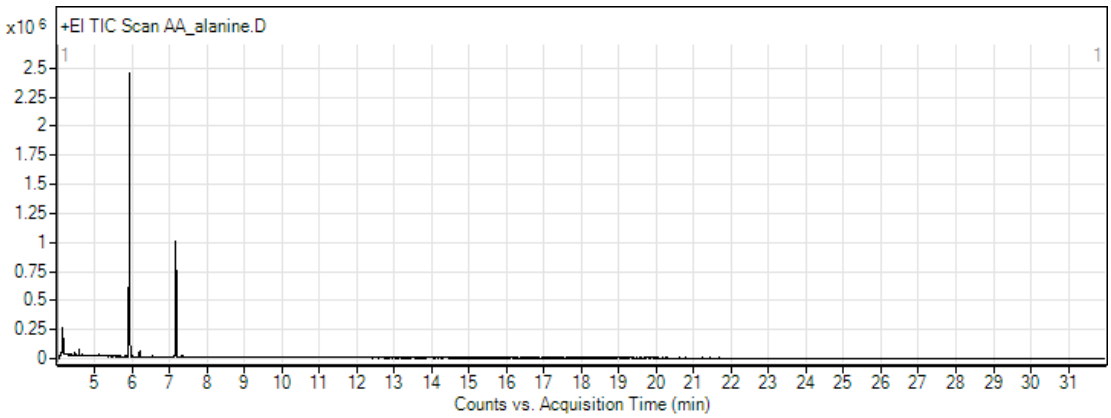
Appendix C: GC-MS Chromatograms

The following Total Ion Chromatograms were generated in Agilent MassHunter Version B.04.00

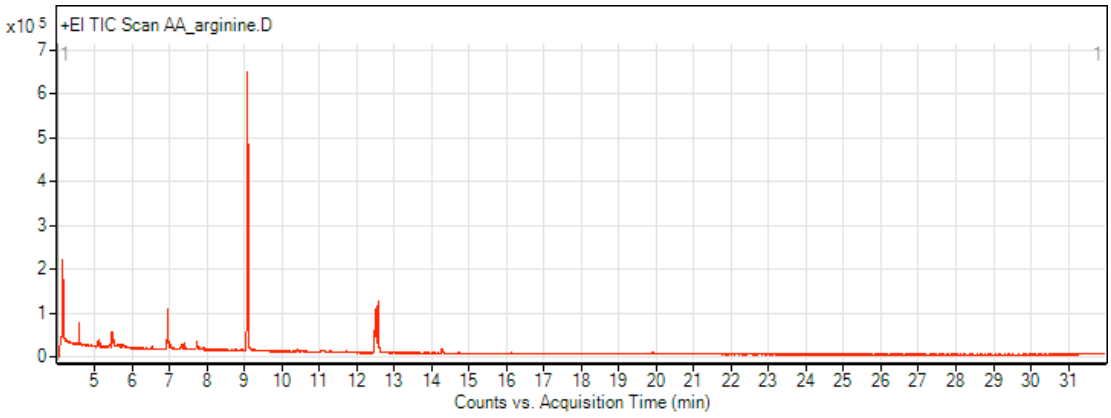
Total Ion Chromatograms Of Standard Compounds

Amino Acids

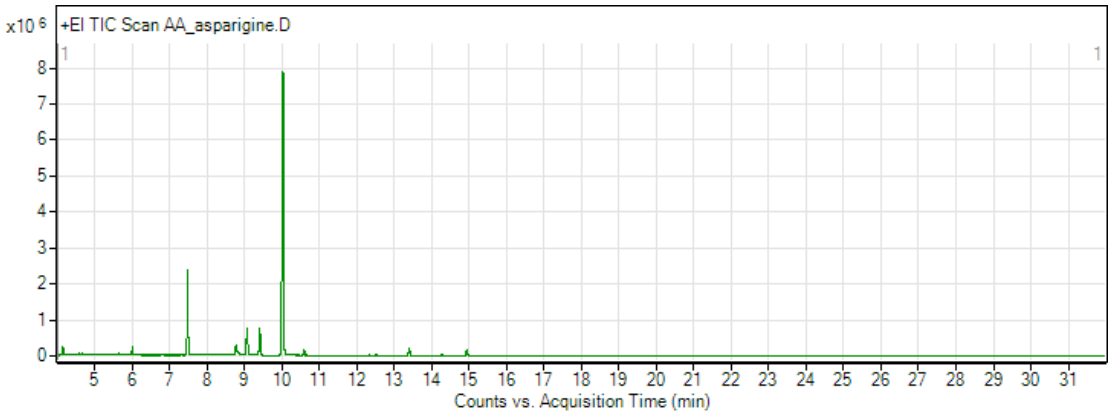
Alanine



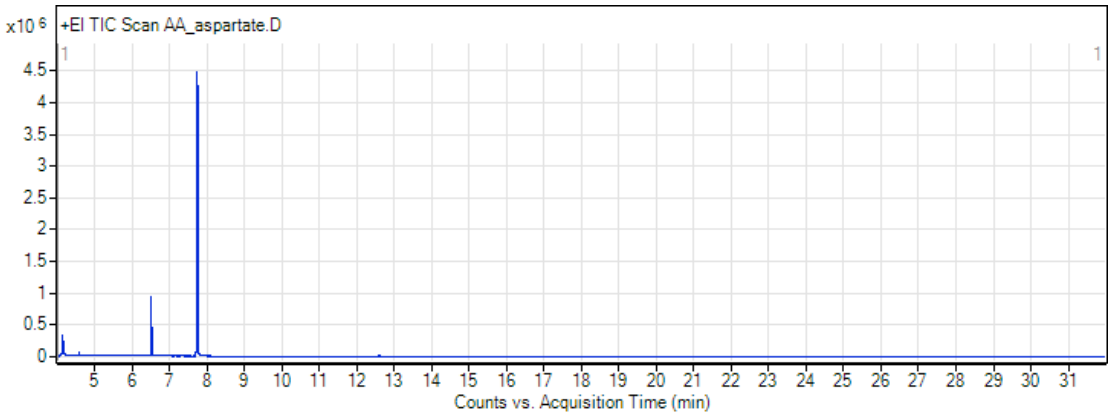
Arginine



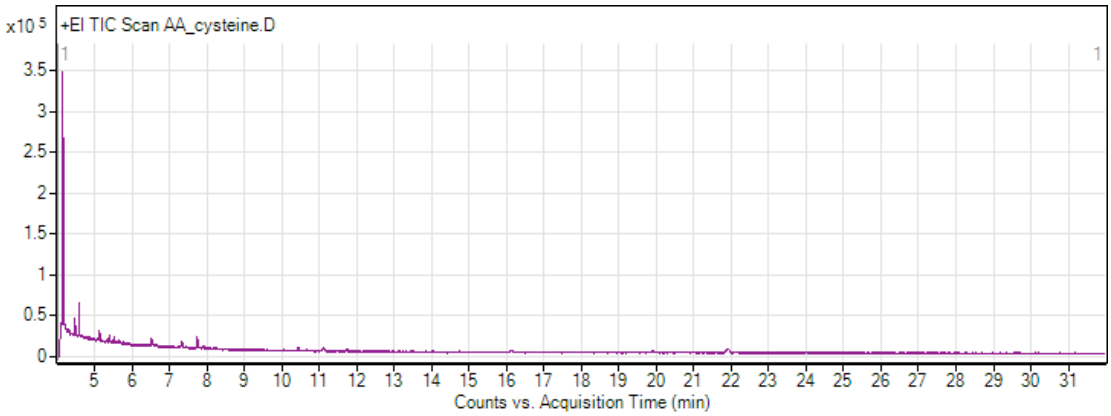
Asparigine



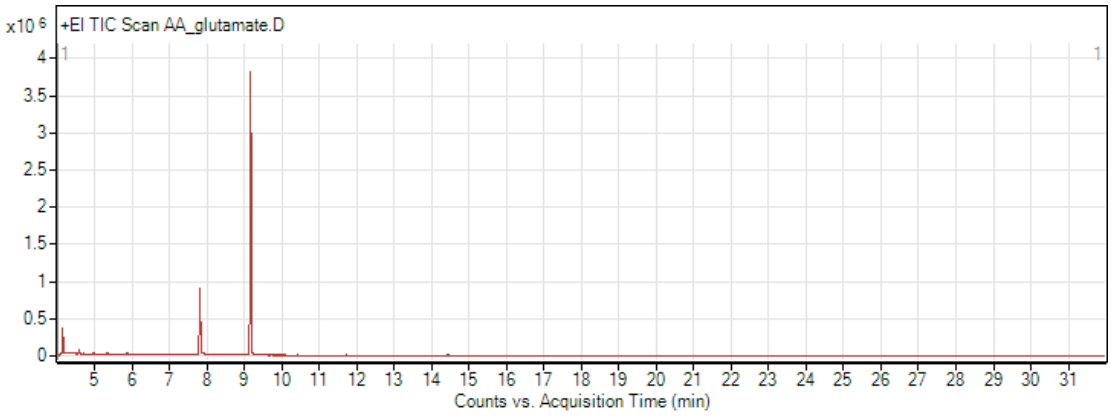
Aspartic Acid



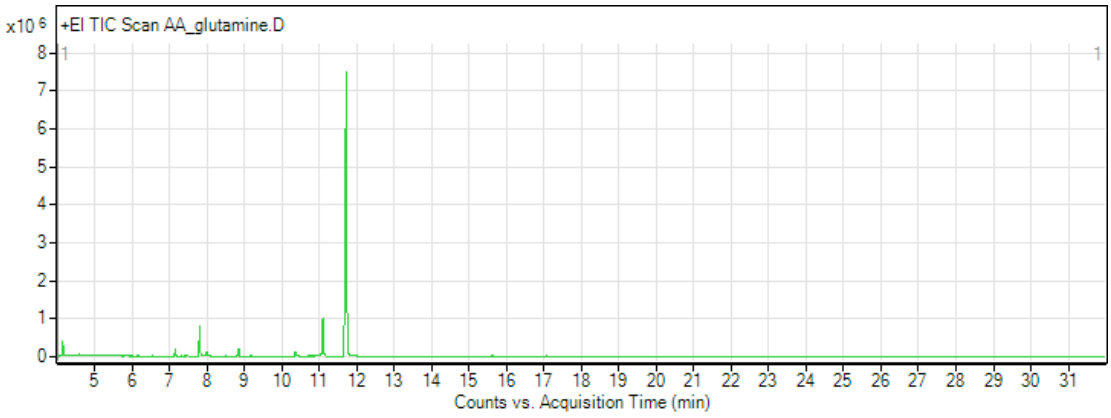
Cysteine



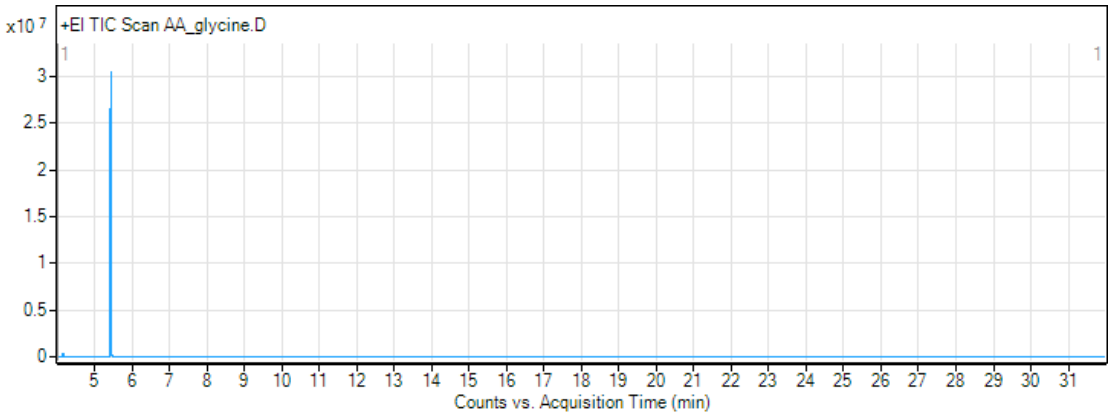
Glutamic Acid



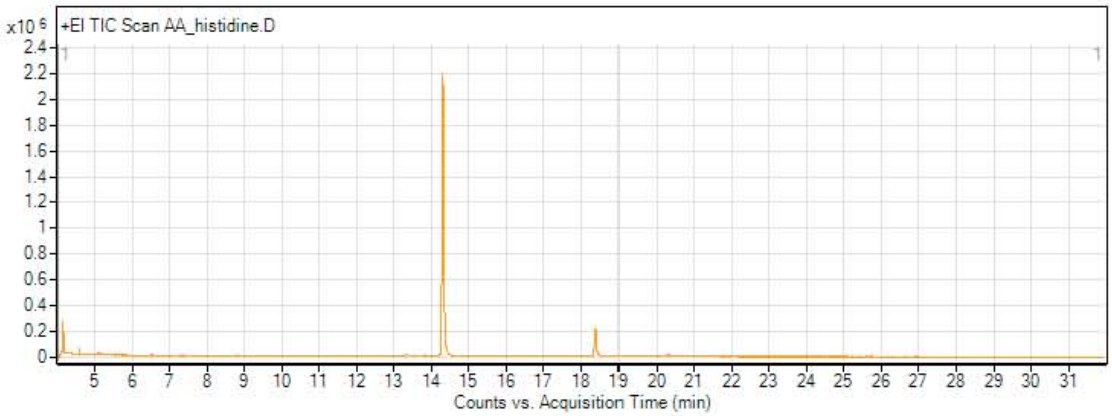
Glutamine



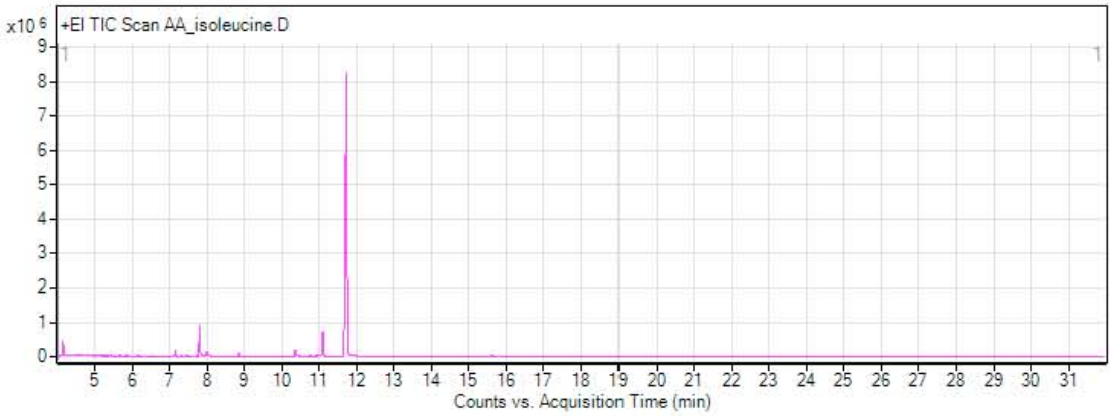
Glycine



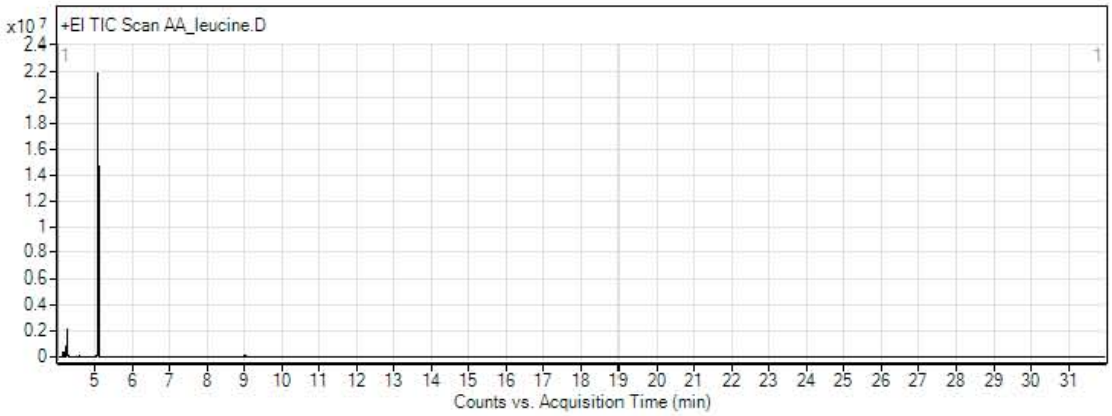
Histidine



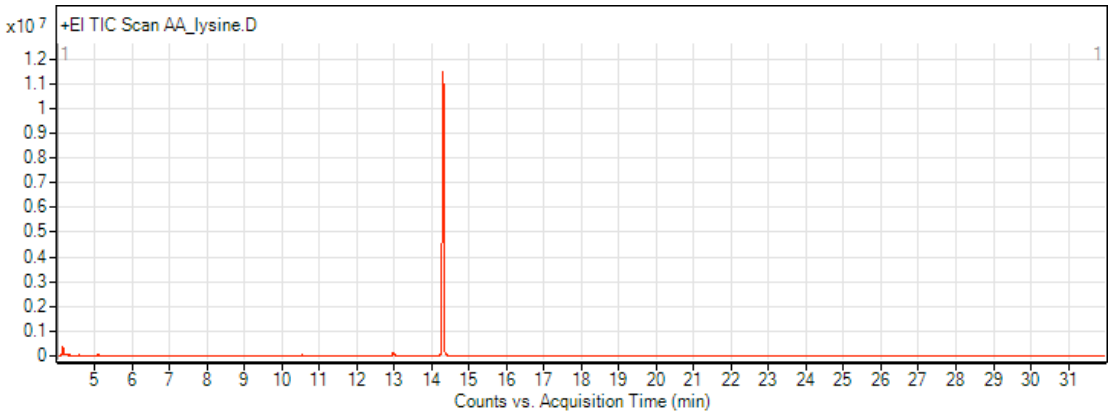
Isoleucine



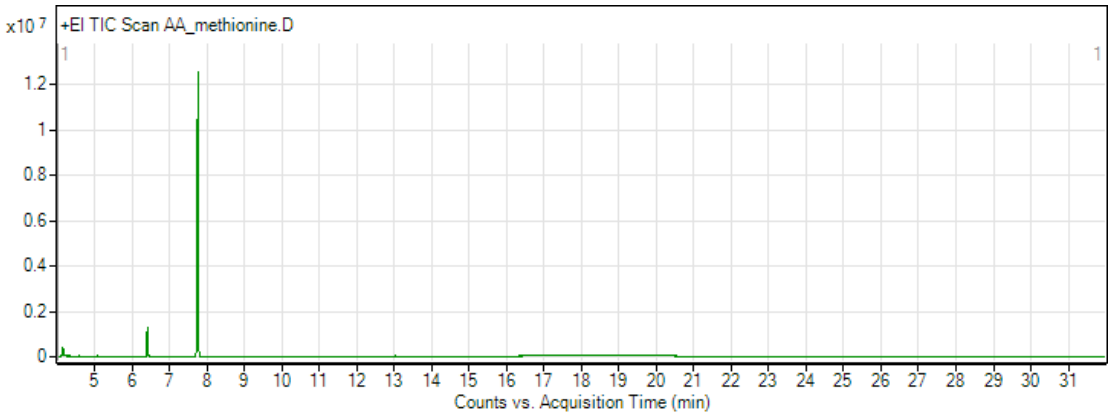
Leucine



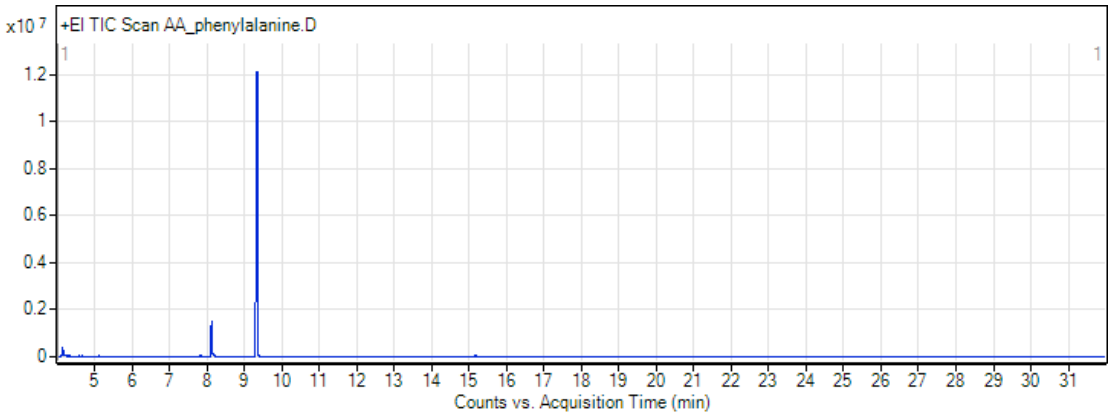
Lysine



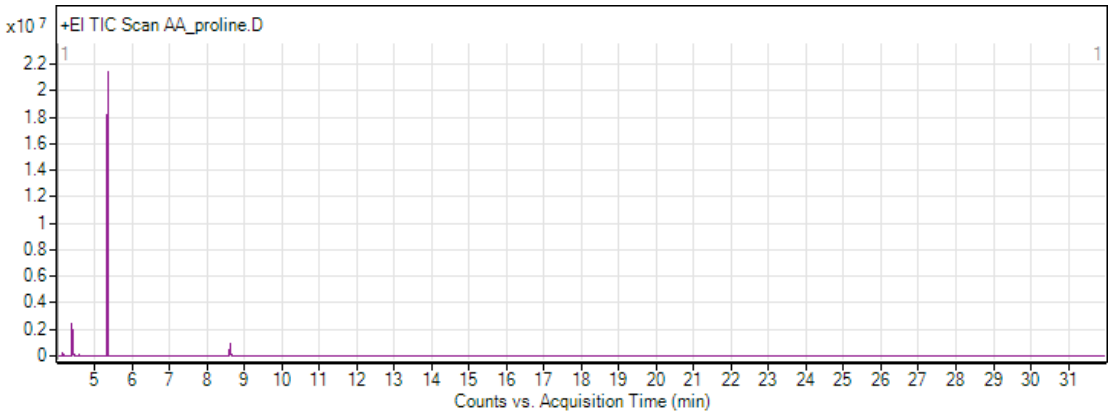
Methionine



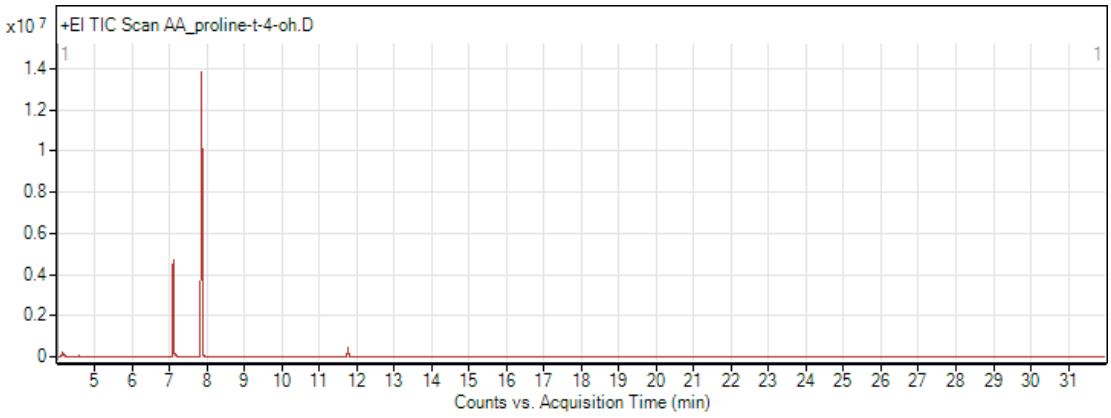
Phenylalanine



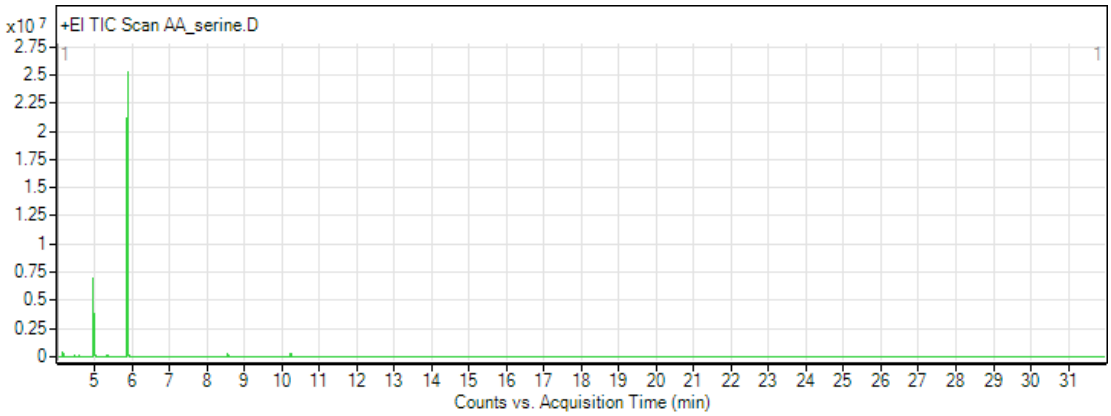
Proline



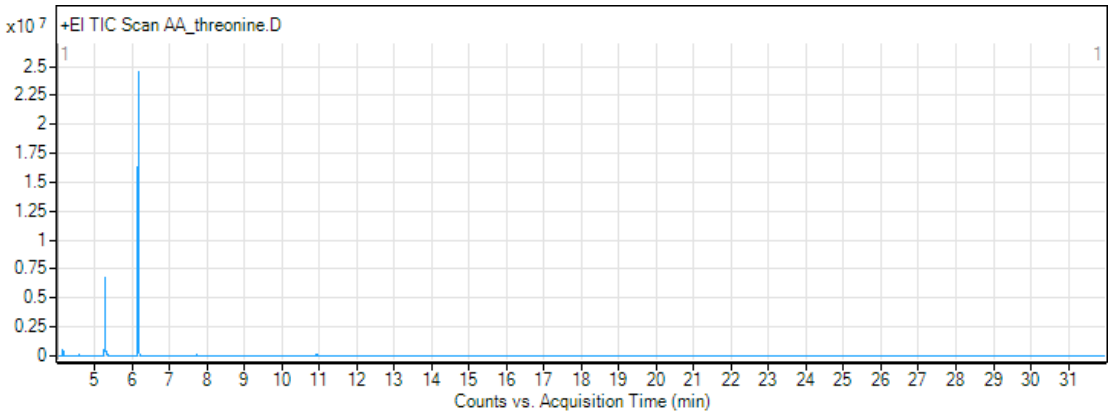
tert-4-hydroxyproline



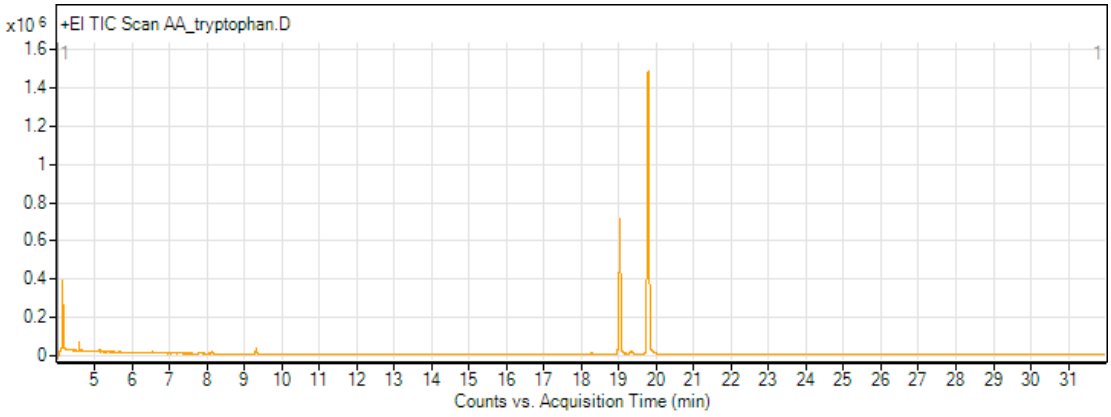
Serine



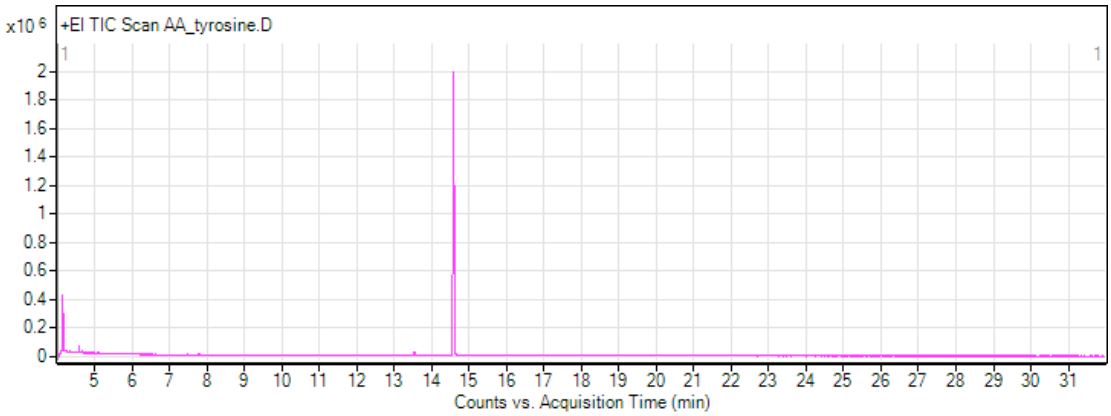
Threonine



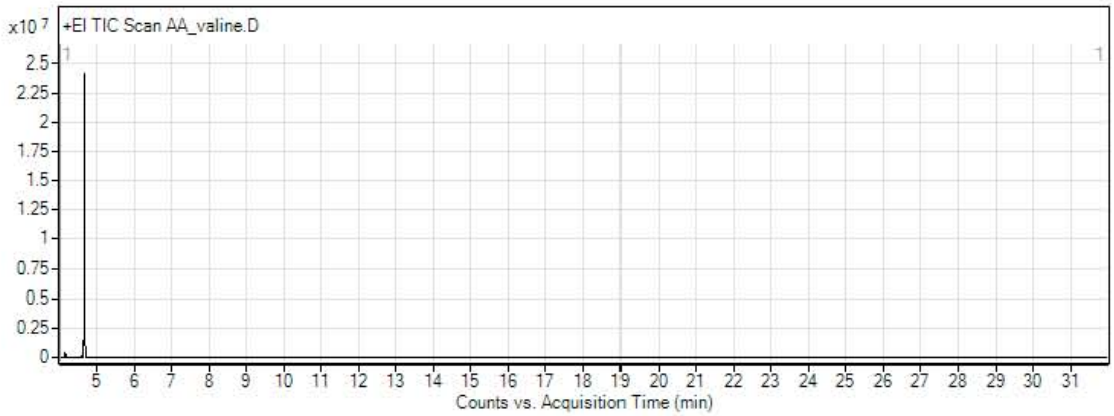
Tryptophan



Tyrosine

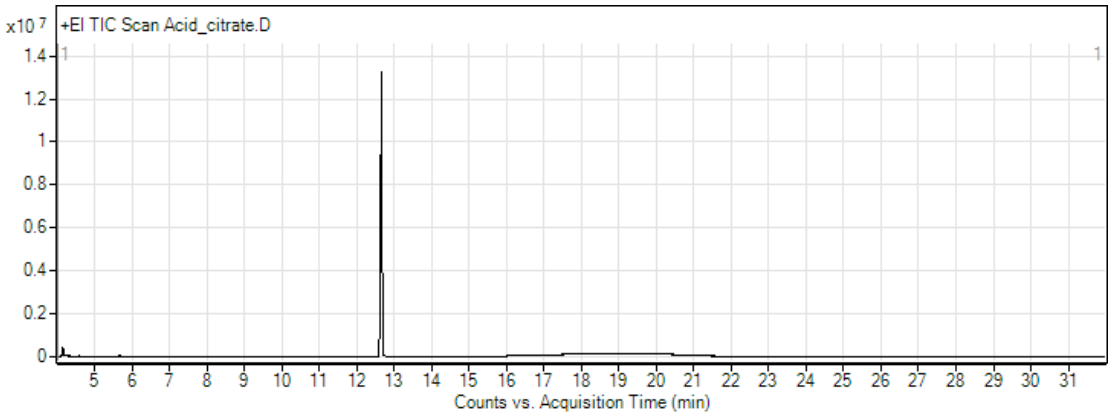


Valine

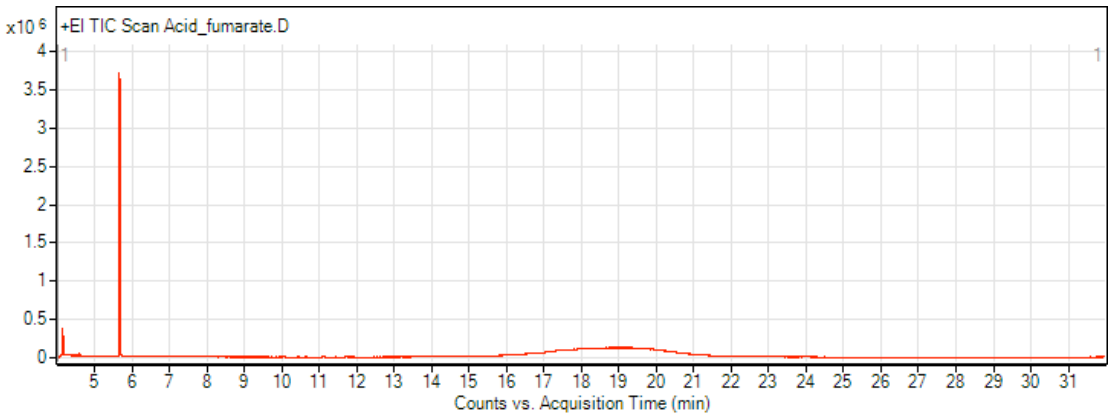

x10 7
+EI TIC Scan AA_valine.D
2.5
2.25
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Counts vs. Acquisition Time (min)

Organic Acids

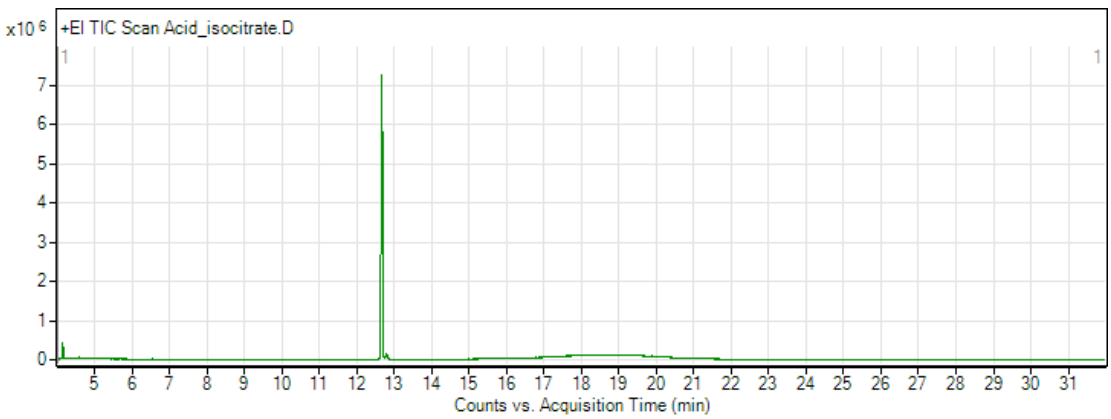
Citric Acid



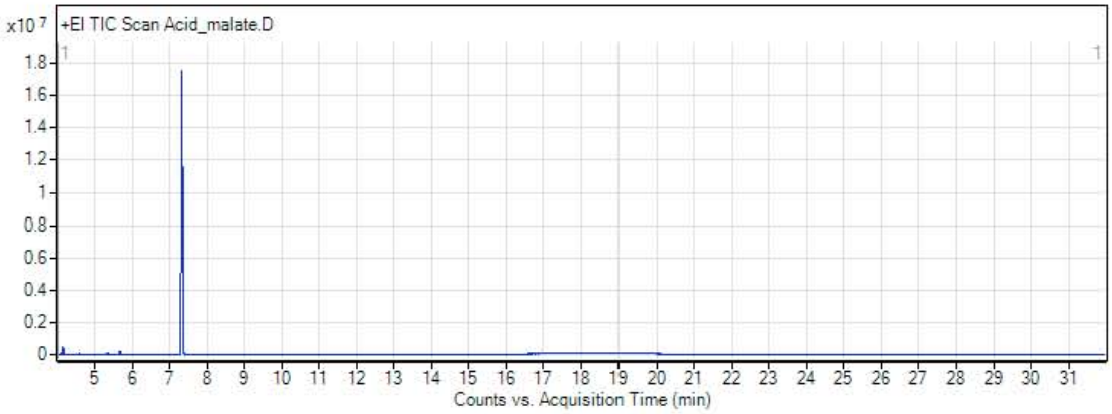
Fumaric Acid



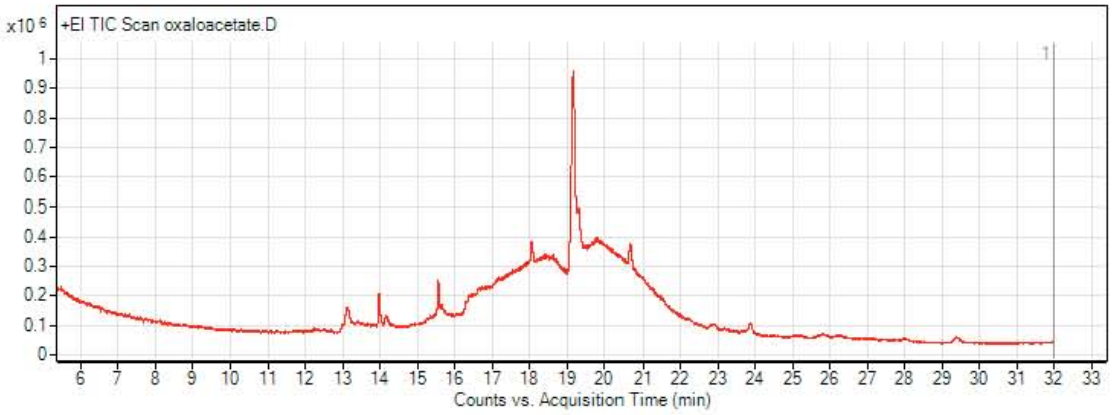
Isocitric Acid



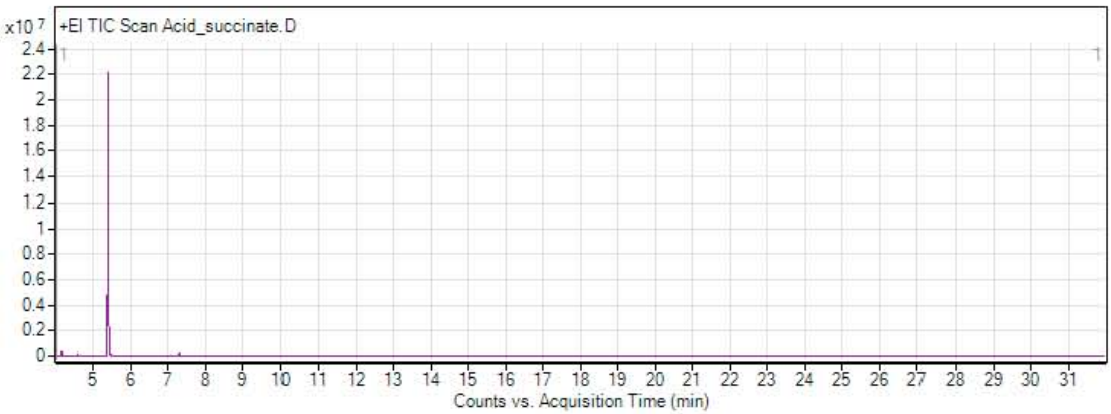
Malic Acid



Oxaloacetic Acid

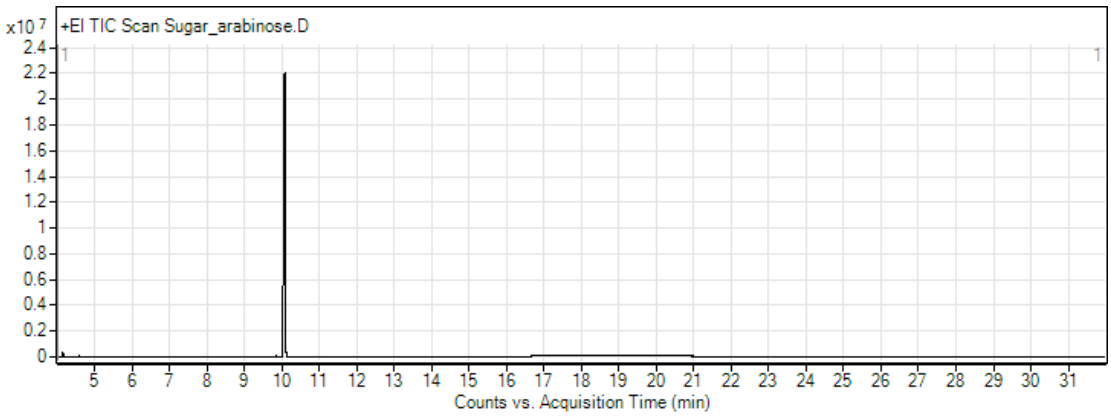


Succinic Acid

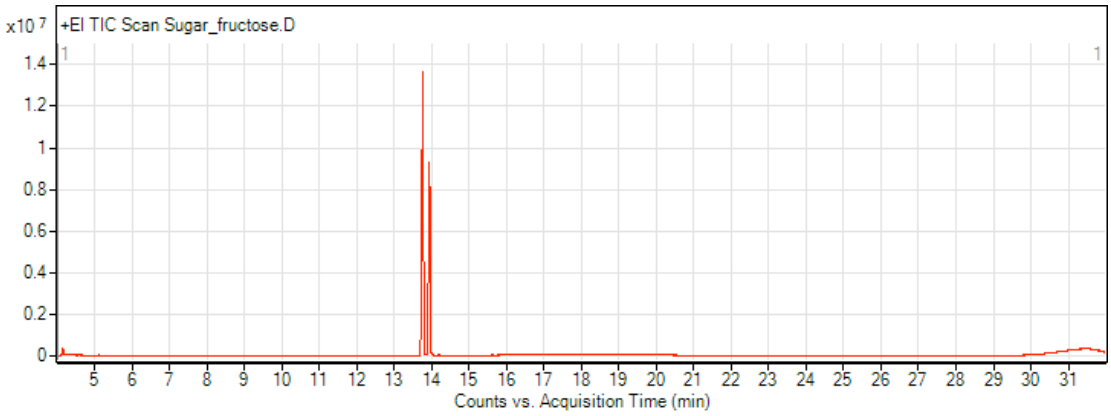


Sugars

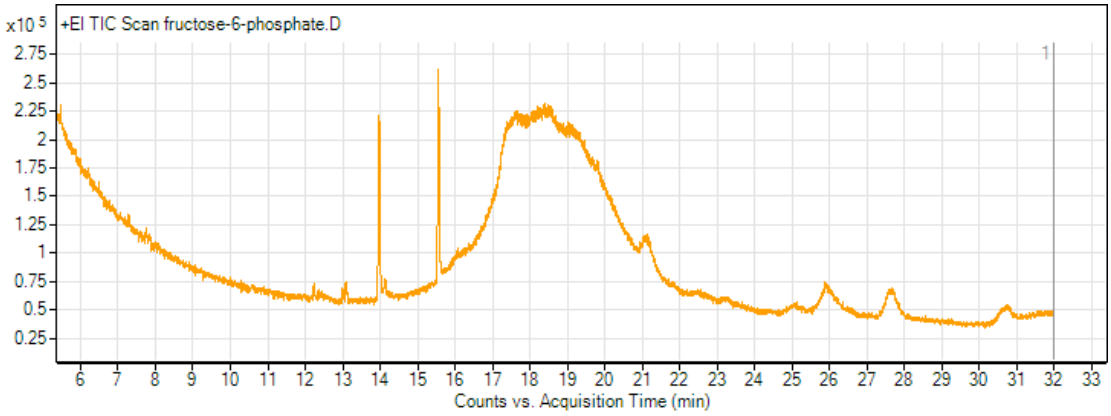
Arabinose



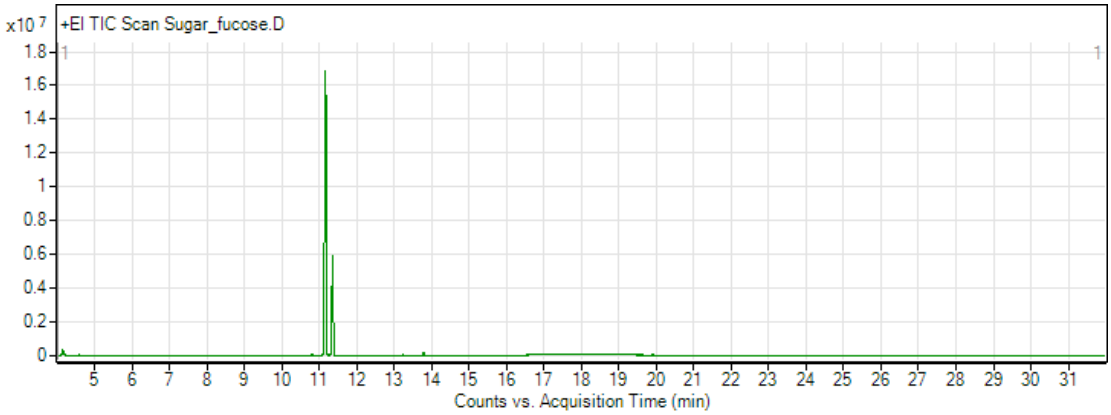
Fructose



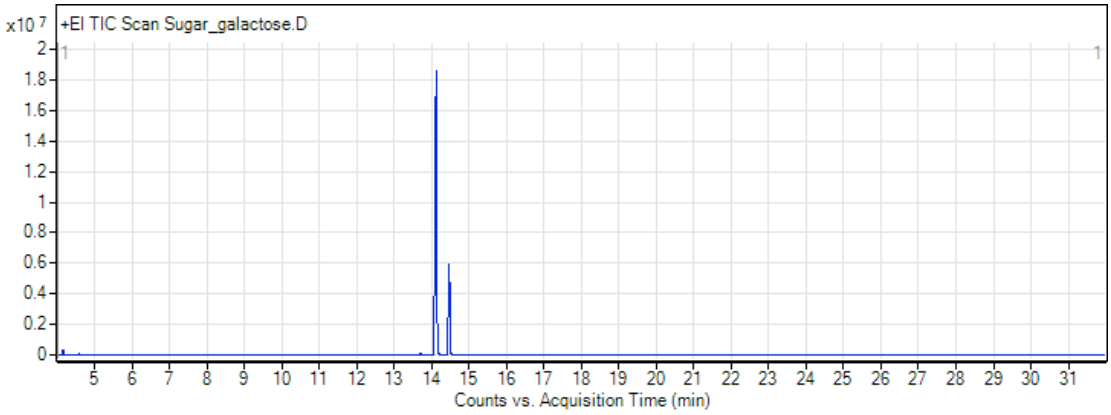
Fructose 6 phosphate



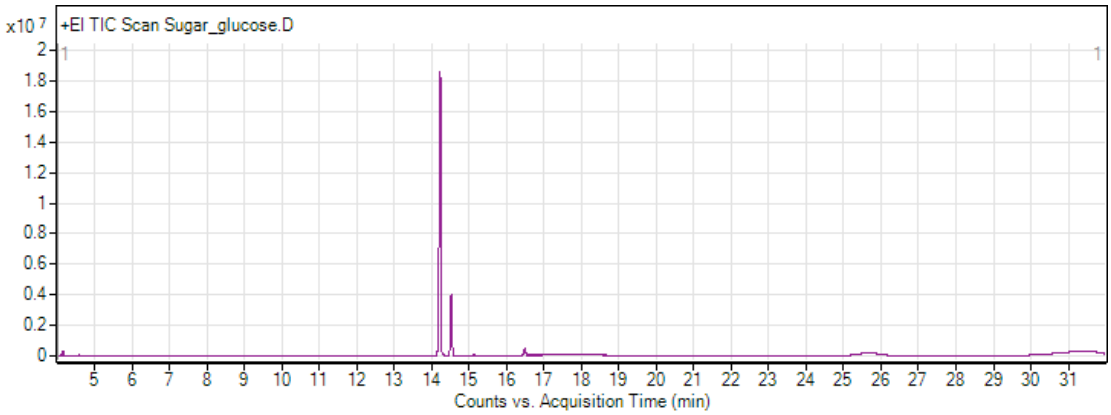
Fucose



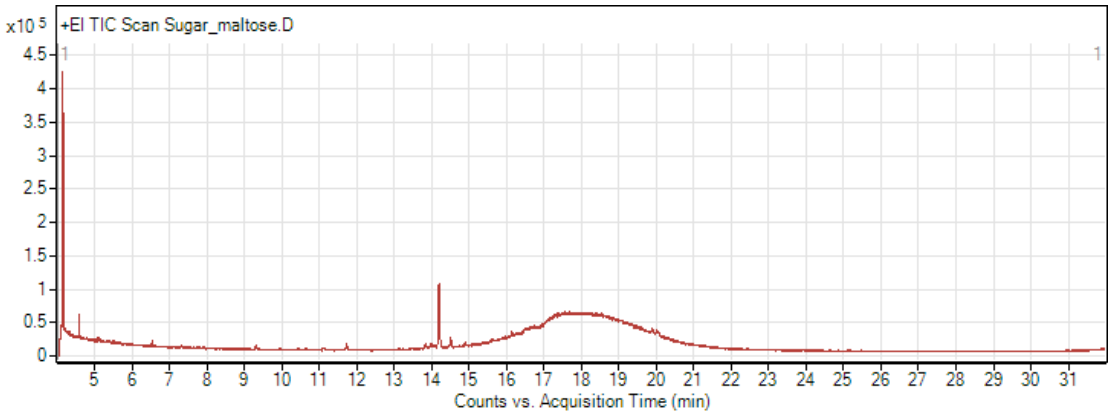
Galactose



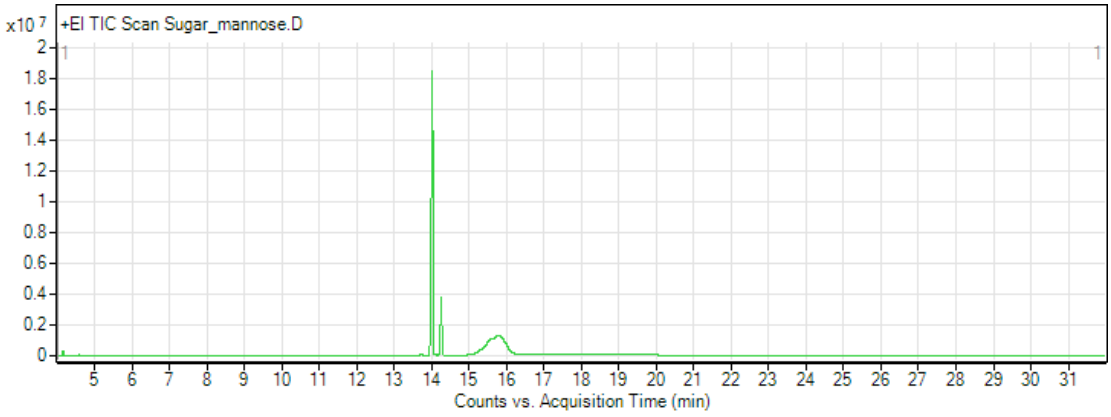
Glucose



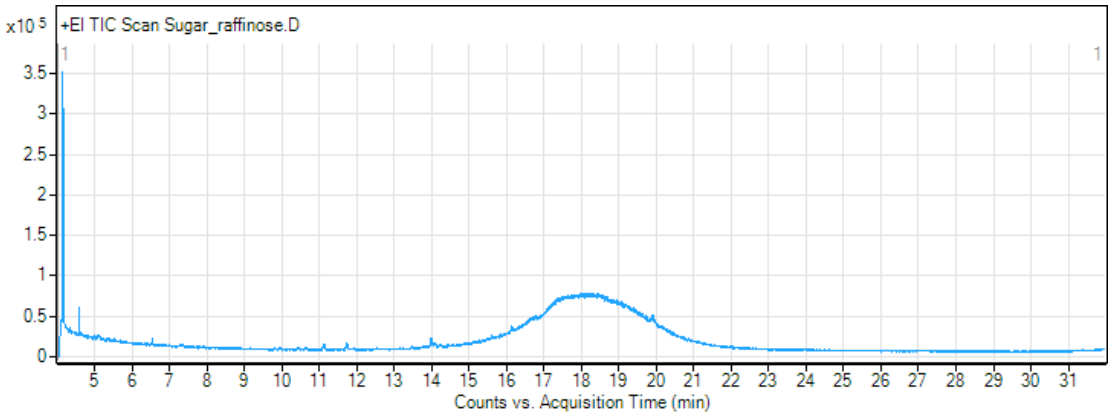
Maltose



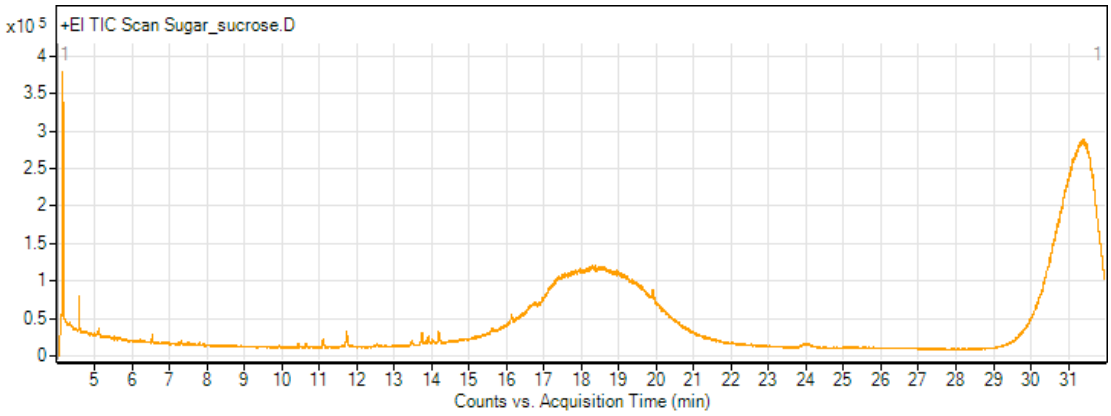
Mannose



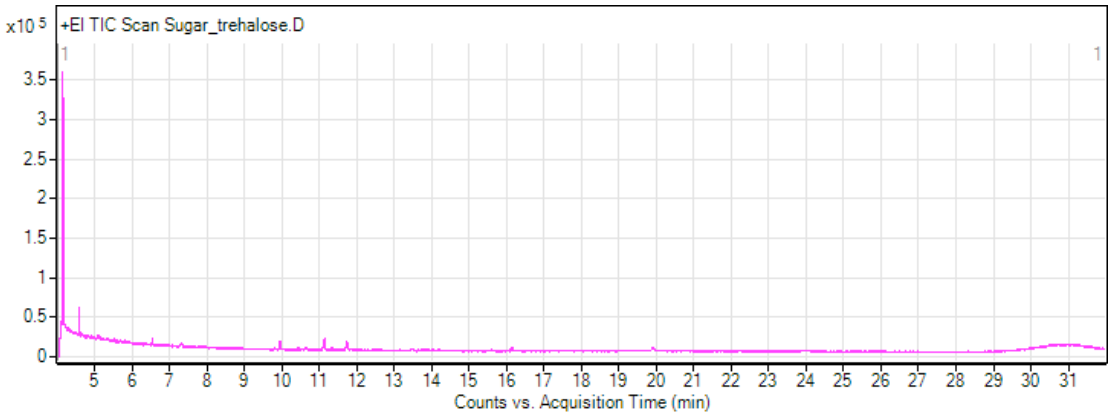
Raffinose



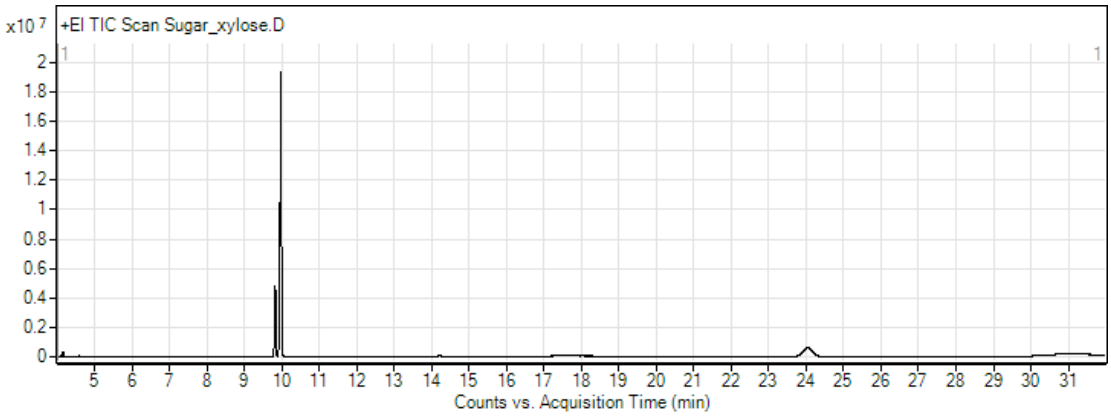
Sucrose



Trehalose

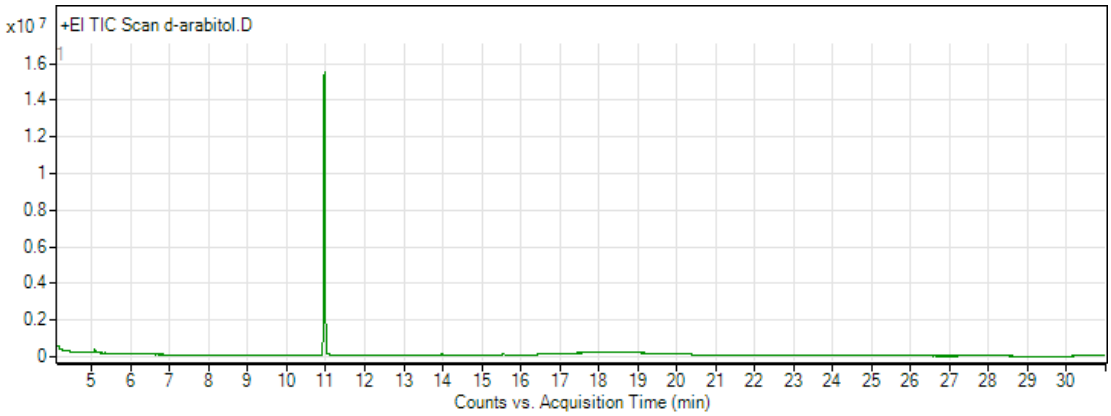


Xylose

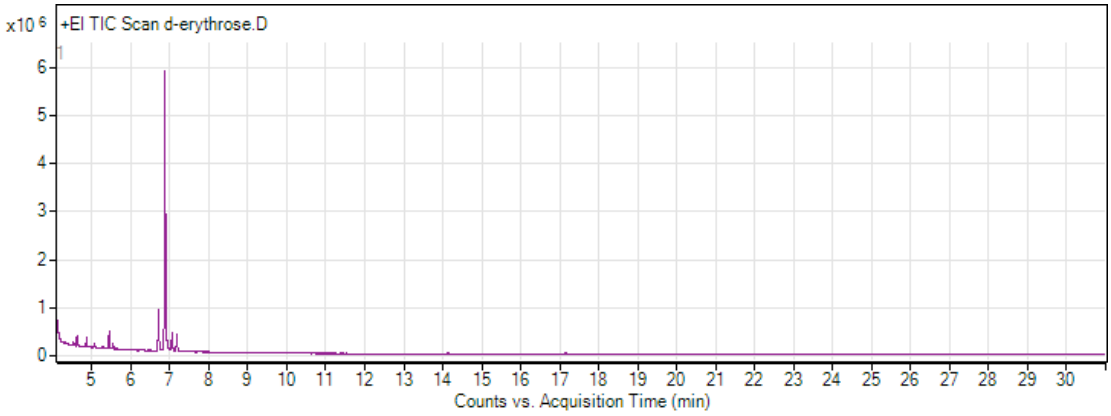


Sugar Alcohols

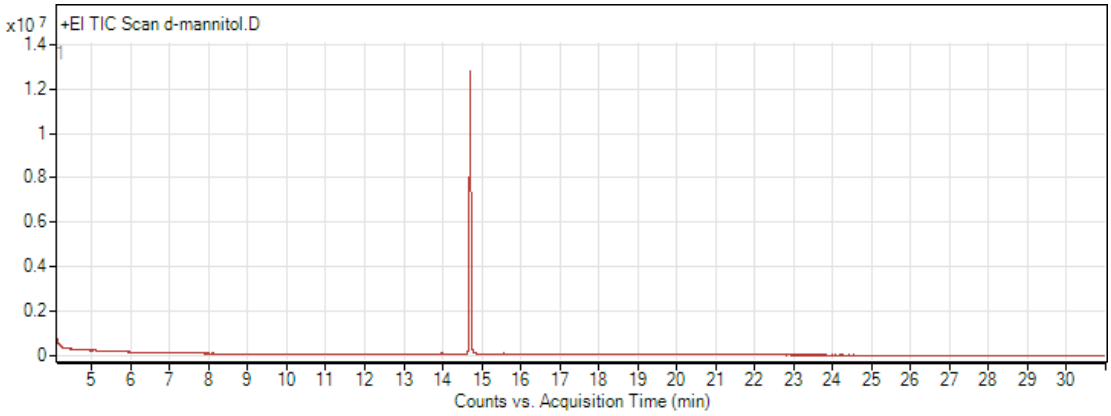
D-Arabitol



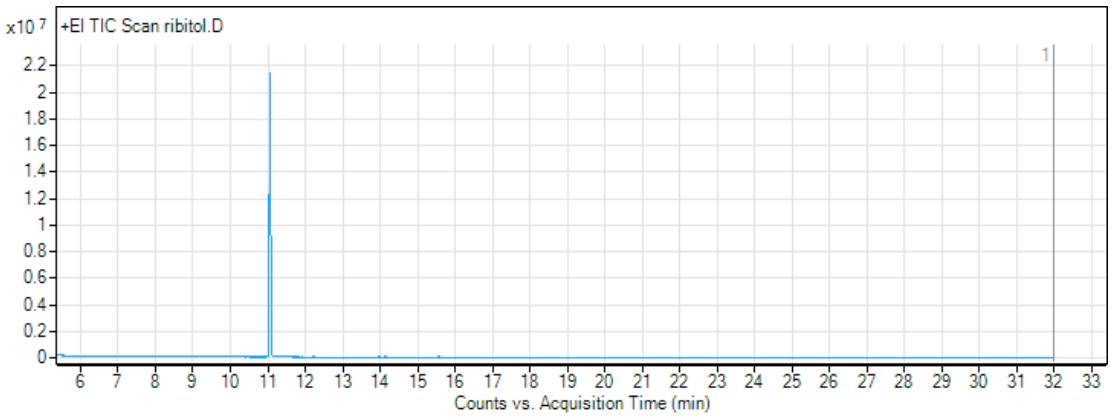
D-Erythrose



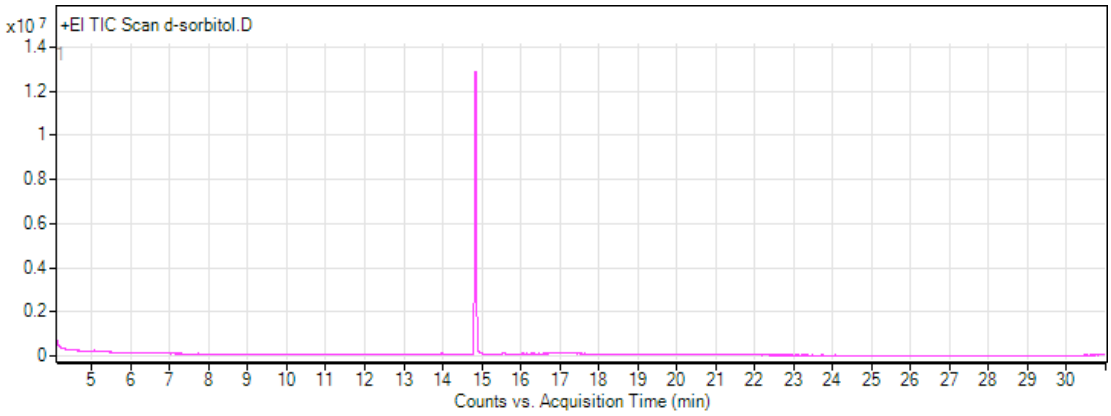
D-Mannitol



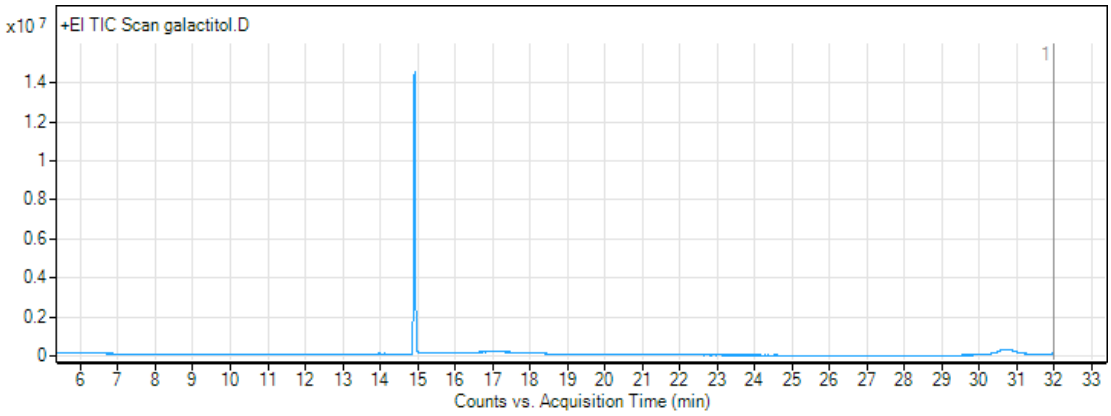
Ribitol (GC internal standard)



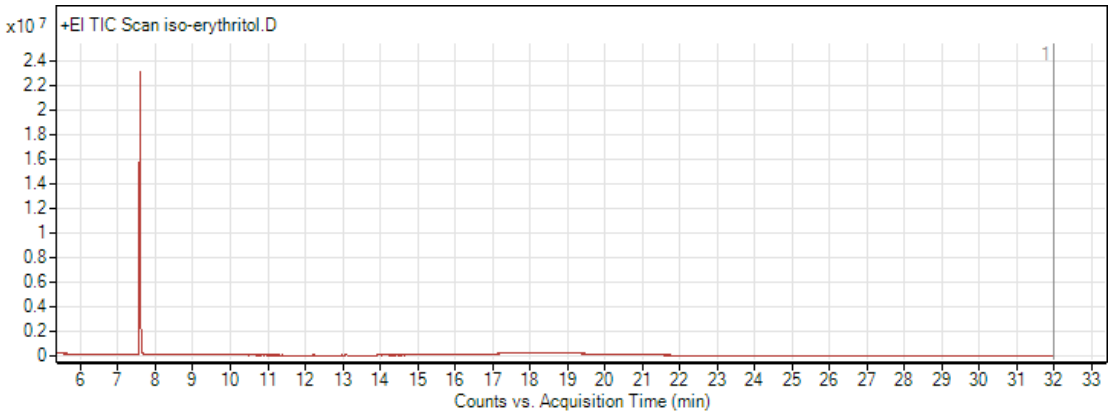
Sorbitol



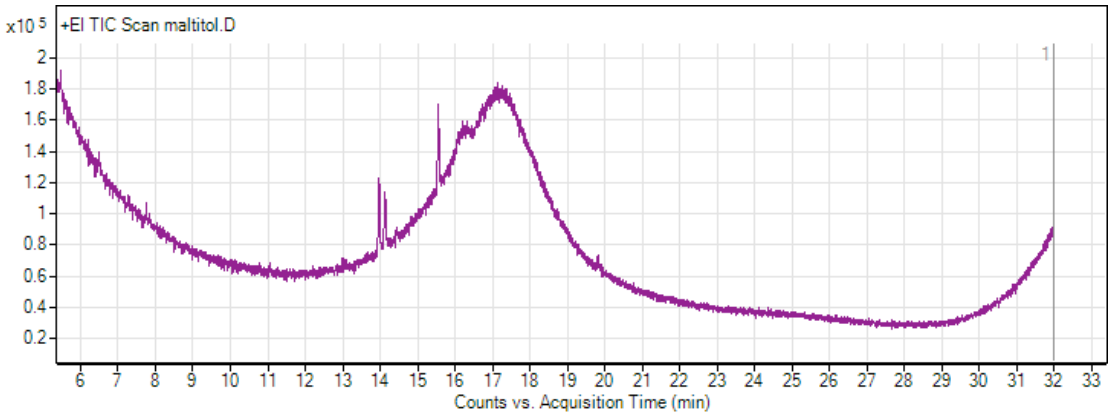
Galactitol



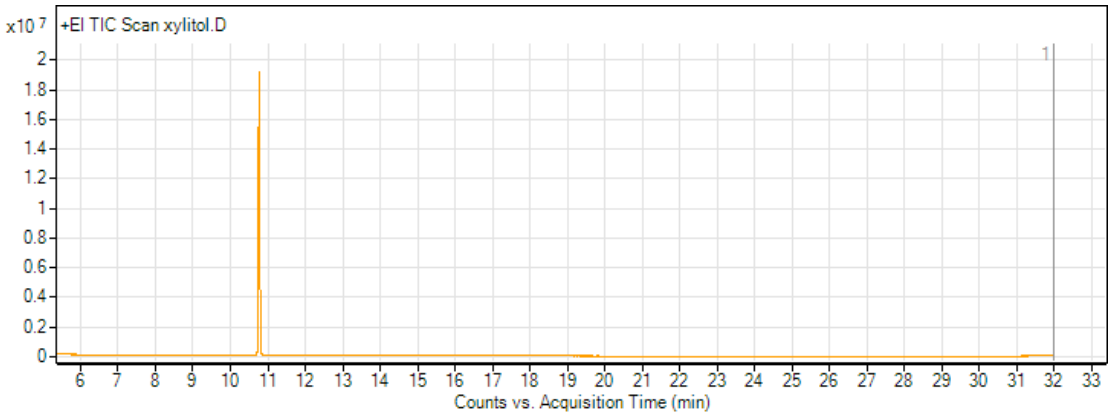
Isoerythritol



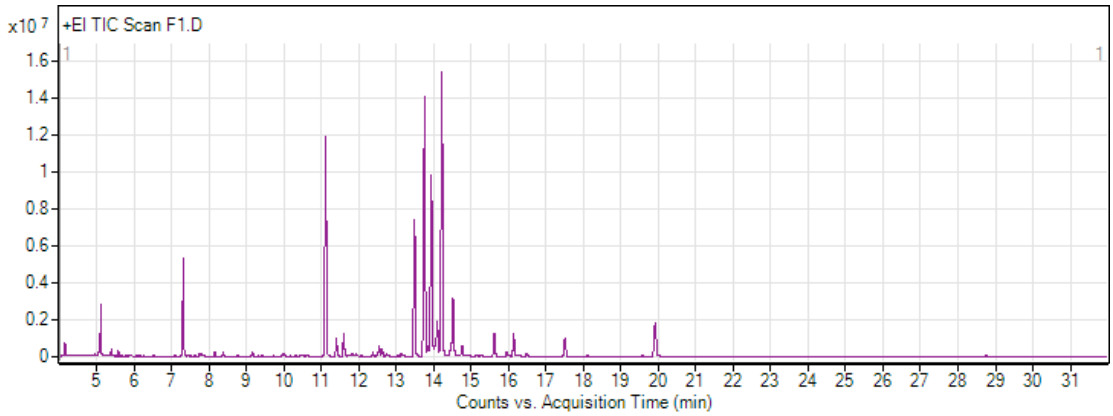
Maltitol

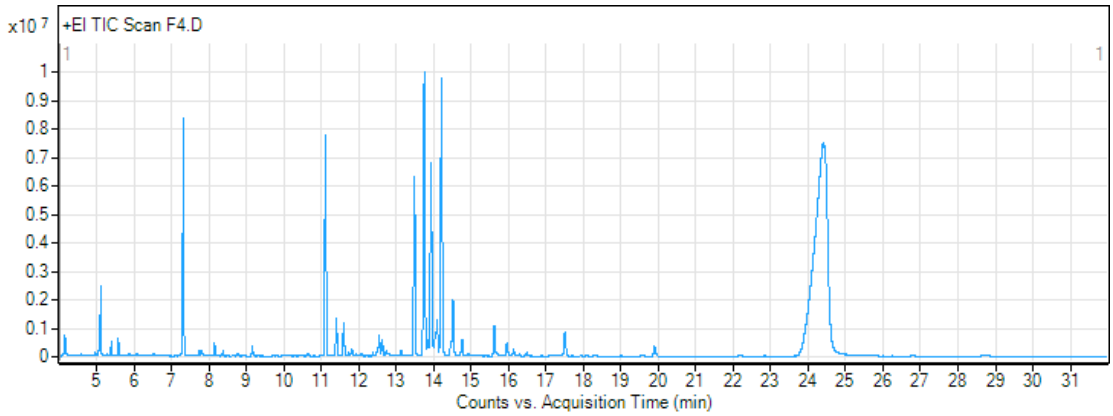


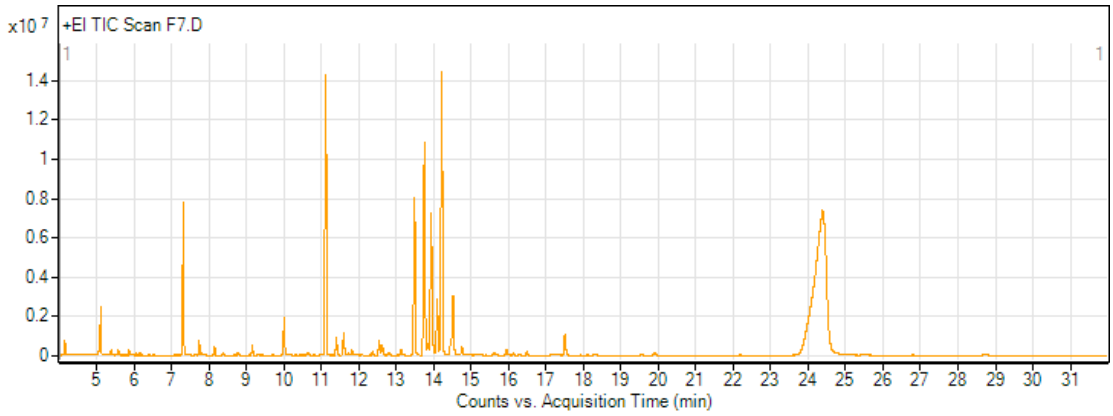
Xylitol

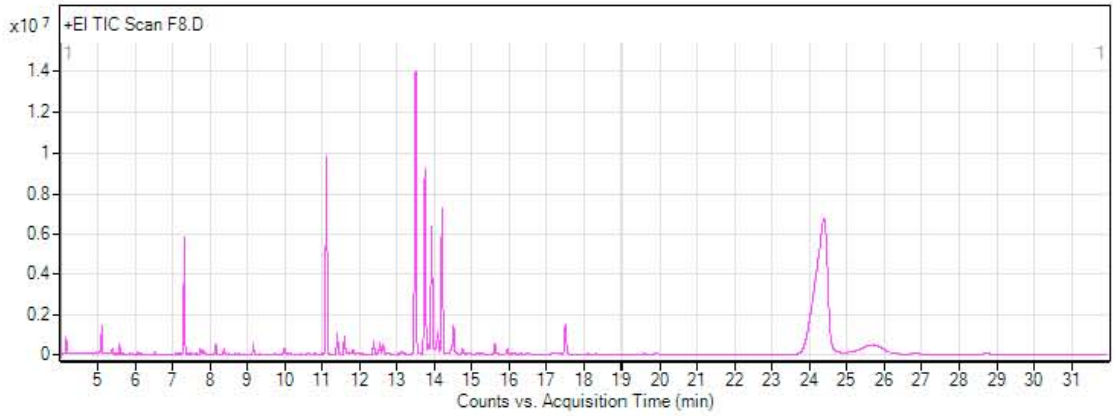


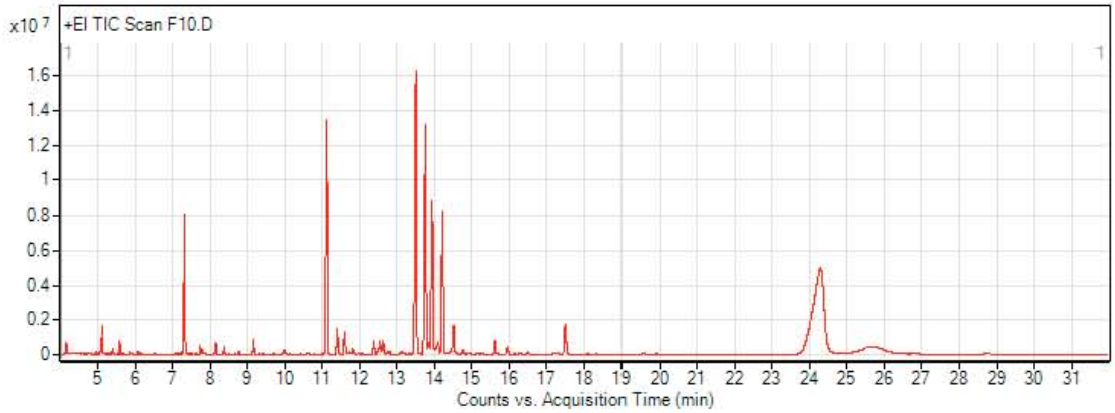
Total Ion Chromatograms of Hydrated Leaf Extracts

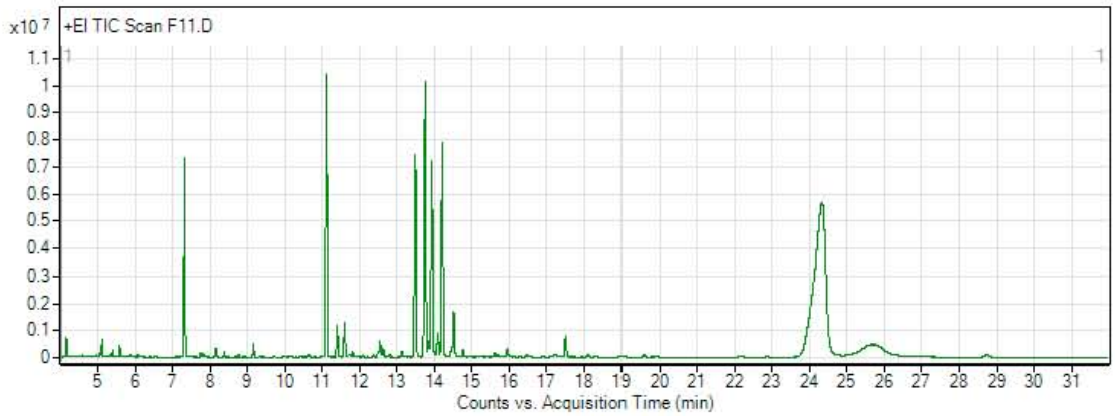


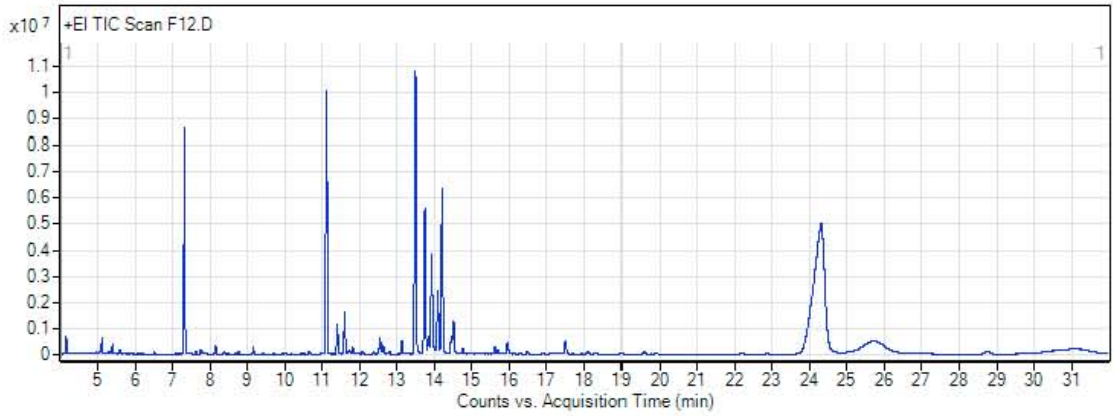


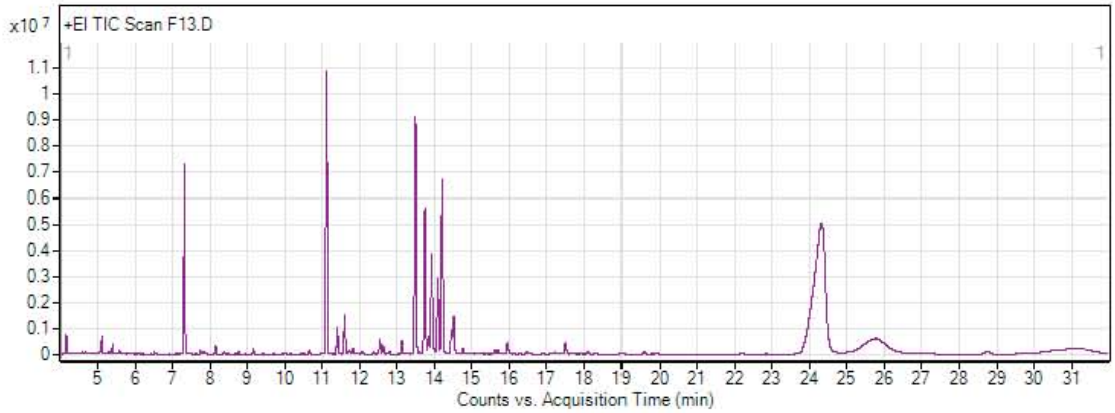


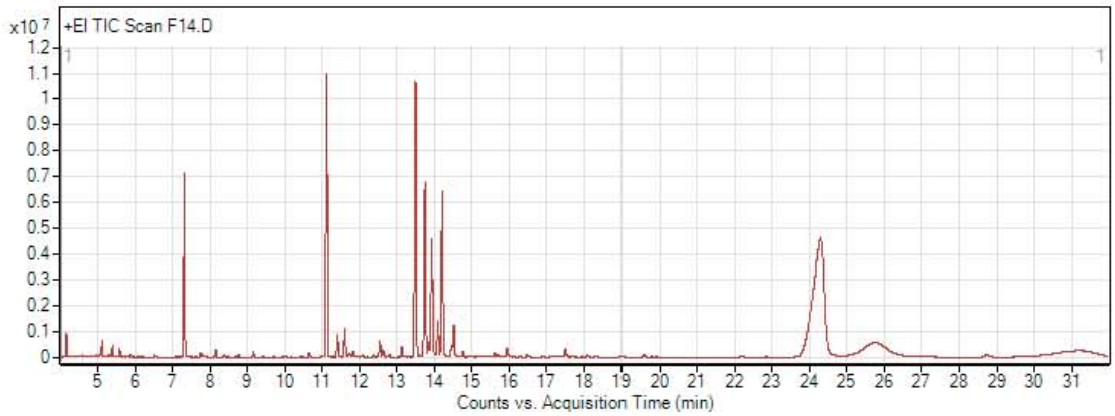

x10 7
+EI TIC Scan F8.D
Counts vs. Acquisition Time (min)

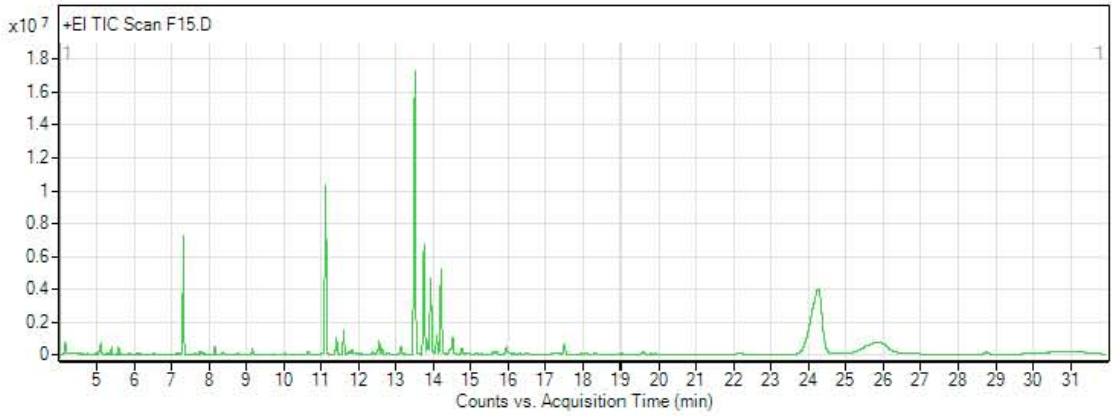

x10 7
+EI TIC Scan F10.D
Counts vs. Acquisition Time (min)


x10 7
+EI TIC Scan F11.D
Counts vs. Acquisition Time (min)

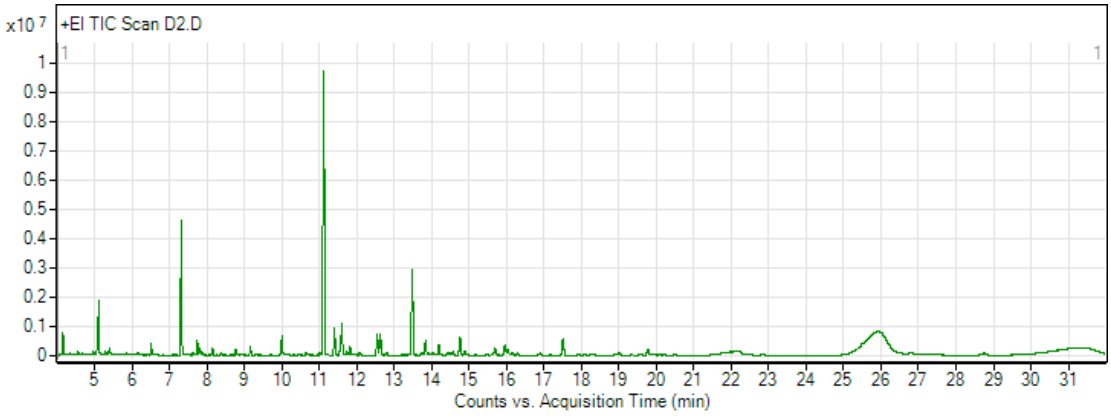

x10 7
+EI TIC Scan F12.D
Counts vs. Acquisition Time (min)

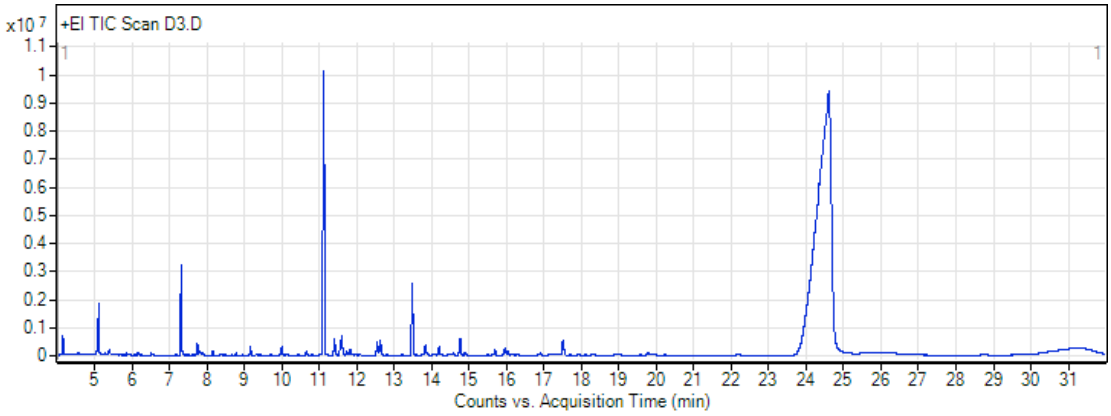

x10 7
+EI TIC Scan F13.D
Counts vs. Acquisition Time (min)

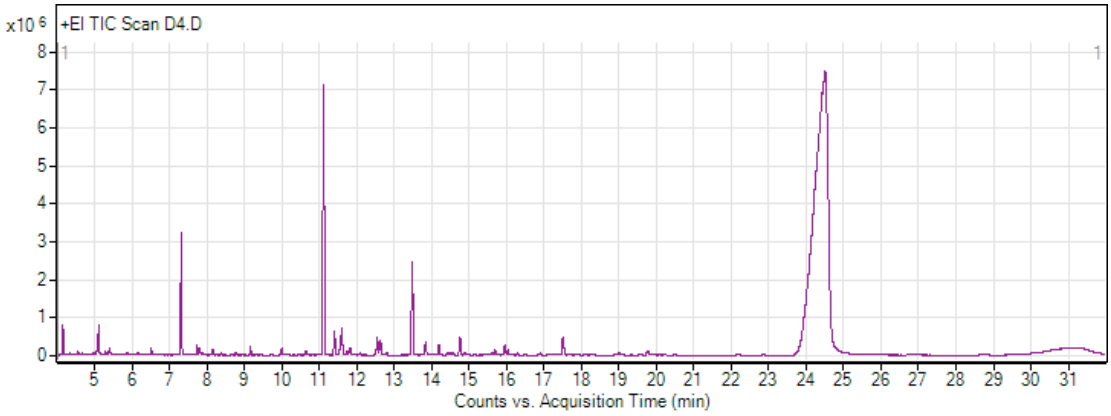

x10 7
+EI TIC Scan F14.D
Counts vs. Acquisition Time (min)

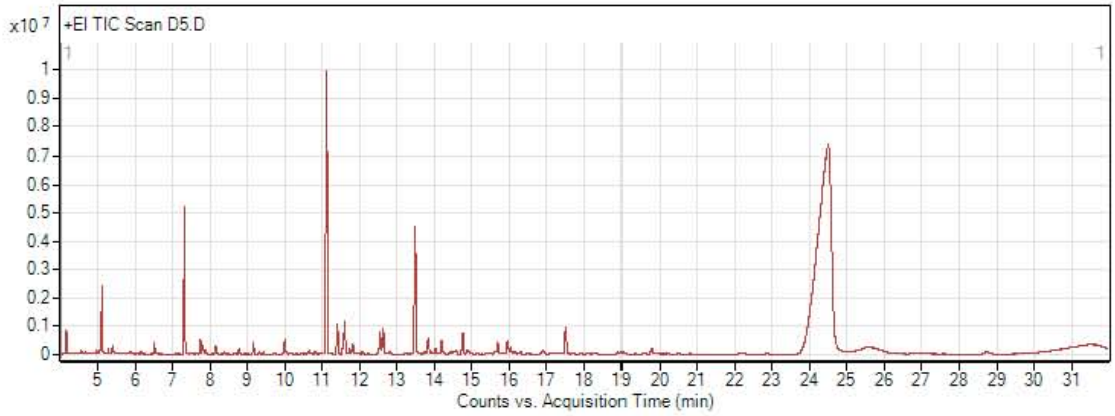

x10 7
+EI TIC Scan F15.D
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Counts vs. Acquisition Time (min)

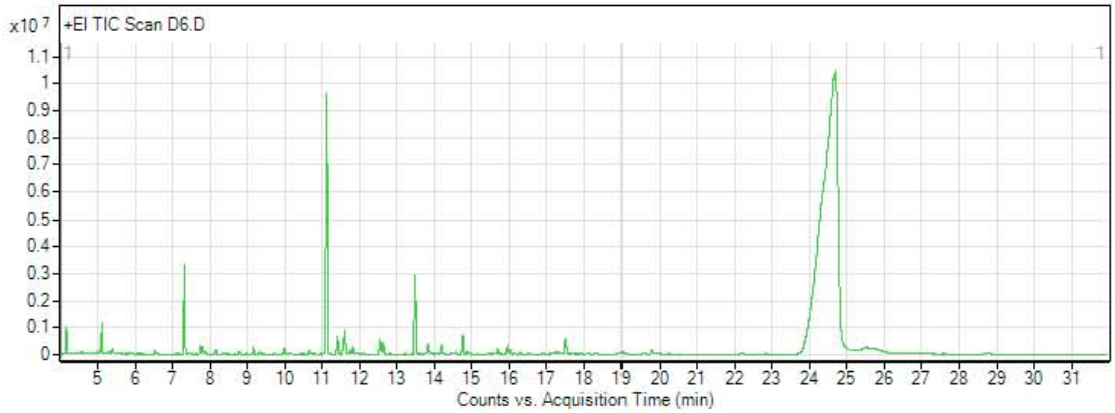
Total Ion Chromatograms of Desiccated Leaf Extracts

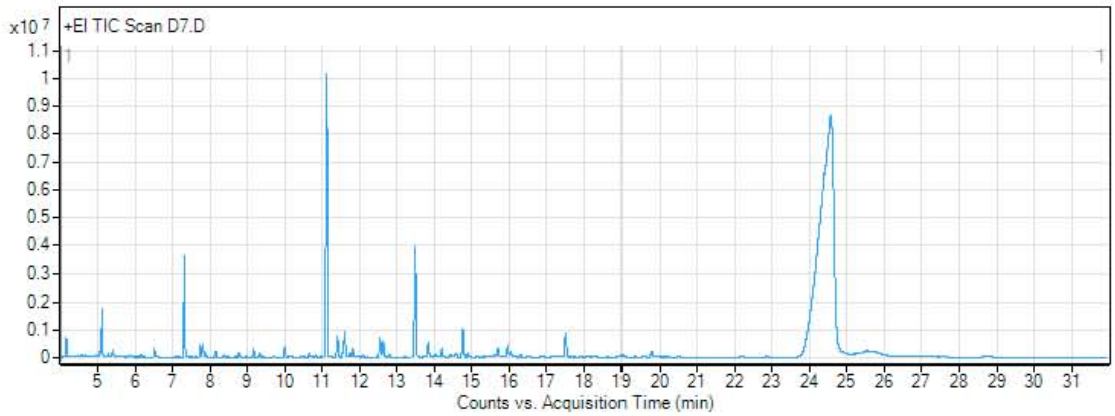


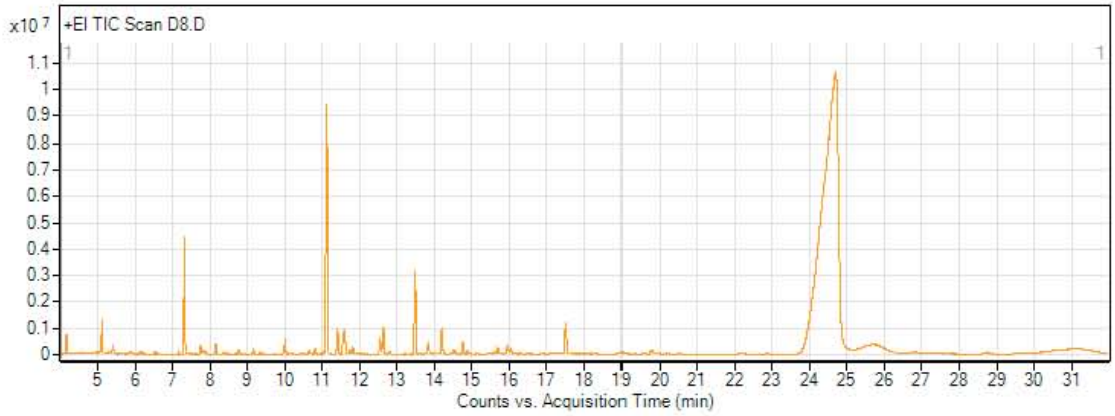


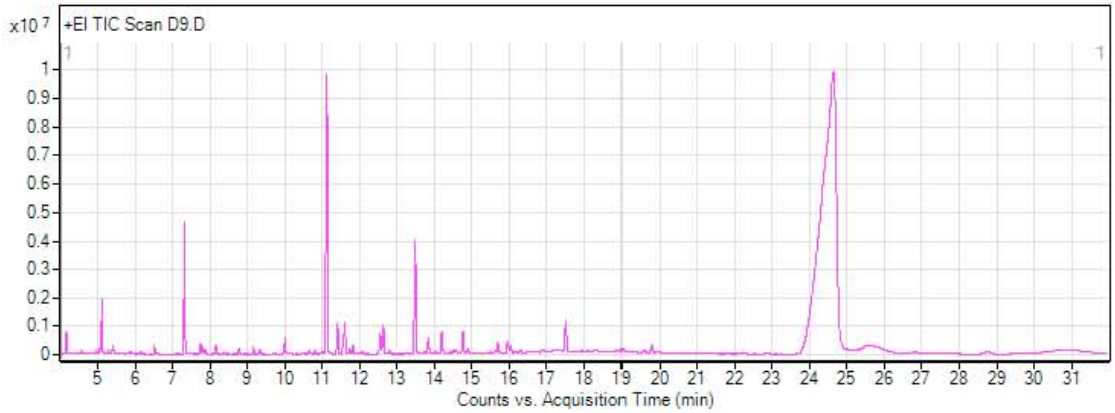


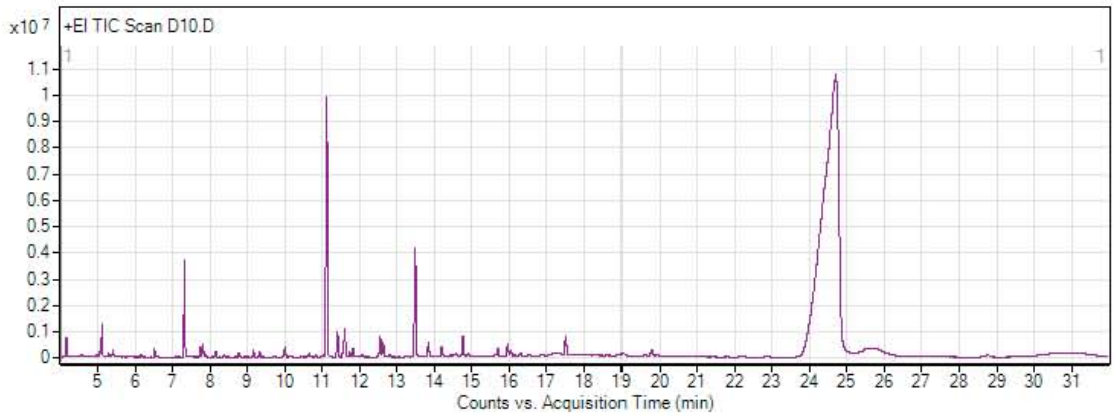

x10 7
+EI TIC Scan D5.D
Counts vs. Acquisition Time (min)

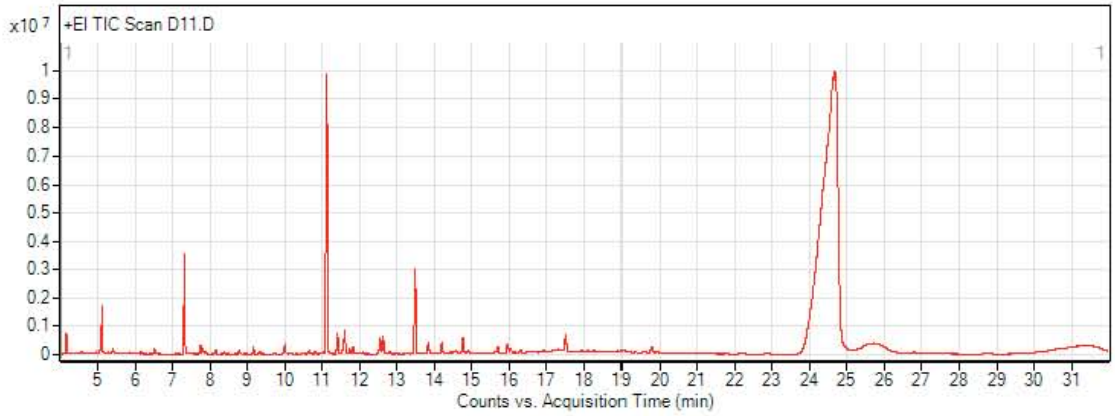

x10 7
+EI TIC Scan D6.D
Counts vs. Acquisition Time (min)


x10 7
+EI TIC Scan D7.D
Counts vs. Acquisition Time (min)

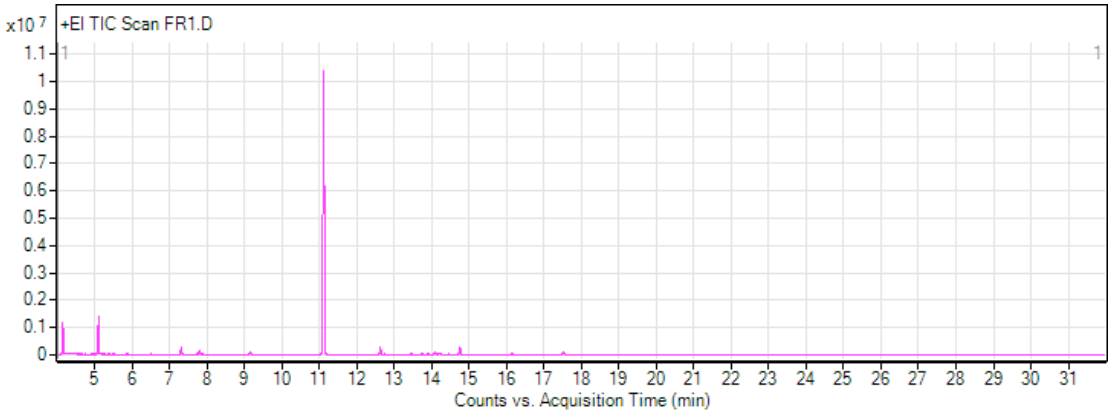

x10 7
+EI TIC Scan D8.D
Counts vs. Acquisition Time (min)

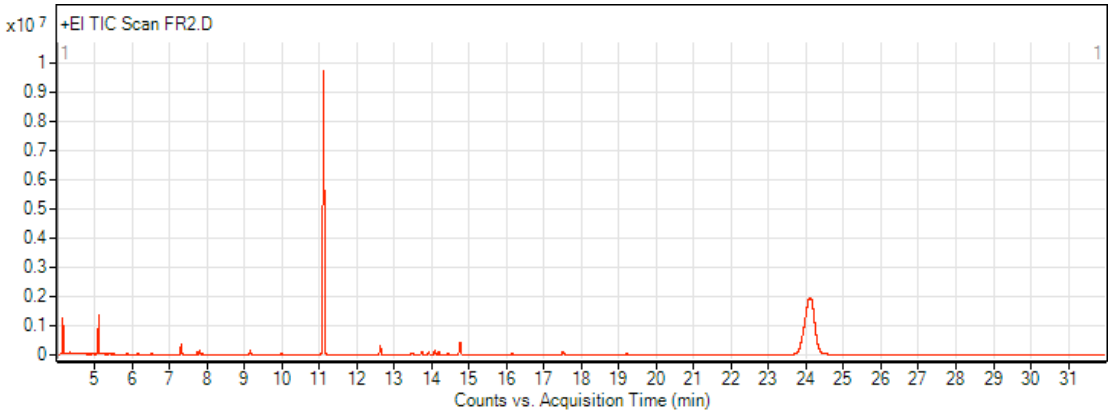

x10 7
+EI TIC Scan D9.D
Counts vs. Acquisition Time (min)

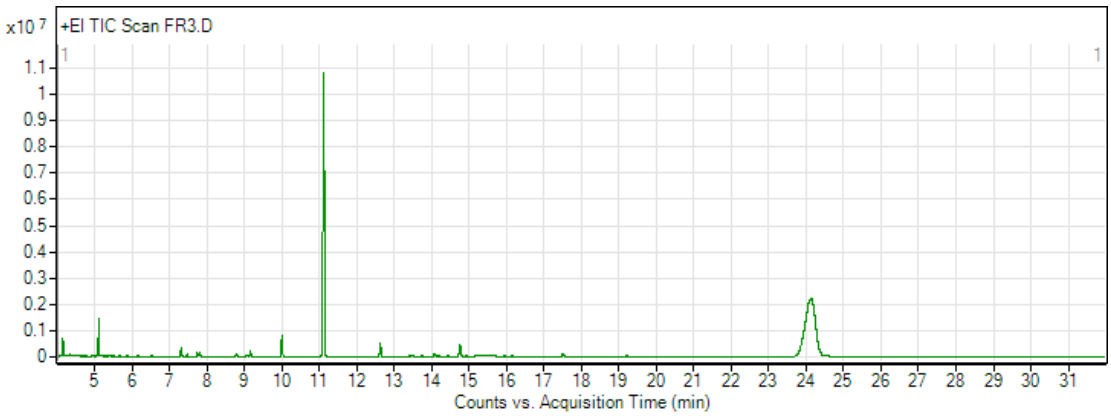

x10 7
+EI TIC Scan D10.D
Counts vs. Acquisition Time (min)

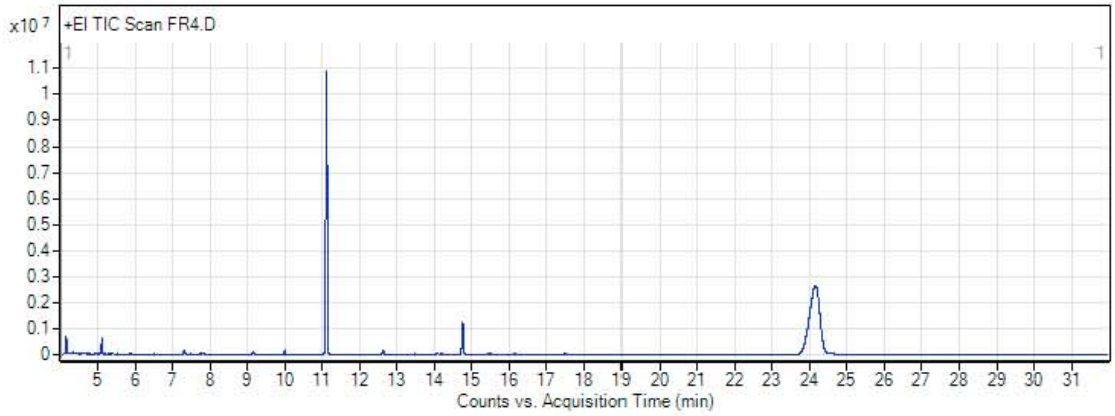

x10 7
+EI TIC Scan D11.D
1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Counts vs. Acquisition Time (min)

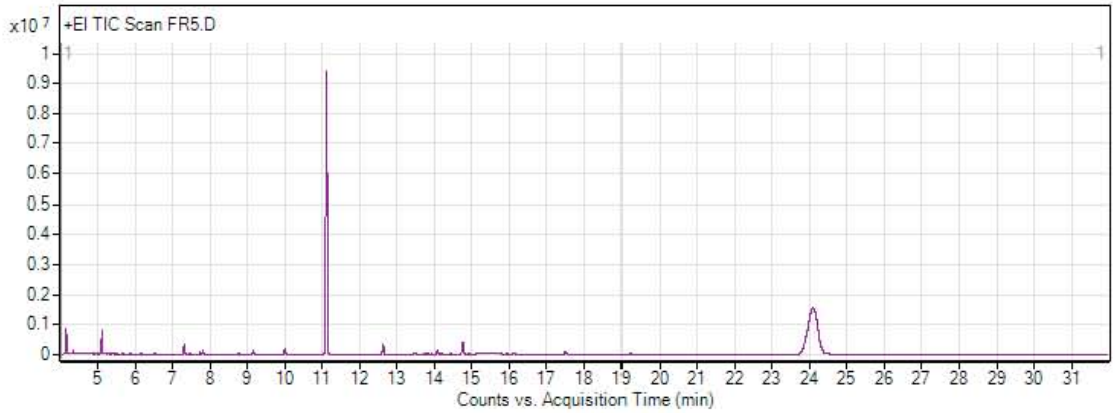
Total Ion Chromatograms of Hydrated Root Extracts

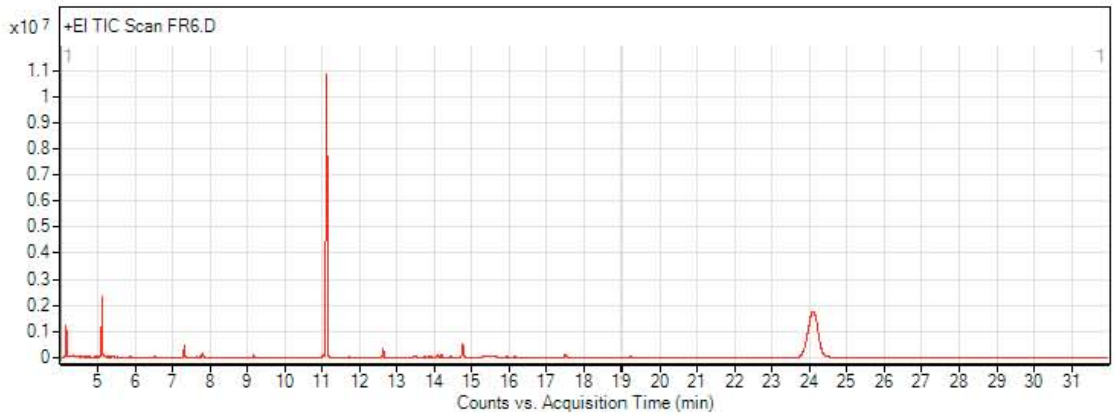


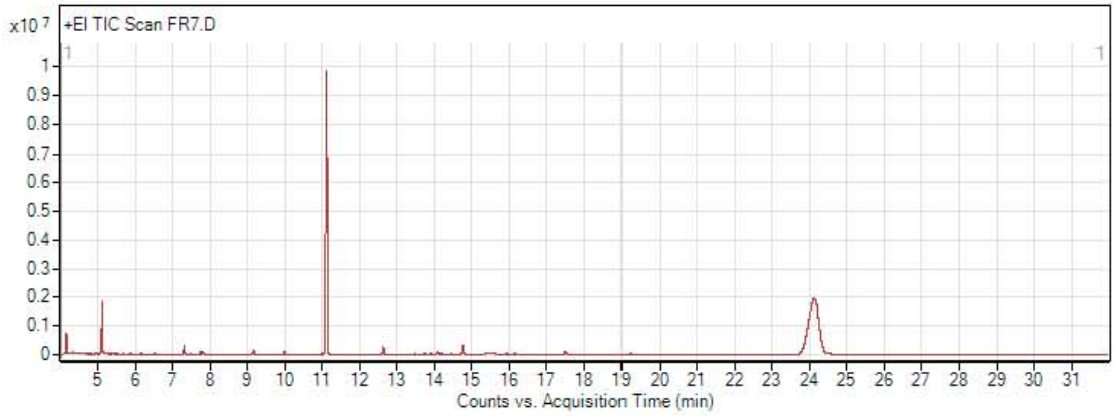


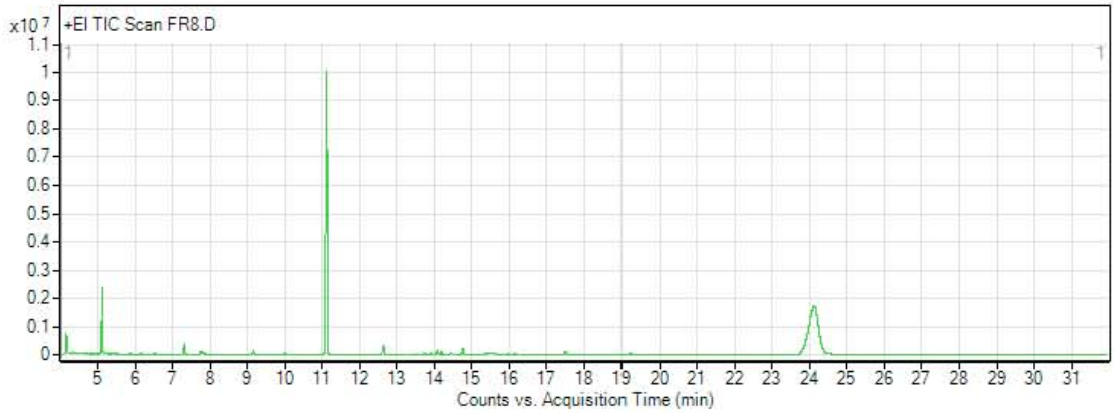


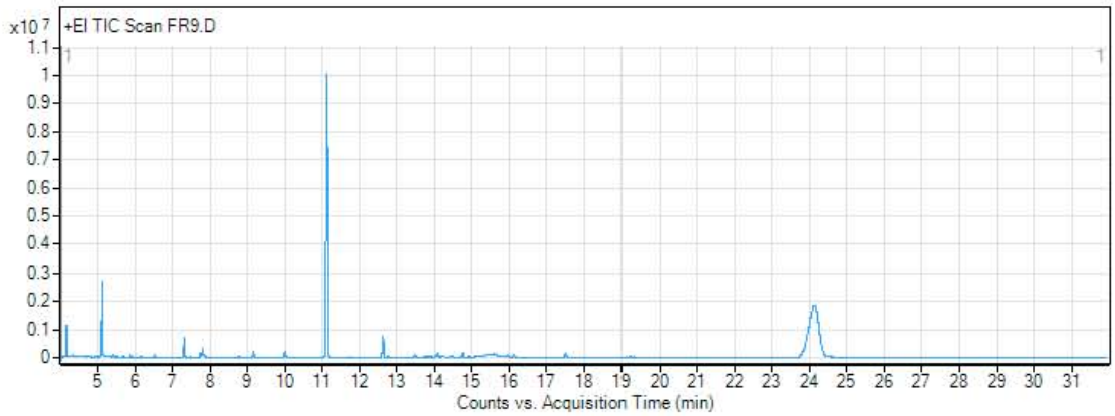

x10 7
+EI TIC Scan FR4.D
Counts vs. Acquisition Time (min)

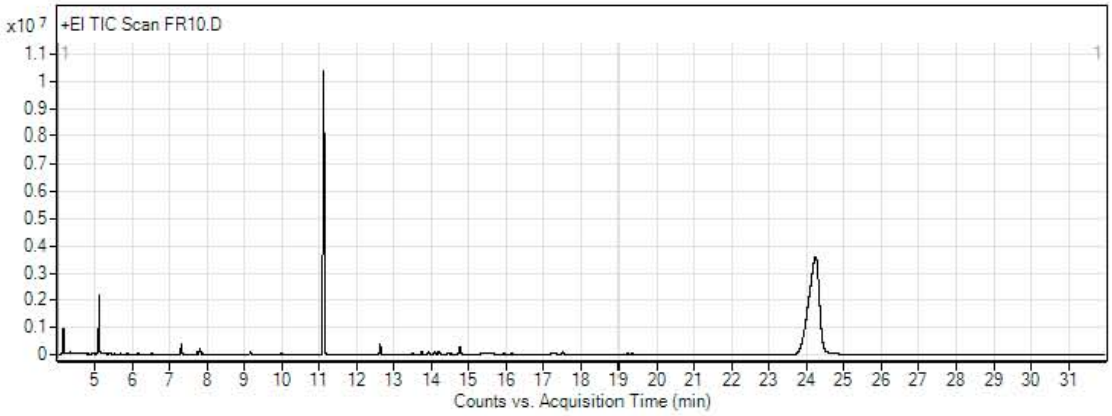

x10 7
+EI TIC Scan FR5.D
Counts vs. Acquisition Time (min)


x10 7
+EI TIC Scan FR6.D
Counts vs. Acquisition Time (min)

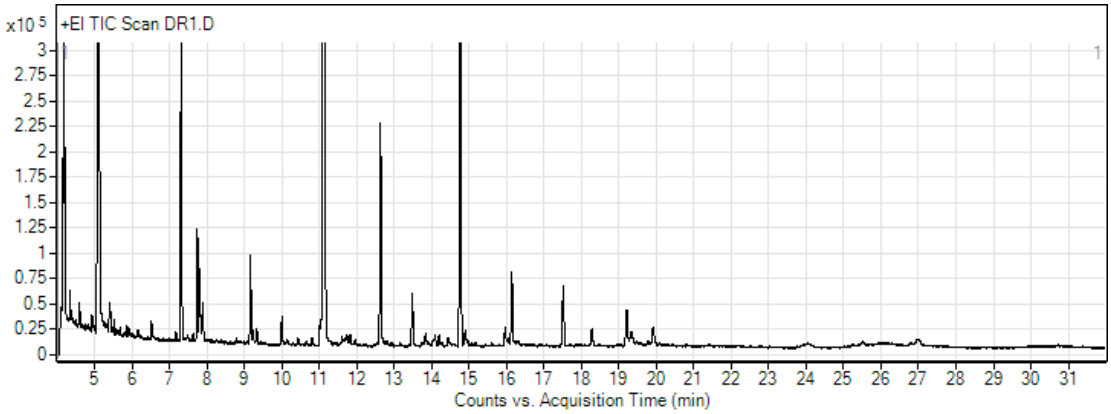

x10 7
+EI TIC Scan FR7.D
Counts vs. Acquisition Time (min)

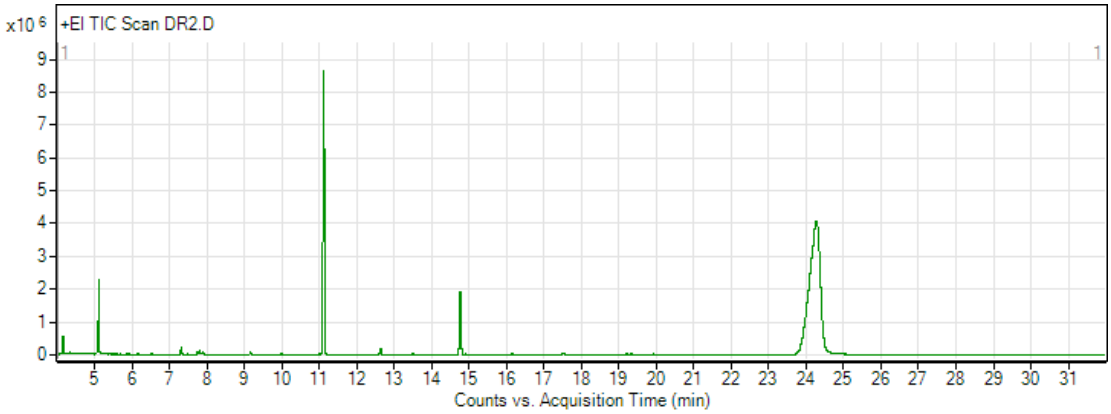

x10 7
+EI TIC Scan FR8.D
Counts vs. Acquisition Time (min)

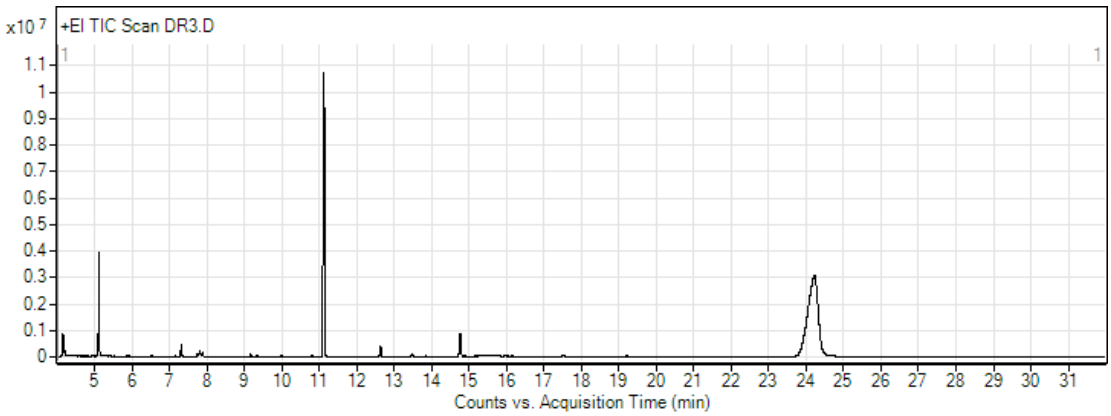

x10 7
+EI TIC Scan FR9.D
Counts vs. Acquisition Time (min)

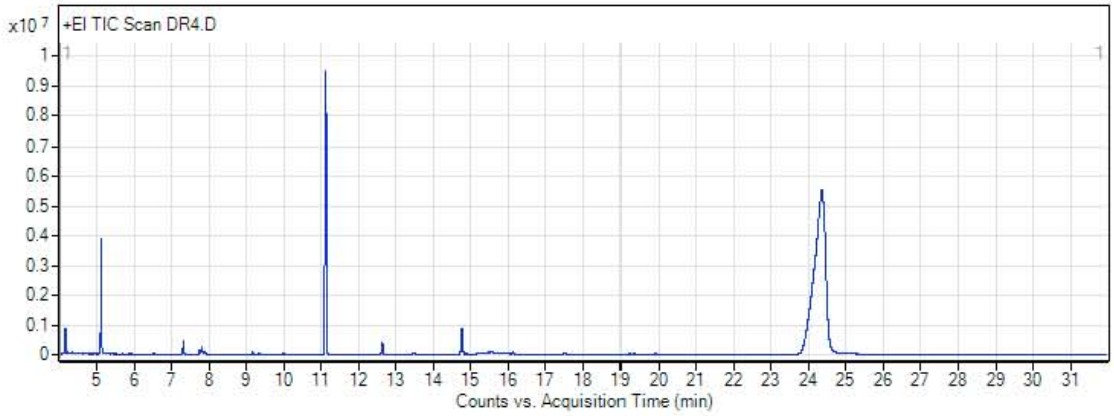

x10 7
+EI TIC Scan FR10.D
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Counts vs. Acquisition Time (min)

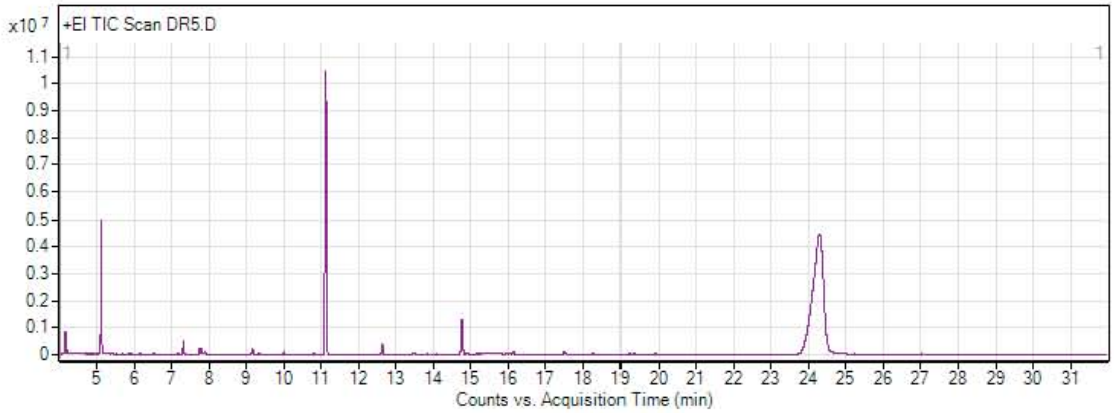
Total Ion Chromatograms of Desiccated Root Extracts

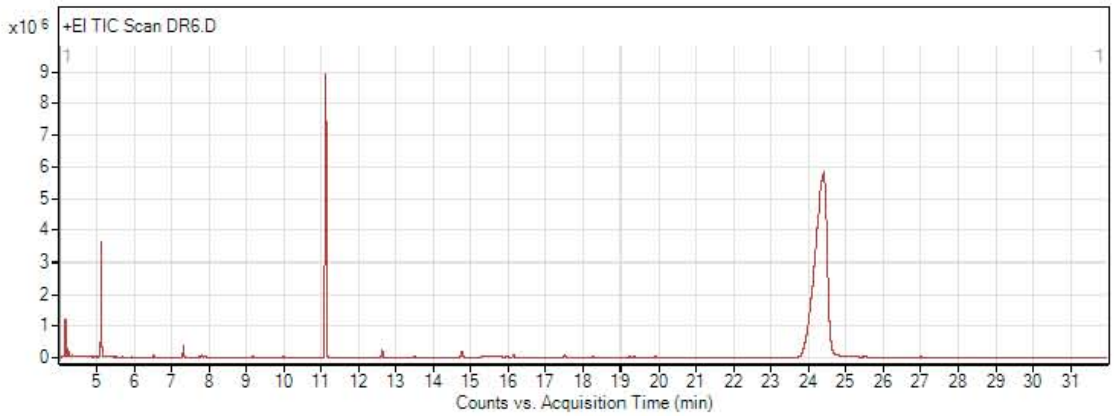


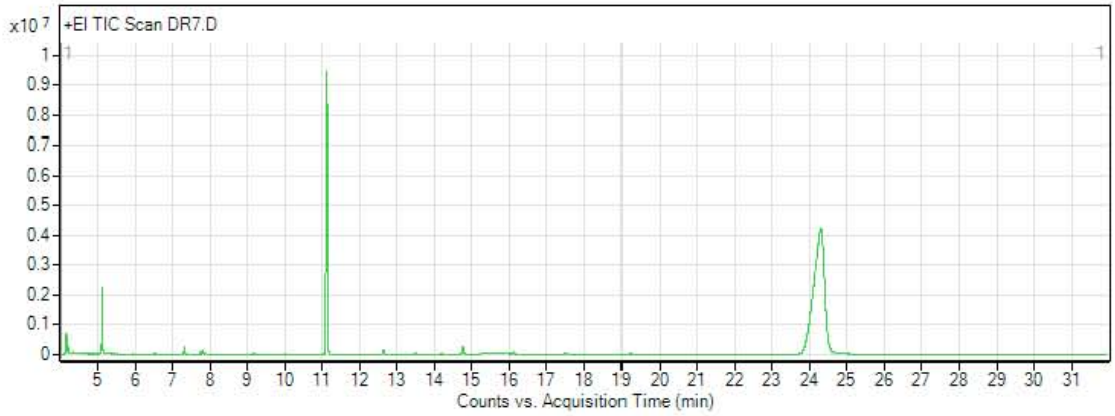


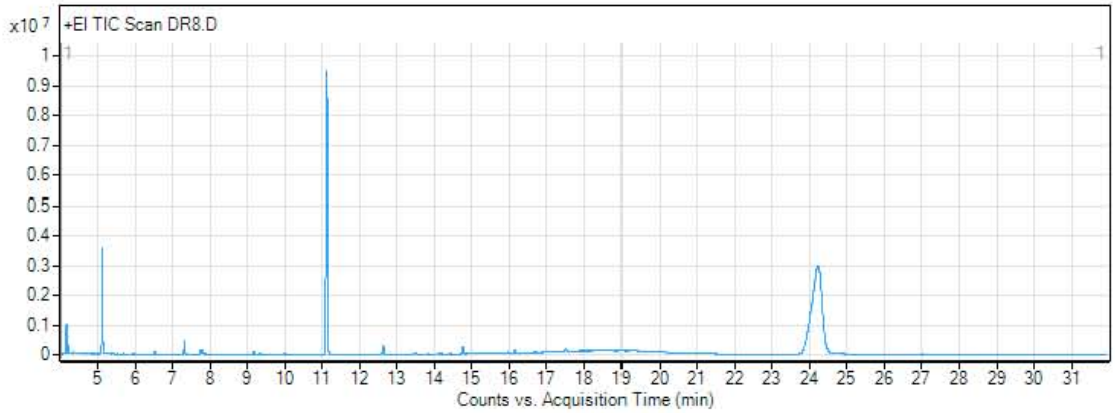


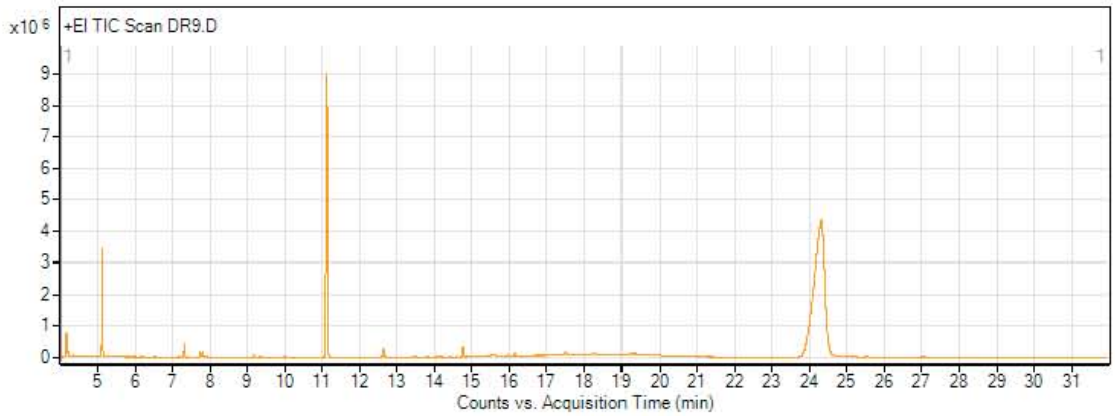

x10 7
+EI TIC Scan DR4.D
Counts vs. Acquisition Time (min)

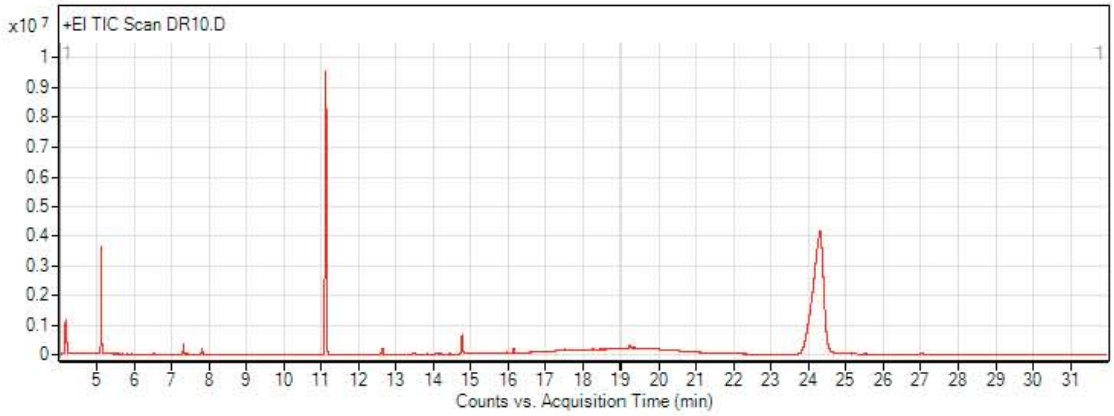

x10 7
+EI TIC Scan DR5.D
Counts vs. Acquisition Time (min)


x10 6
+EI TIC Scan DR6.D
Counts vs. Acquisition Time (min)


x10 7
+EI TIC Scan DR7.D
Counts vs. Acquisition Time (min)


x10 7
+EI TIC Scan DR8.D
Counts vs. Acquisition Time (min)

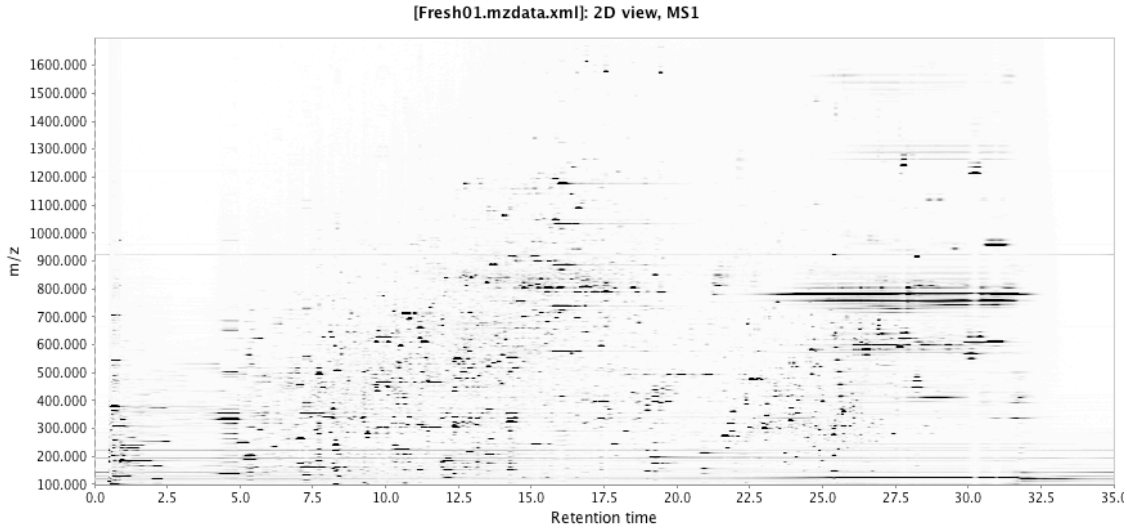

x10 6
+EI TIC Scan DR9.D
Counts vs. Acquisition Time (min)

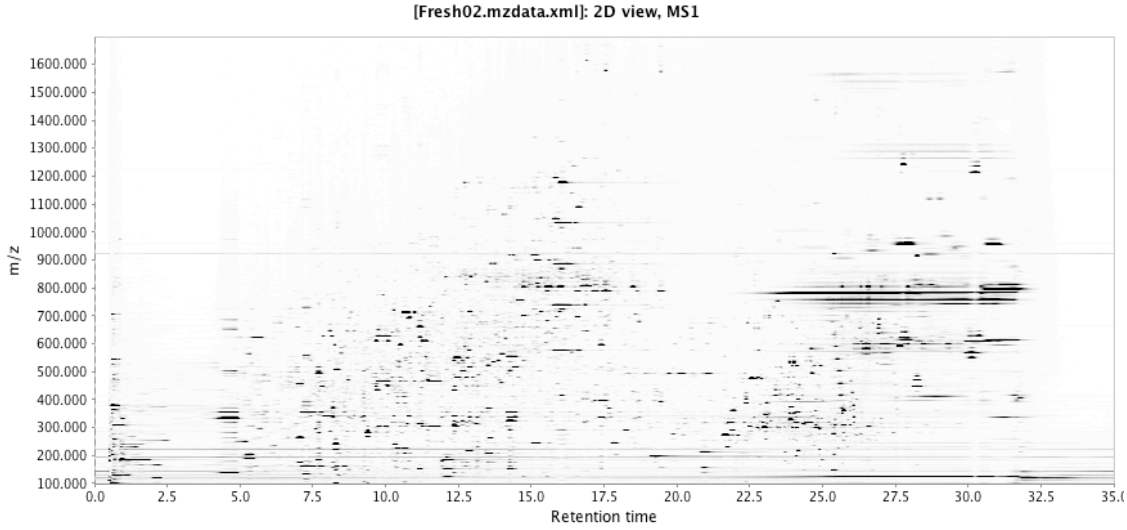

x10 7
+EI TIC Scan DR10.D
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Counts vs. Acquisition Time (min)

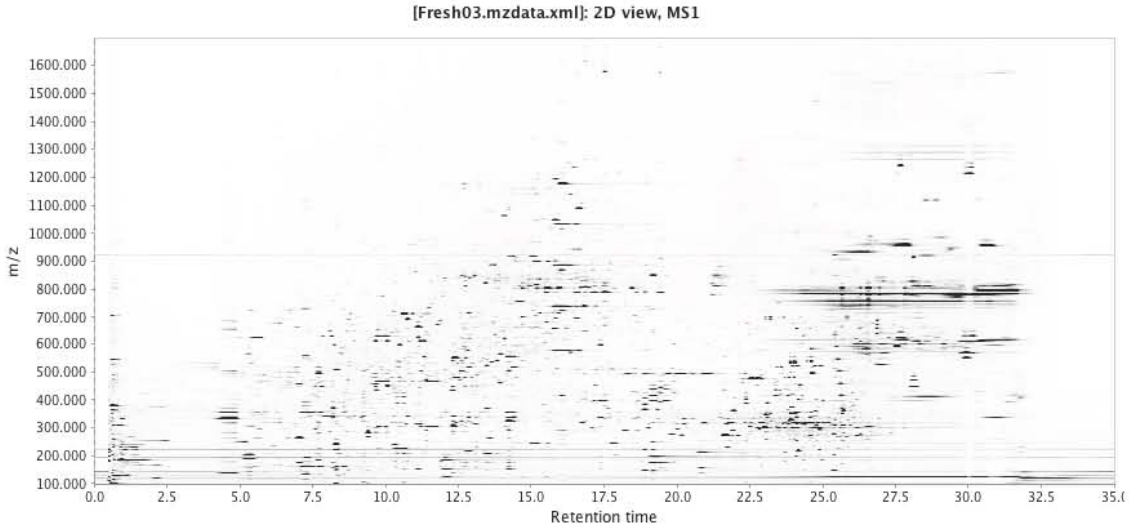
Appendix D: LC-MS Chromatograms

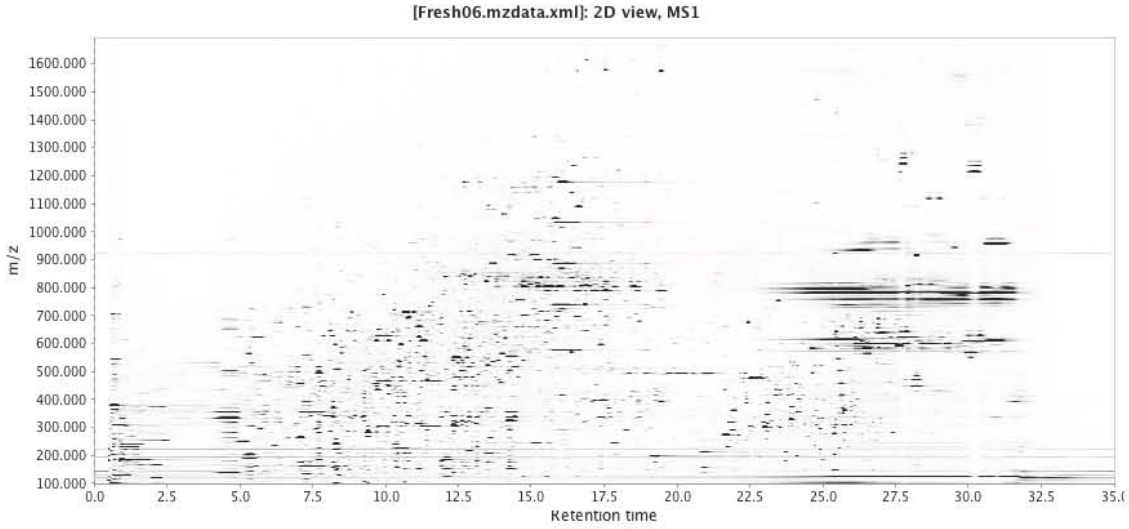
The following 2D visualisations of the LC-MS chromatograms were generated in MZmine 2.10.

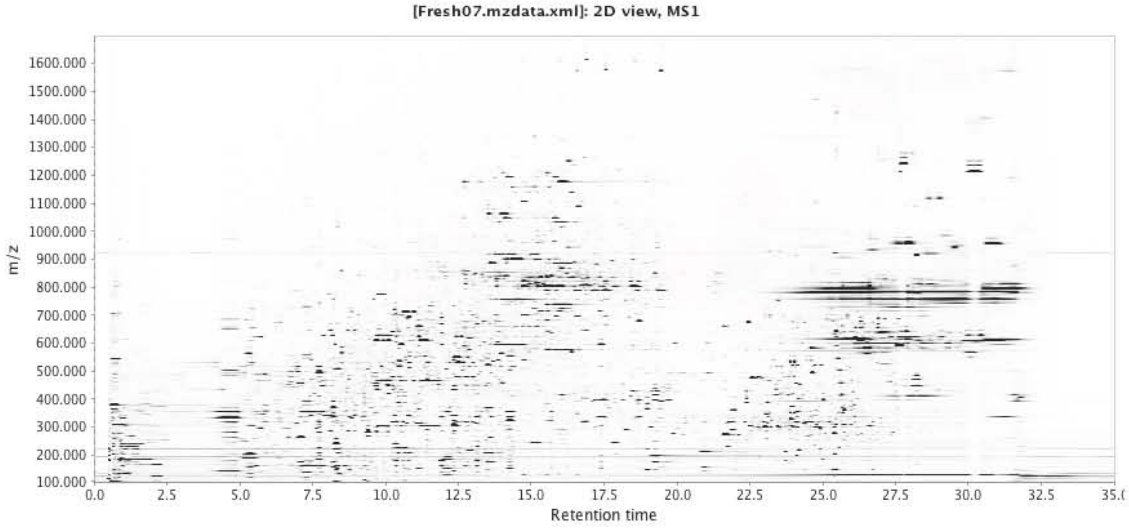
2D Chromatograms of Hydrated Leaf Extracts

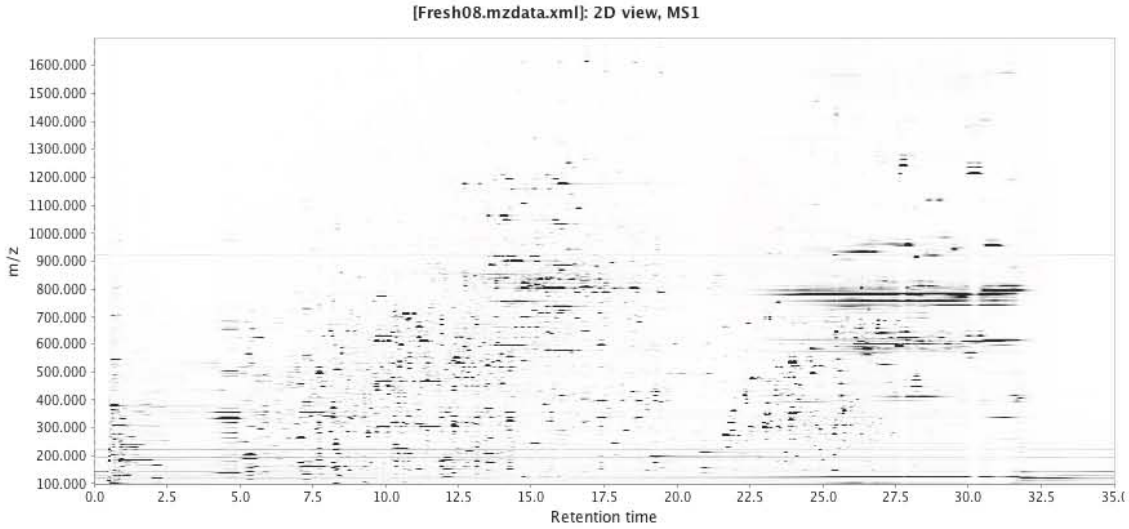


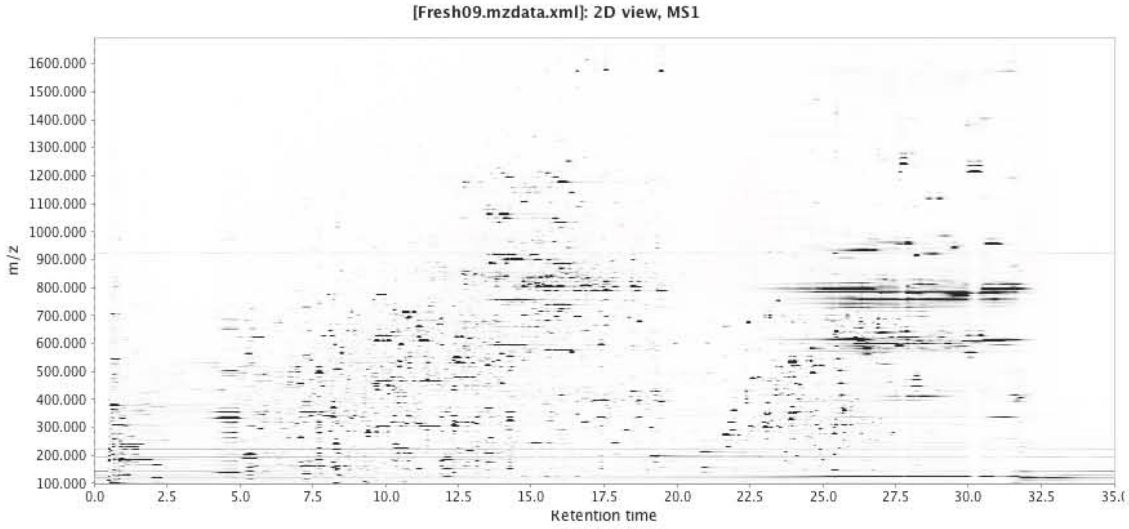


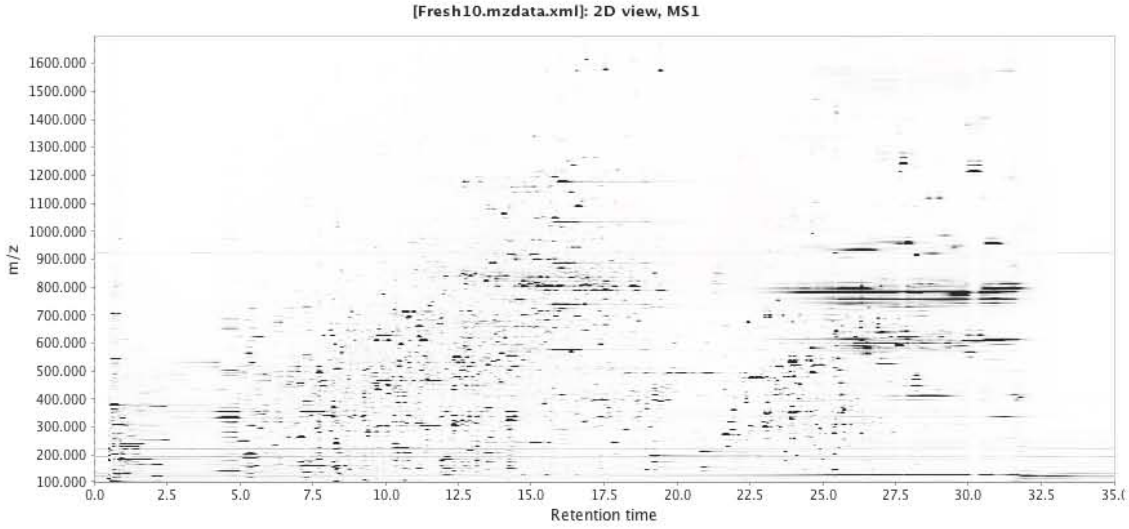

[Fresh03.mzdata.xml]: 2D view, MS1
m/z
Retention time


[Fresh06.mzdata.xml]: 2D view, MS1
m/z
Retention time

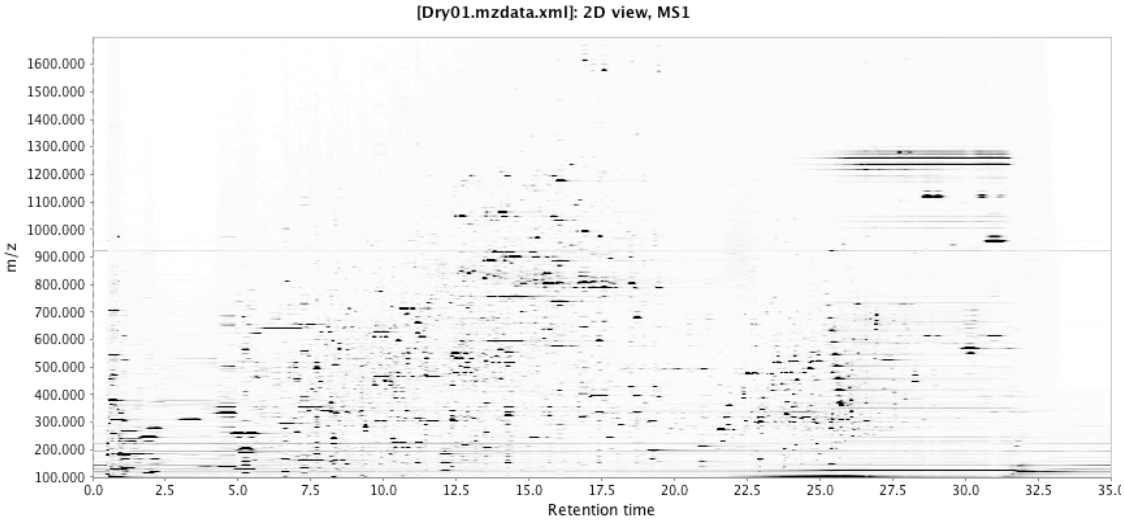

[Fresh07.mzdata.xml]: 2D view, MS1
m/z
Retention time

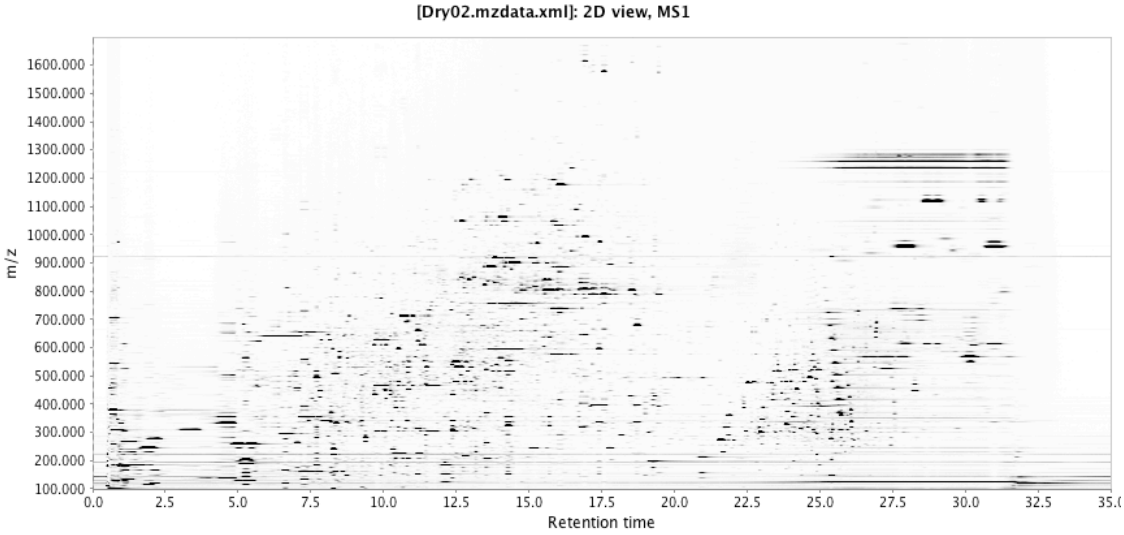

[Fresh08.mzdata.xml]: 2D view, MS1
m/z
Retention time

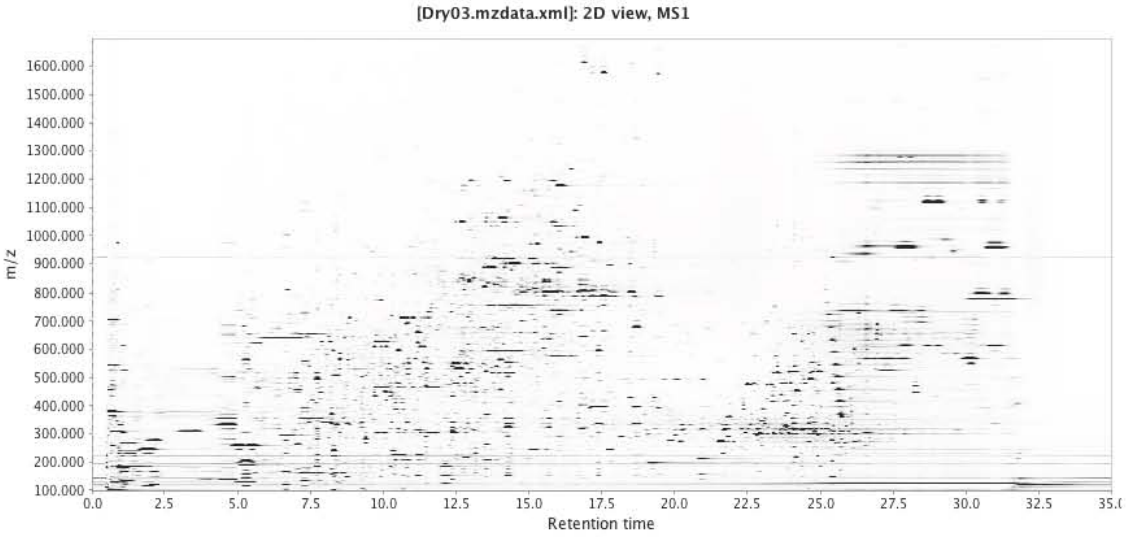

[Fresh09.mzdata.xml]: 2D view, MS1
m/z
Retention time

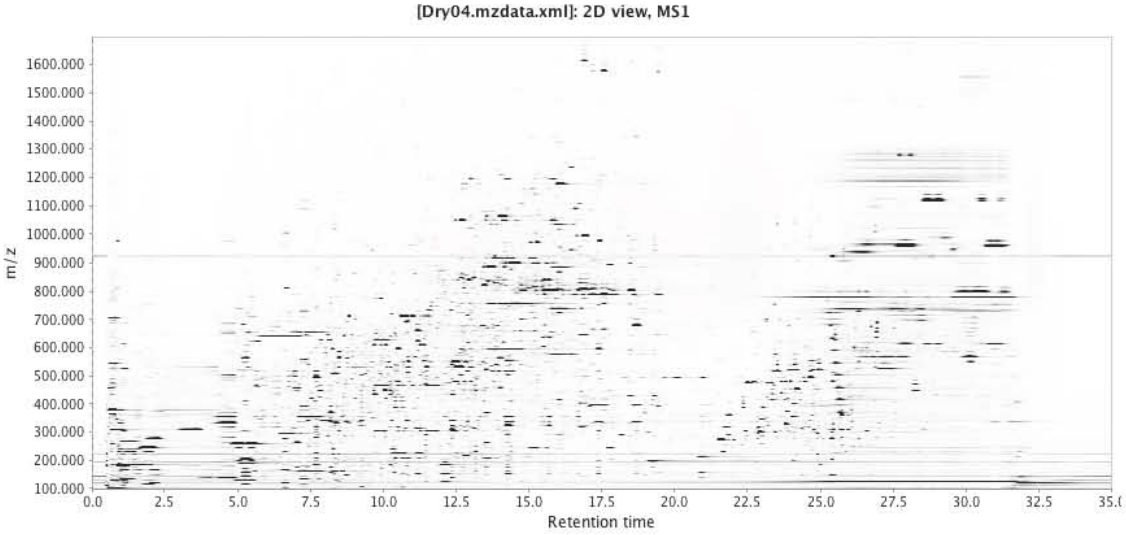

[Fresh10.mzdata.xml]: 2D view, MS1
m/z
Retention time

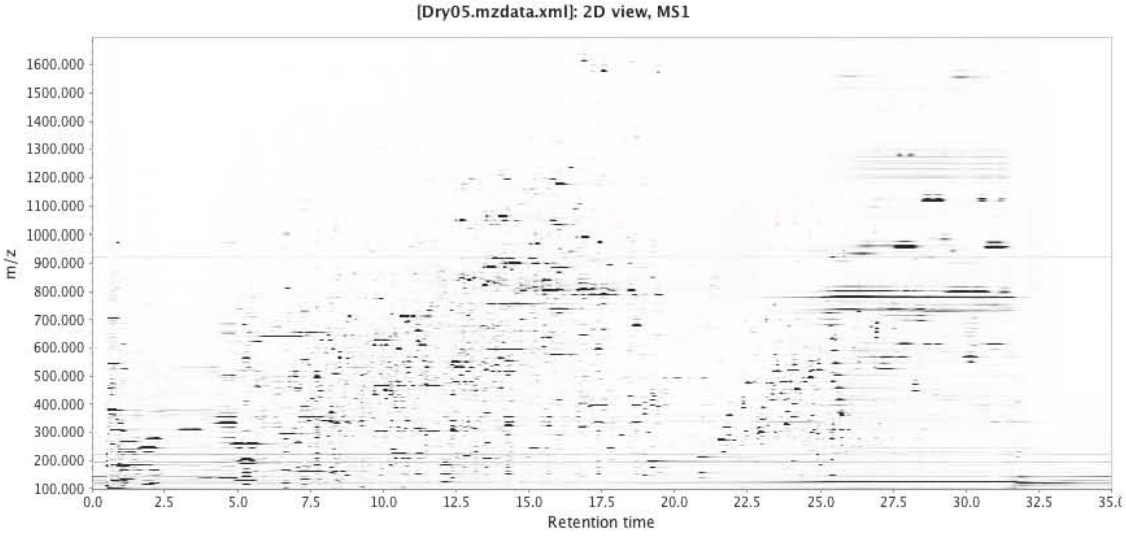
2D Chromatograms of Desiccated Leaf Extracts

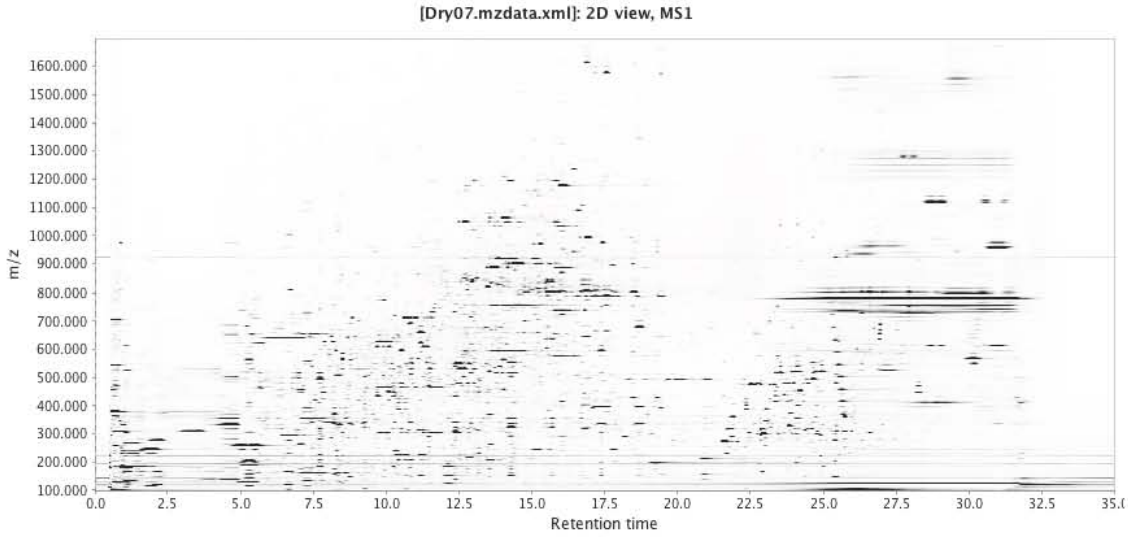


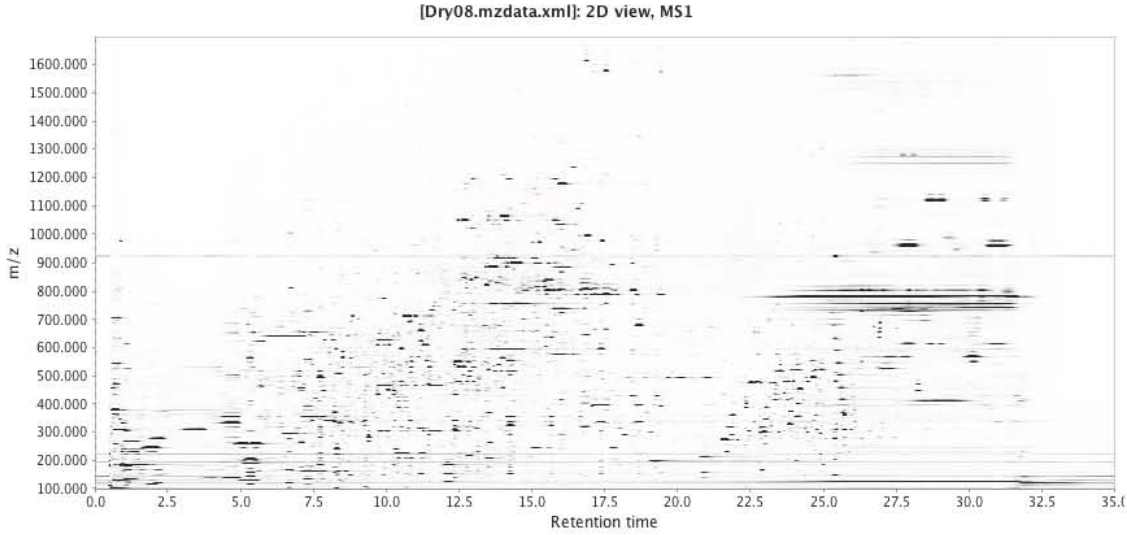


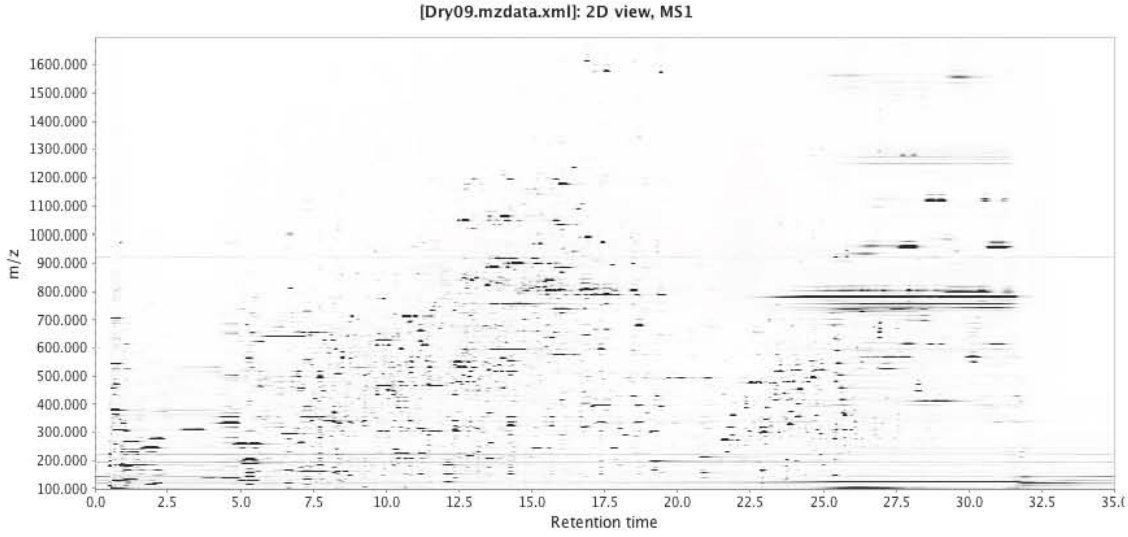

[Dry03.mzdata.xml]: 2D view, MS1
m/z
Retention time

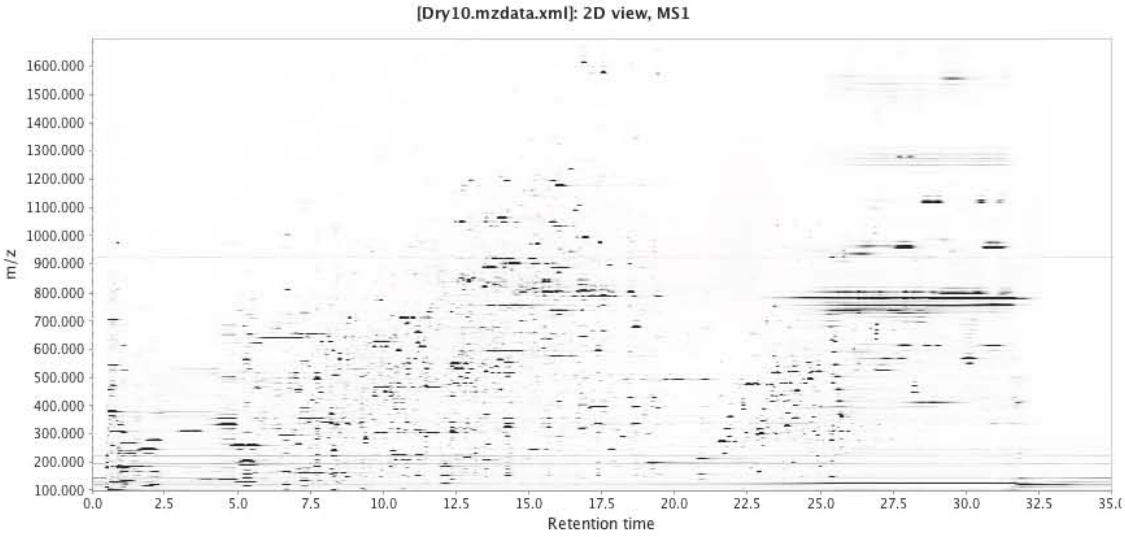

[Dry04.mzdata.xml]: 2D view, MS1
m/z
Retention time


[Dry05.mzdata.xml]: 2D view, MS1
m/z
Retention time

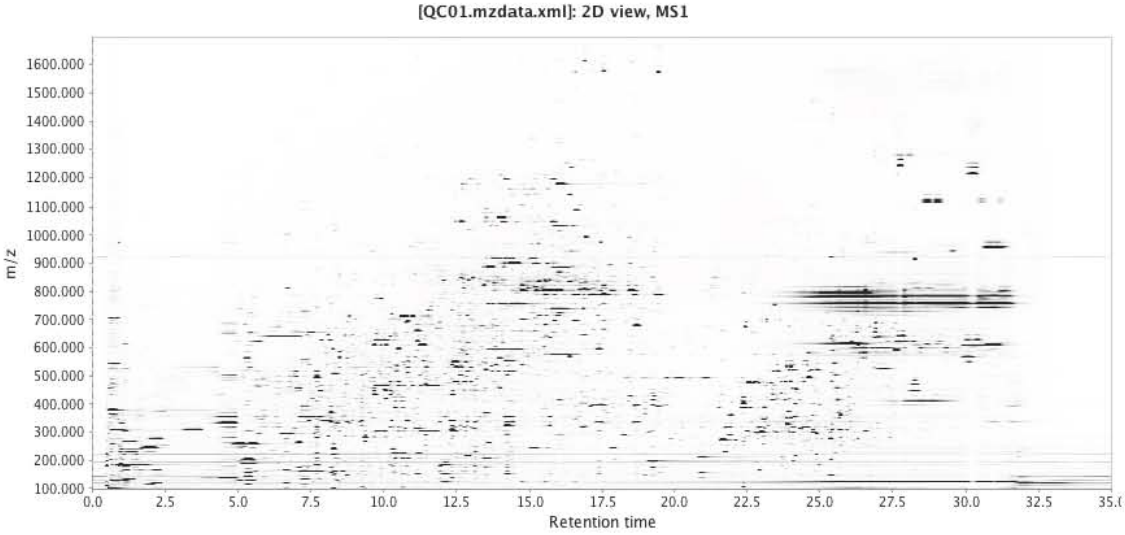

[Dry07.mzdata.xml]: 2D view, MS1
m/z
Retention time

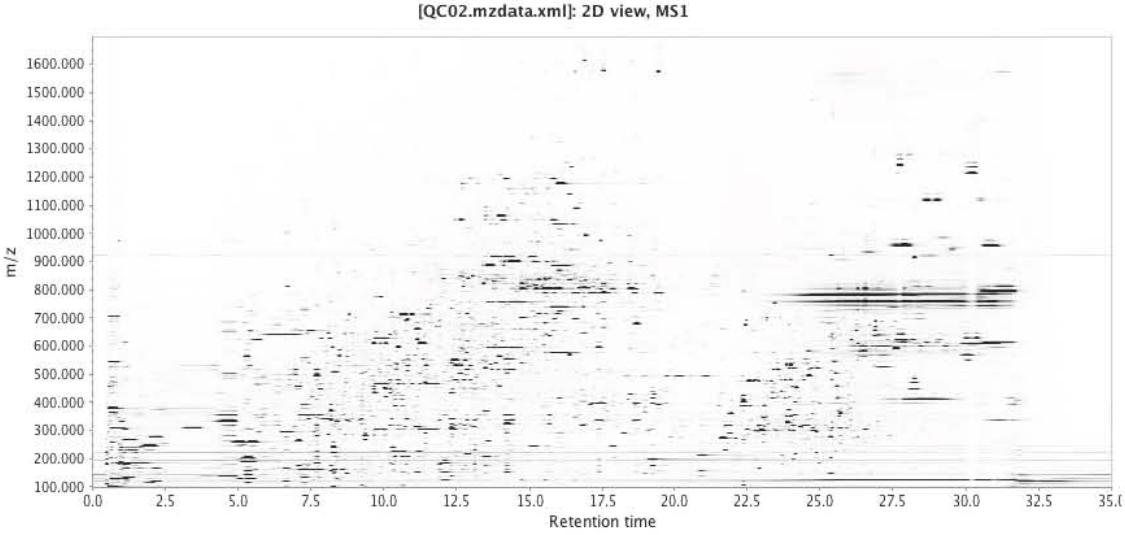

[Dry08.mzdata.xml]: 2D view, MS1
m/z
Retention time

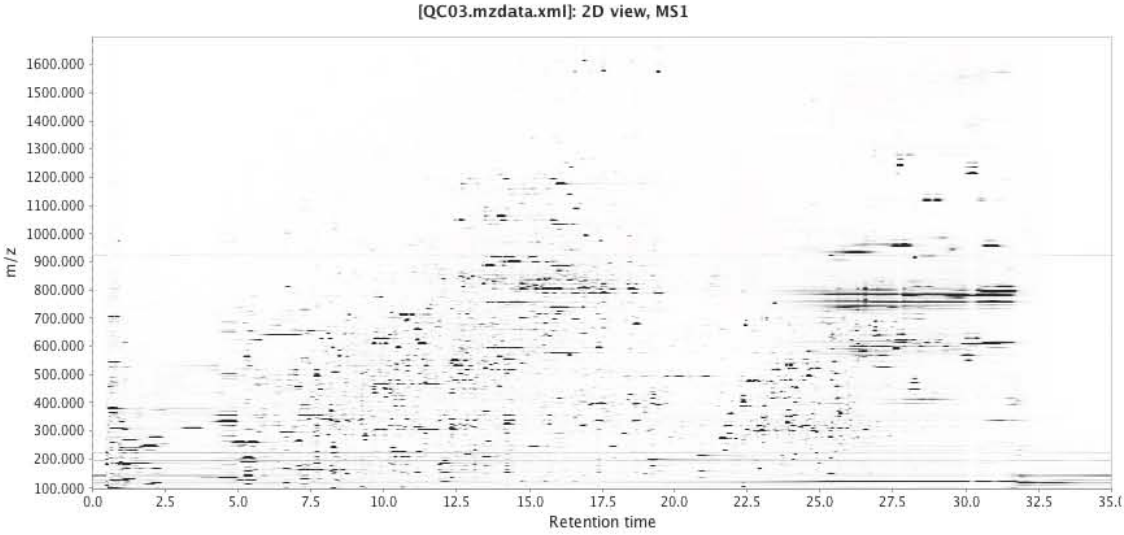

[Dry09.mzdata.xml]: 2D view, MS1
m/z
Retention time


[Dry10.mzdata.xml]: 2D view, MS1
m/z
1600.000
1500.000
1400.000
1300.000
1200.000
1100.000
1000.000
900.000
800.000
700.000
600.000
500.000
400.000
300.000
200.000
100.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
Retention time

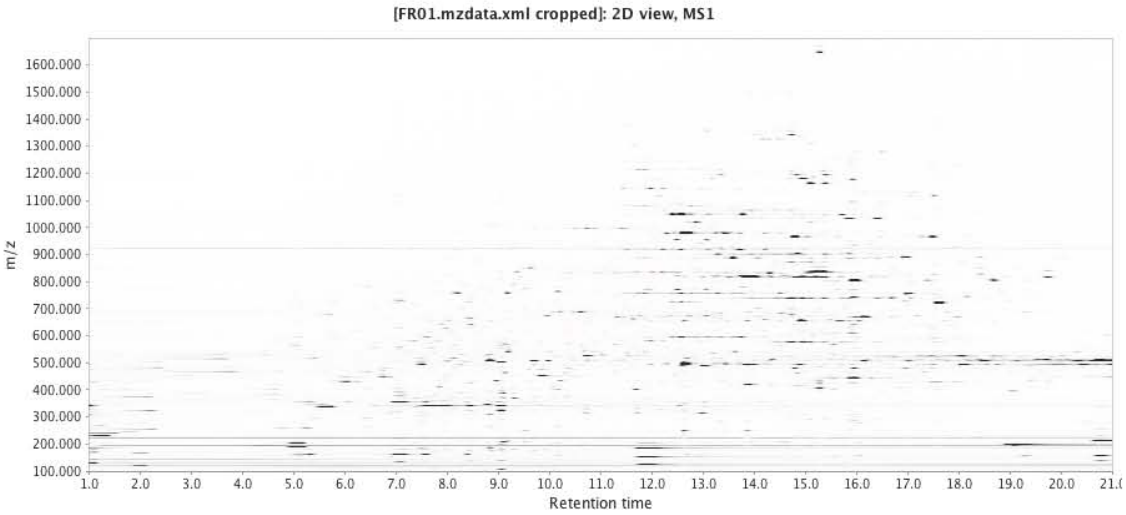
2D Chromatograms of QC Leaf Samples

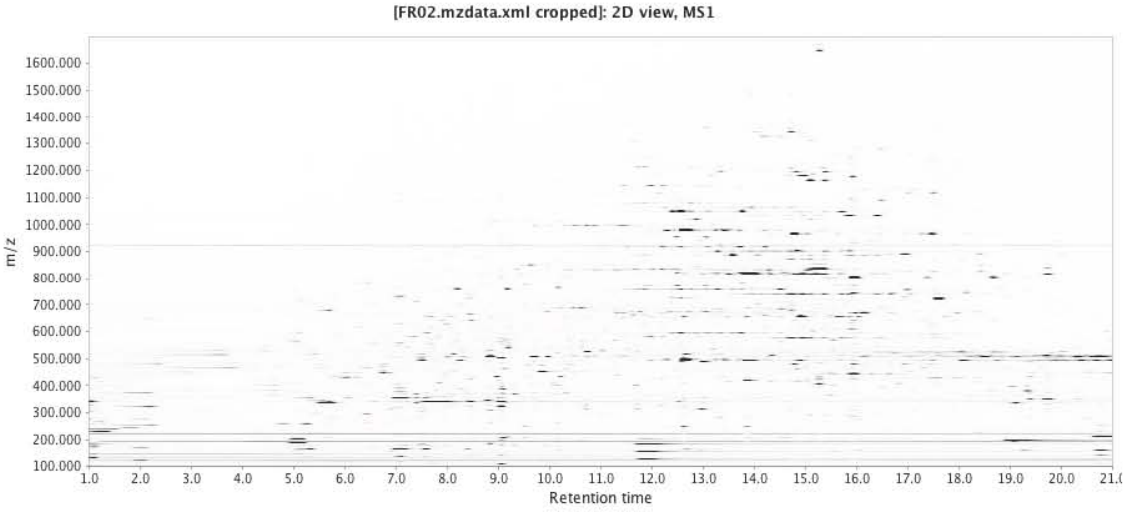


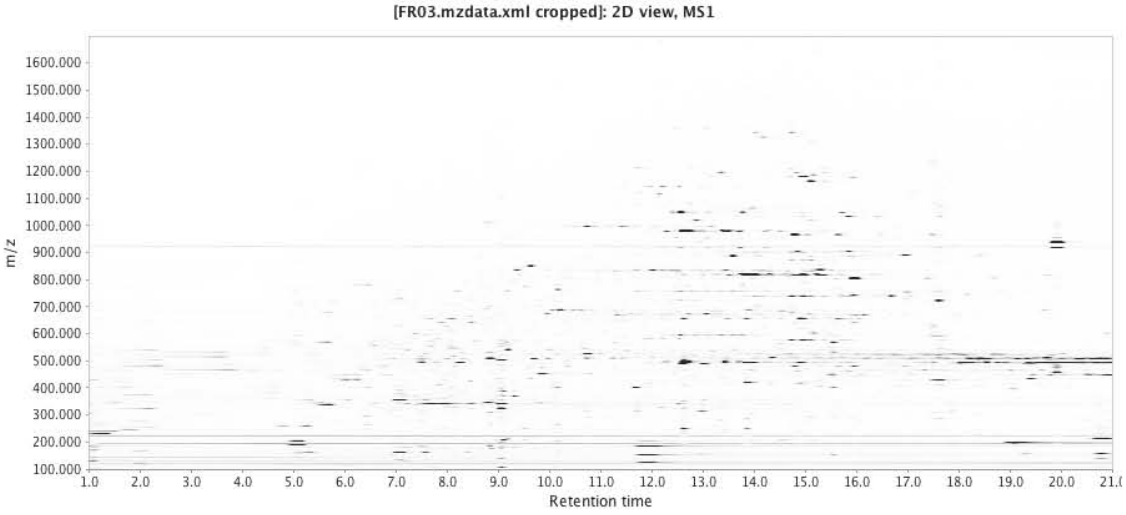


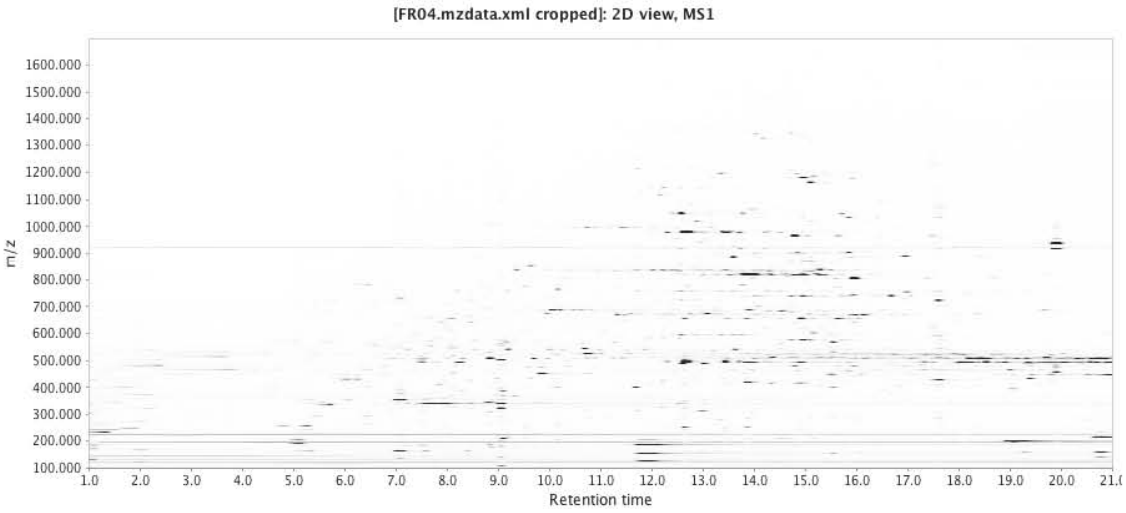


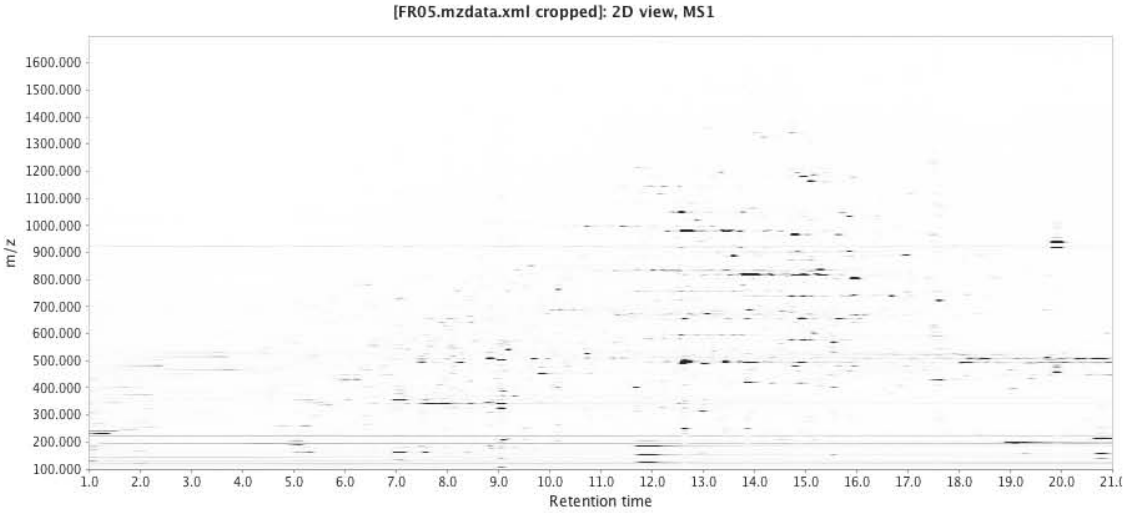
2D Chromatograms Of Hydrated Root Extracts

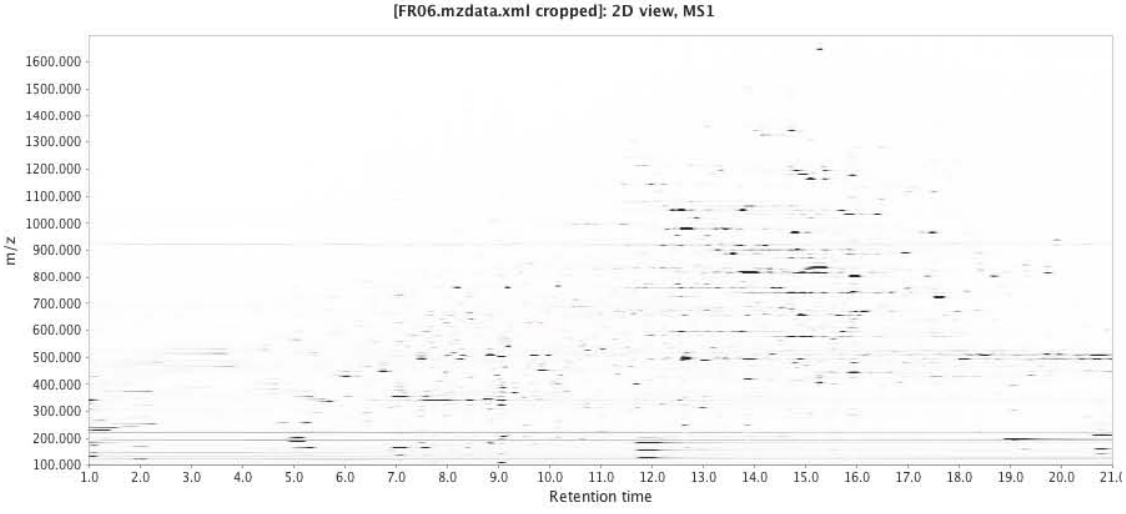


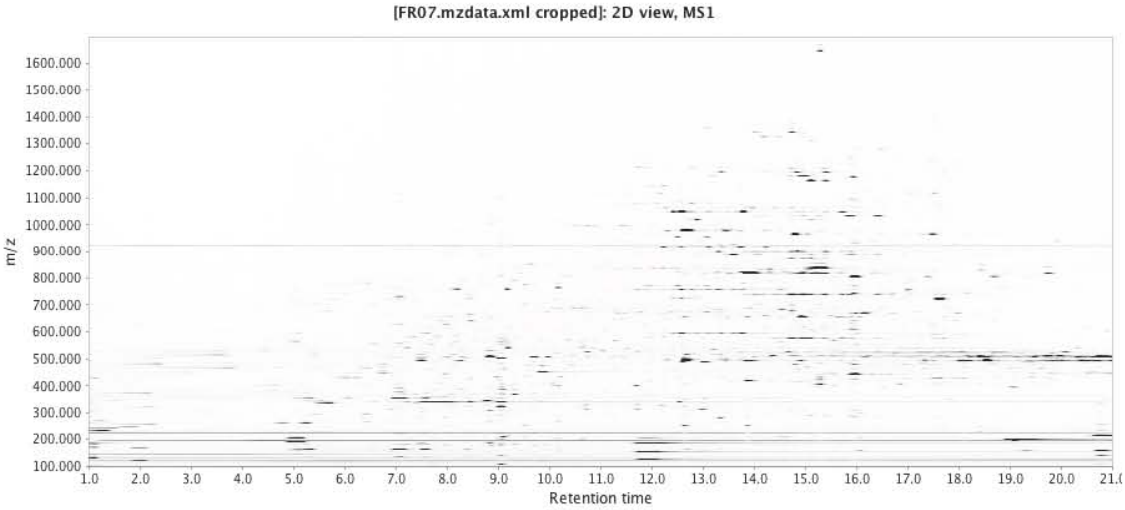


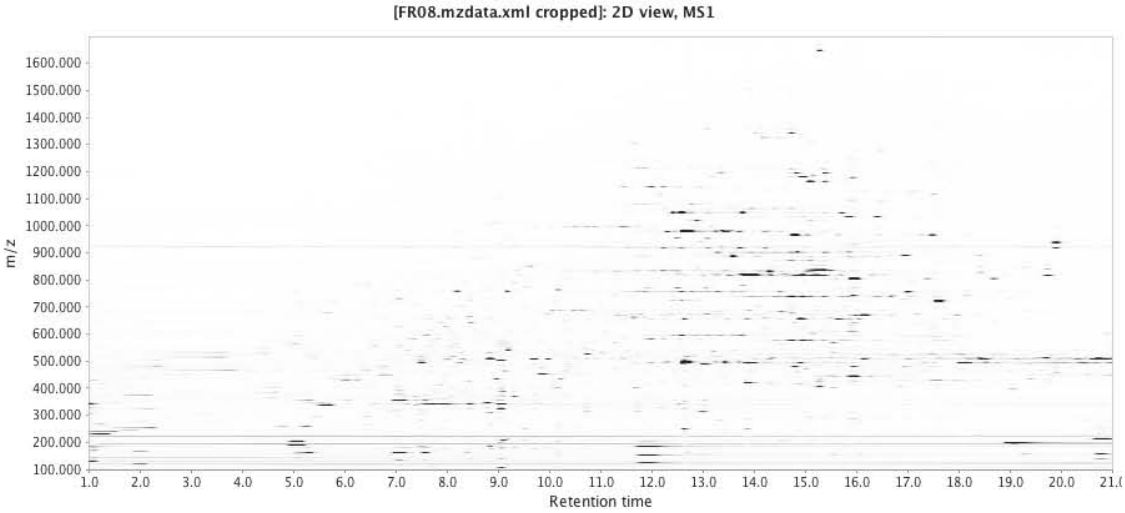


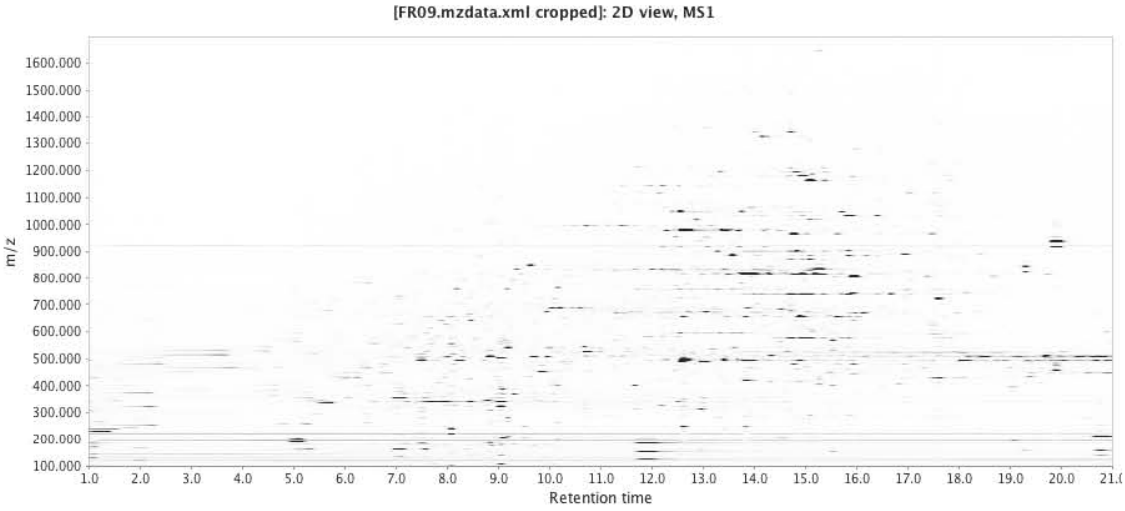

[FR04.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

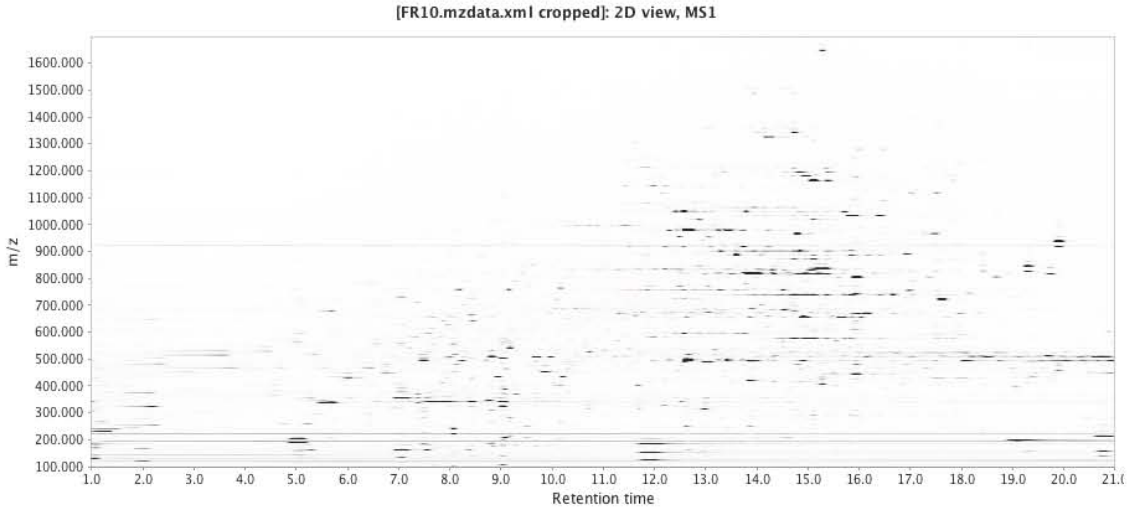

[FR05.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time


[FR06.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

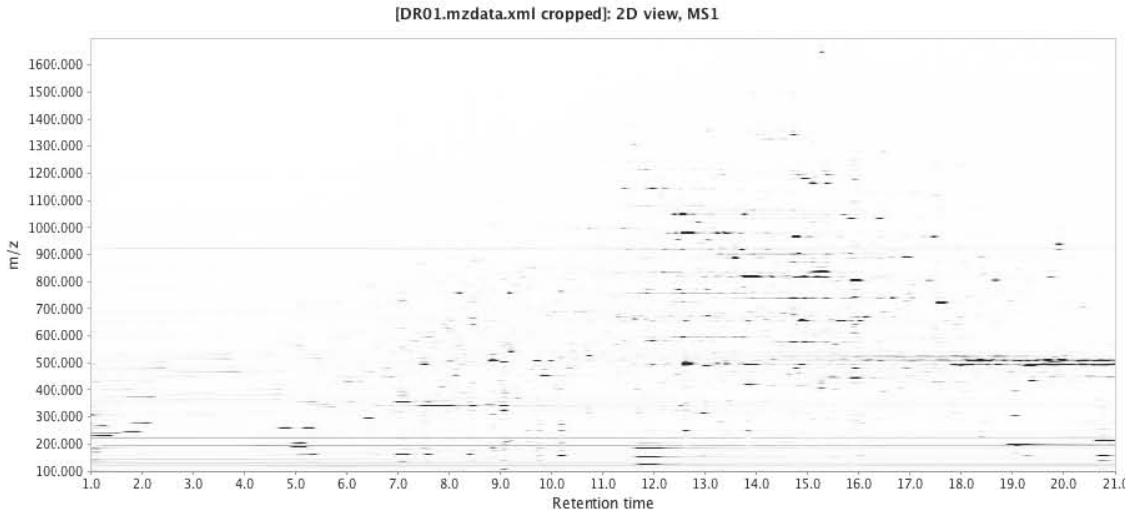

[FR07.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

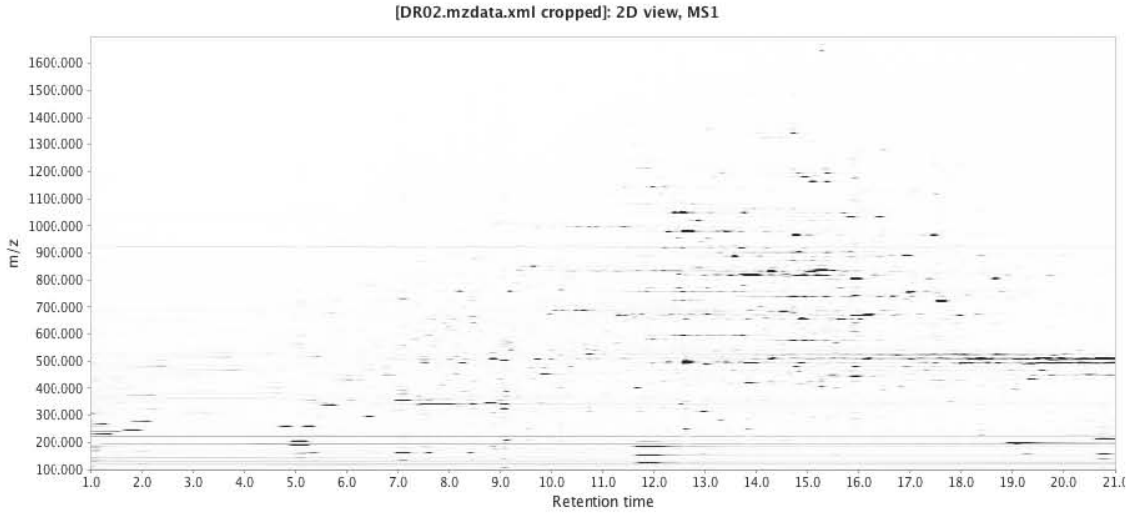

[FR08.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

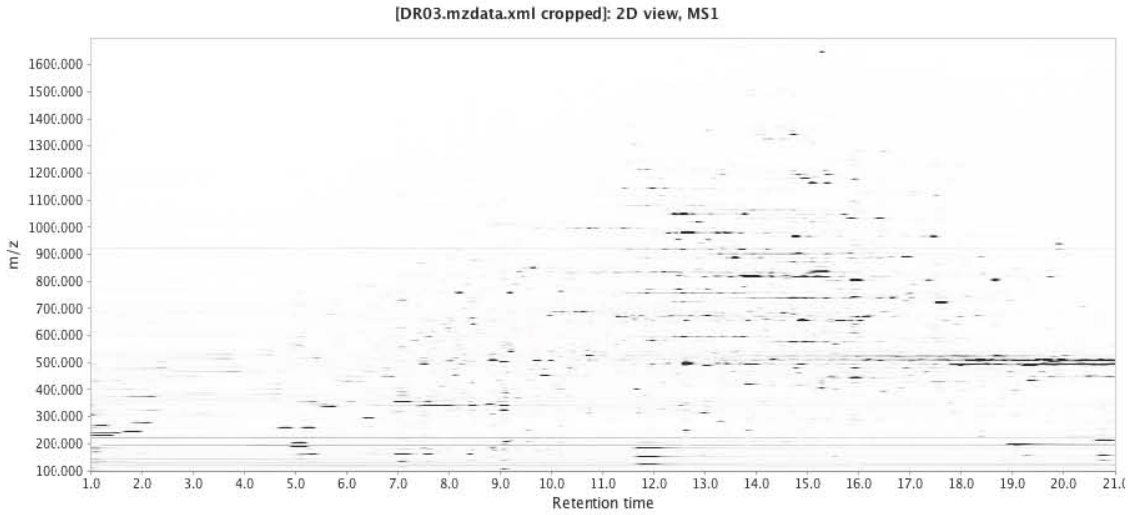

[FR09.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

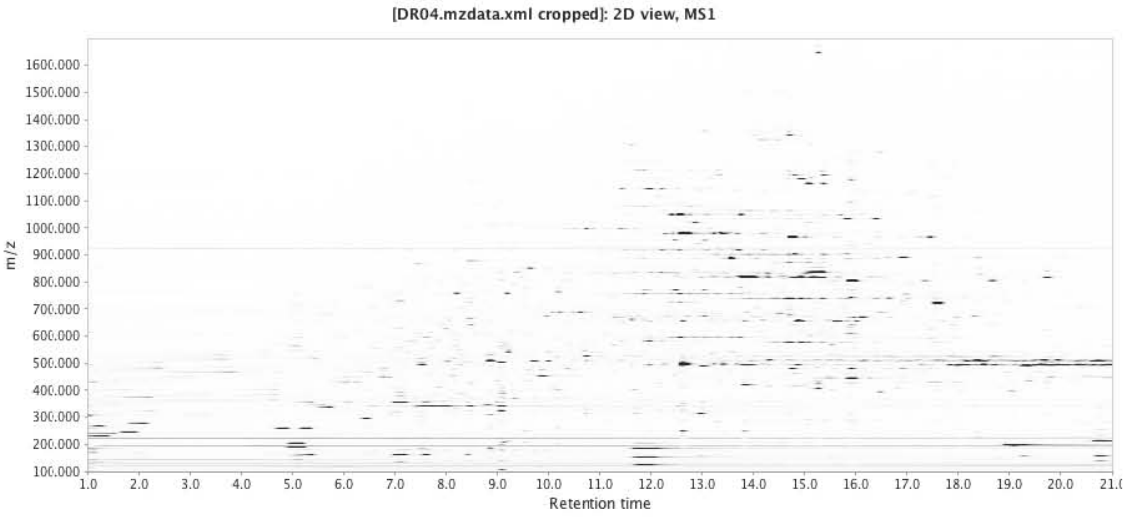

[FR10.mzdata.xml cropped]: 2D view, MS1
m/z
1600.000
1500.000
1400.000
1300.000
1200.000
1100.000
1000.000
900.000
800.000
700.000
600.000
500.000
400.000
300.000
200.000
100.000
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
Retention time

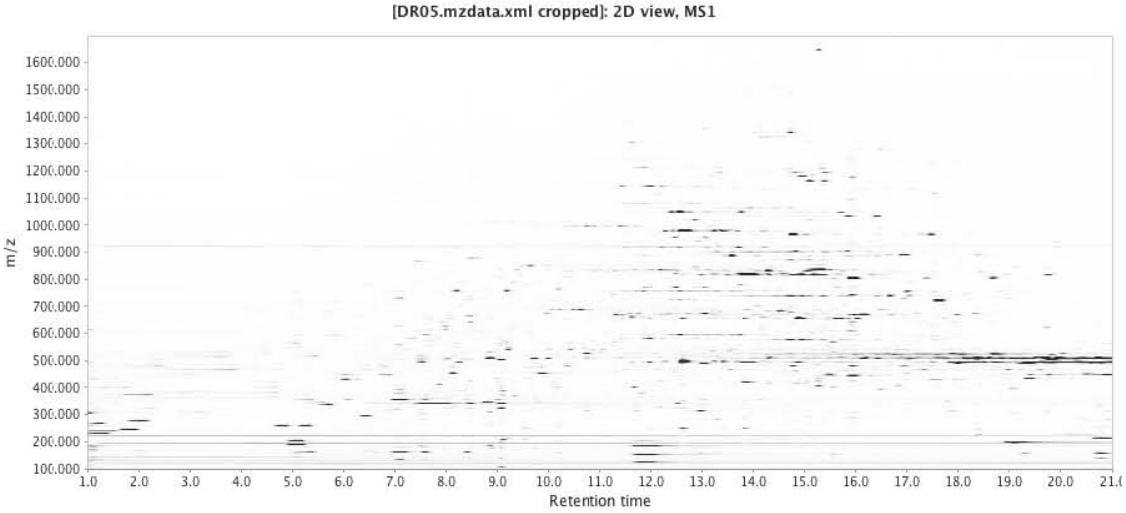
2D Chromatograms of Desiccated Root Extracts

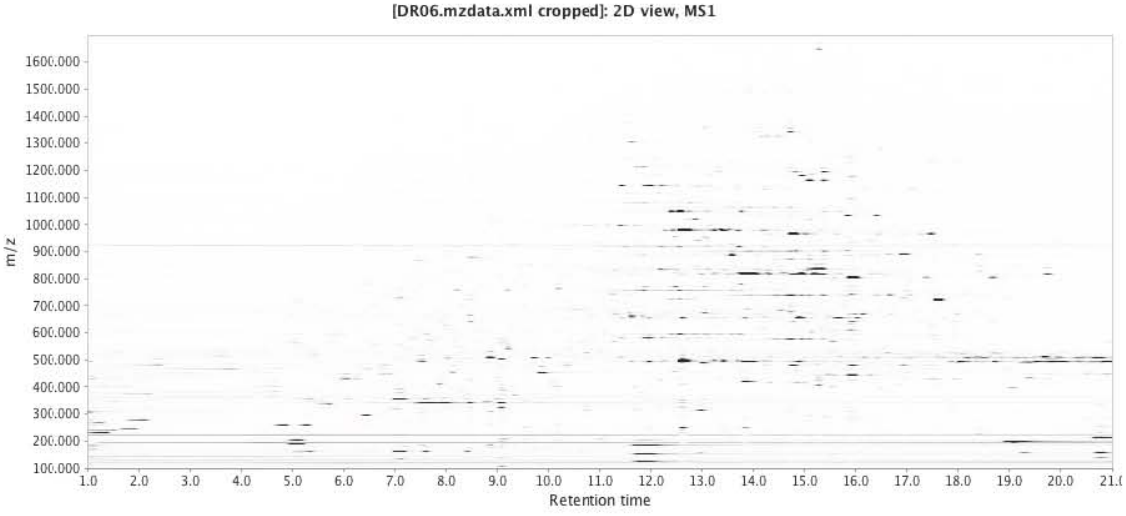


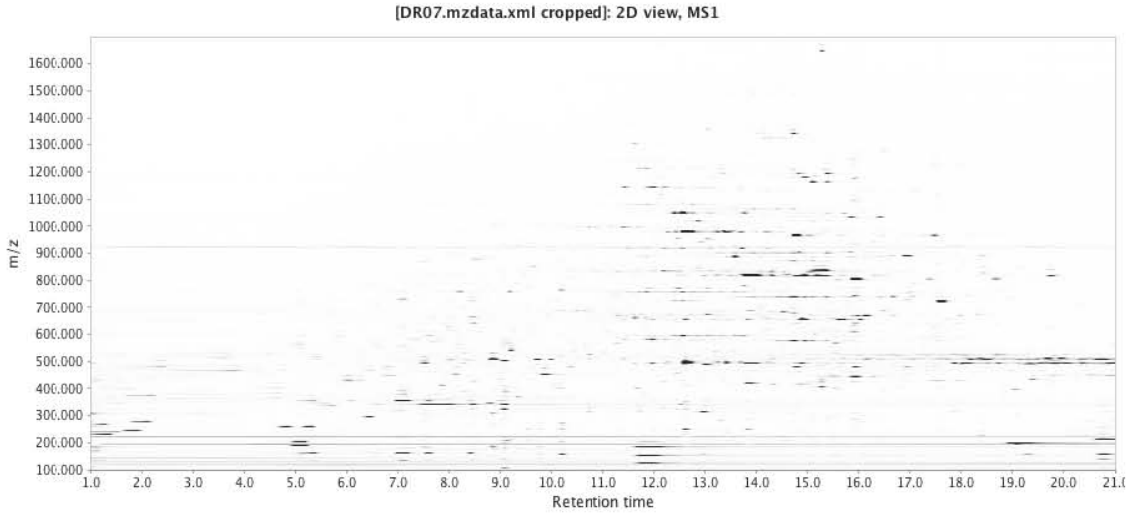


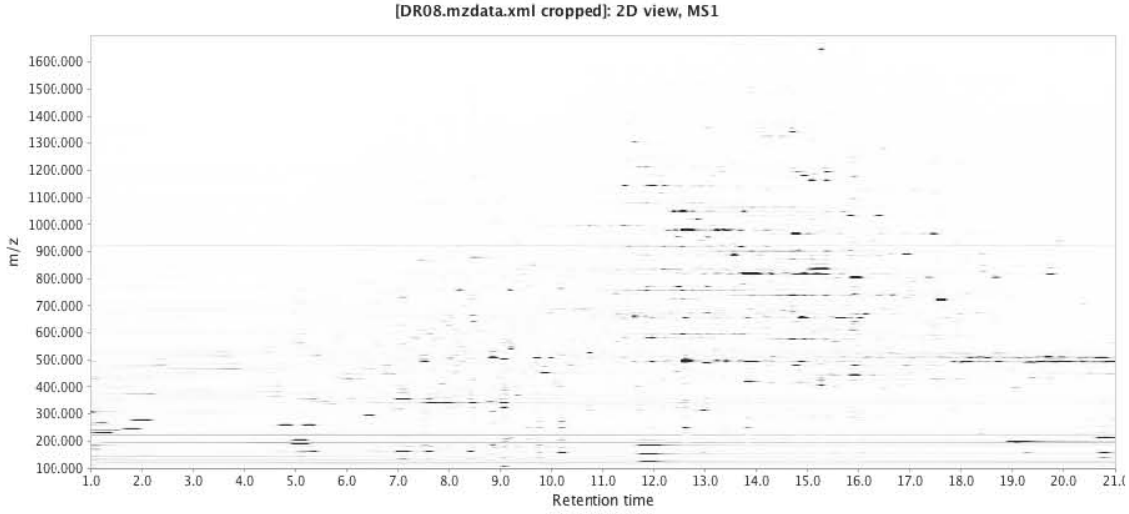


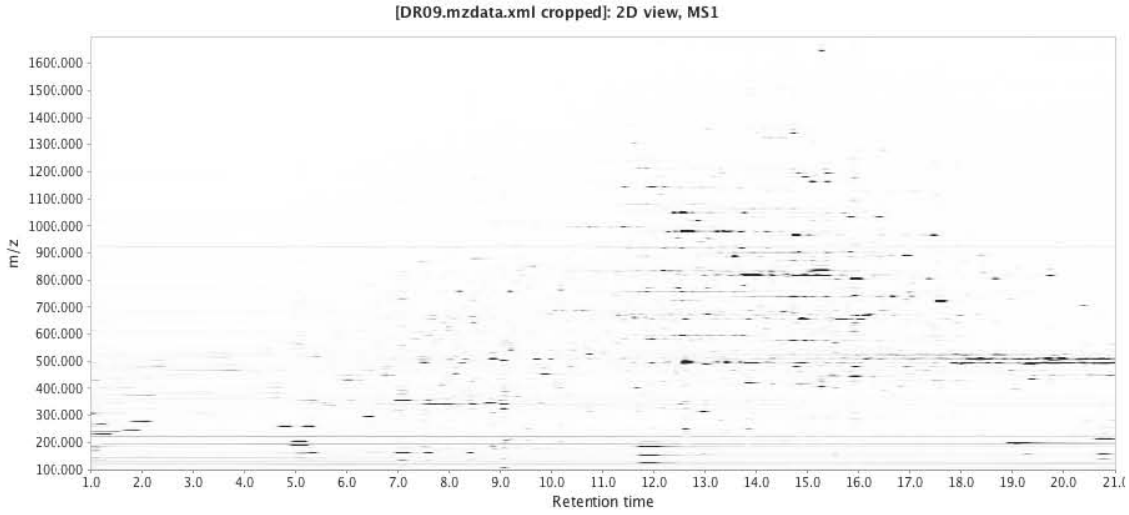

[DR04.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

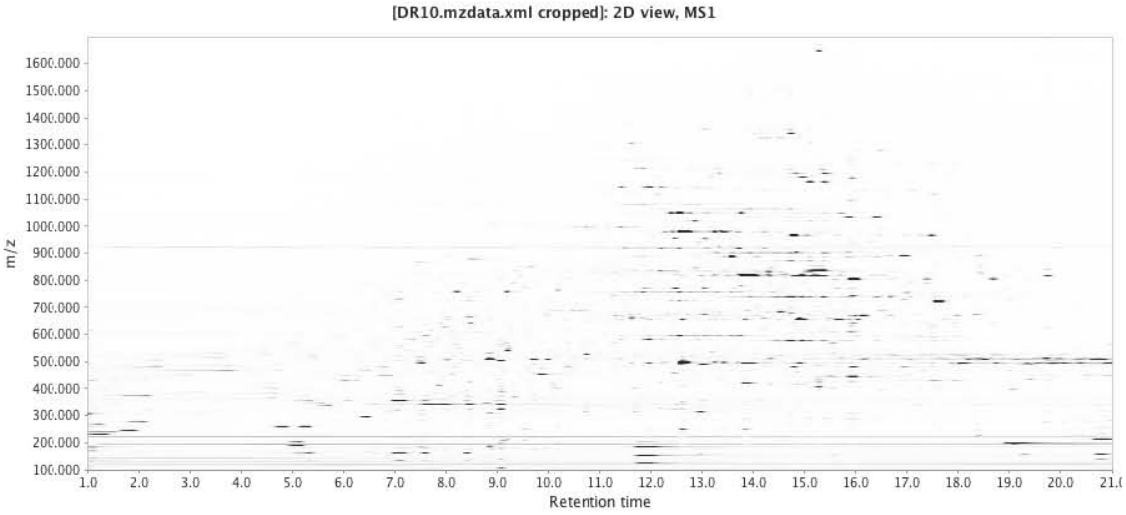

[DR05.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time


[DR06.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

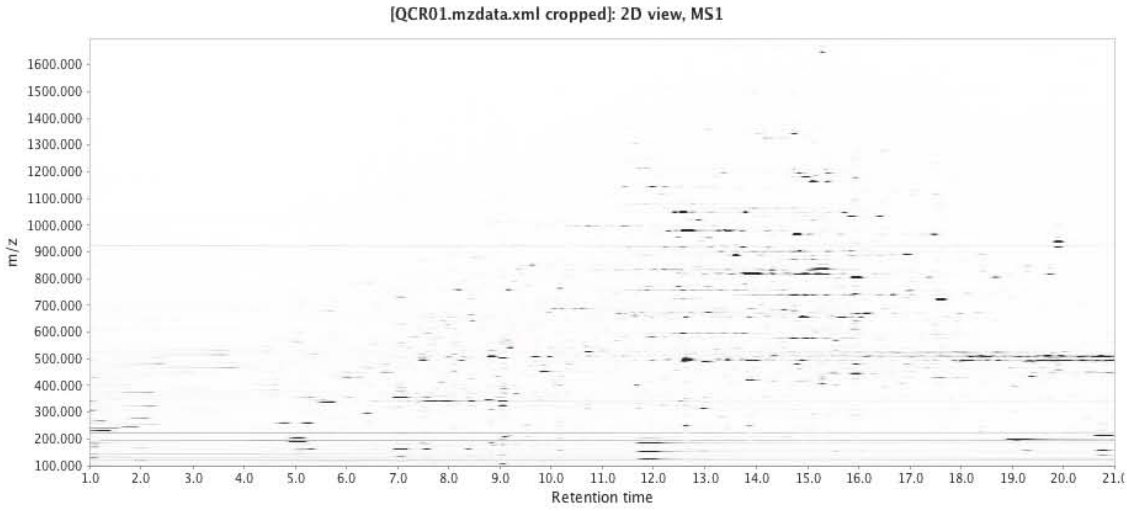

[DR07.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

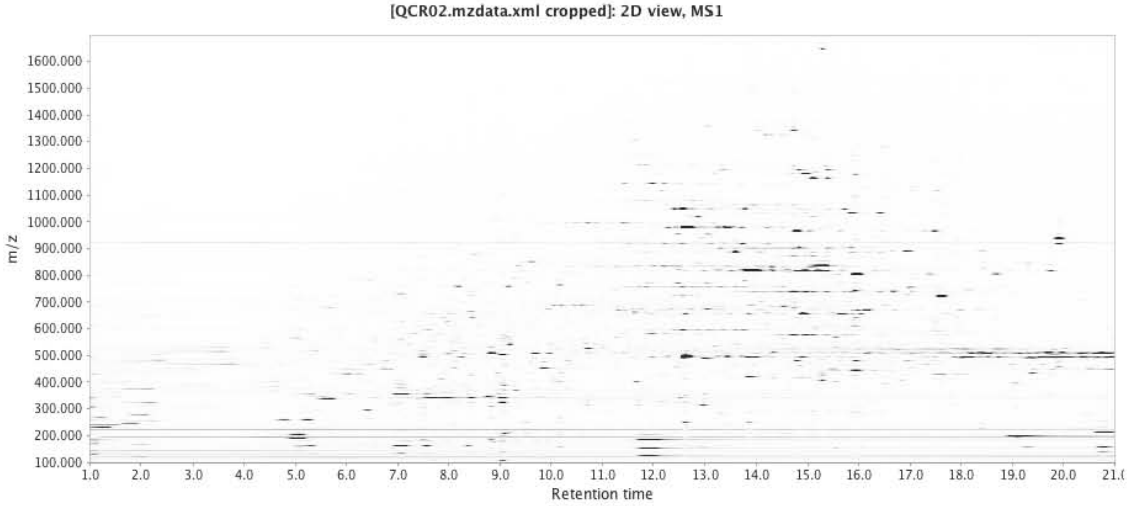

[DR08.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time

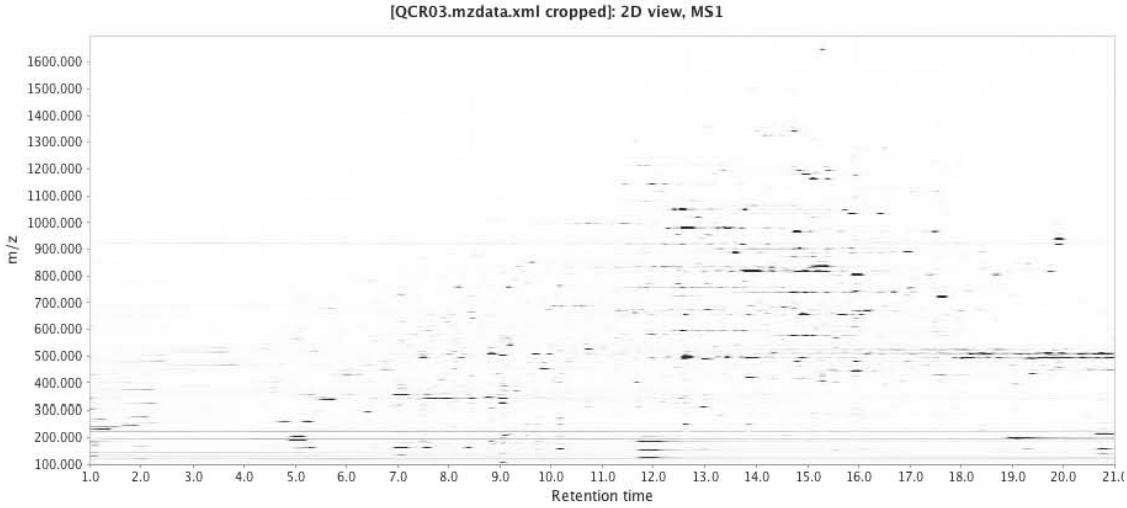

[DR09.mzdata.xml cropped]: 2D view, MS1
m/z
Retention time


[DR10.mzdata.xml cropped]: 2D view, MS1
m/z
1600.000
1500.000
1400.000
1300.000
1200.000
1100.000
1000.000
900.000
800.000
700.000
600.000
500.000
400.000
300.000
200.000
100.000
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
Retention time

2D Chromatograms of QC Root Samples





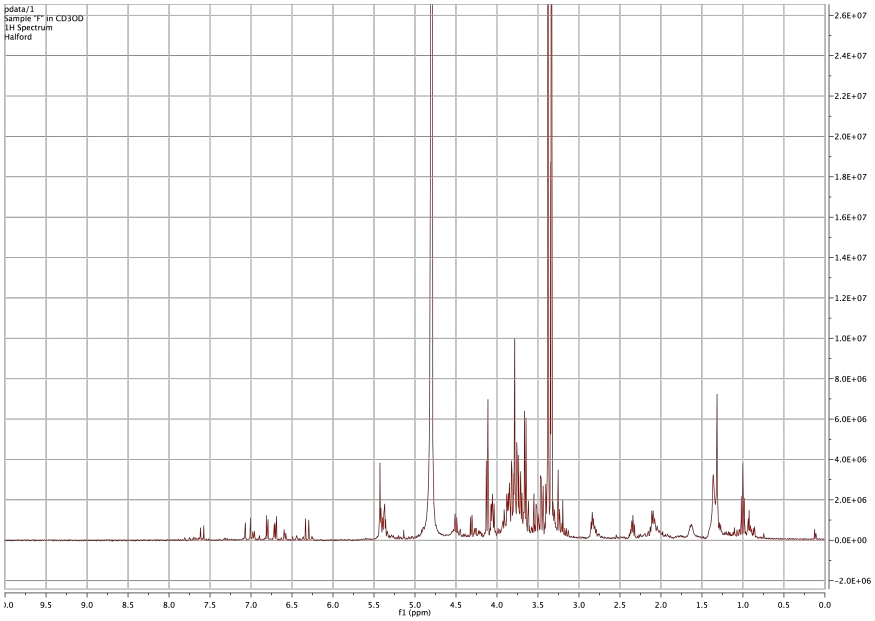


Appendix E: NMR Spectra

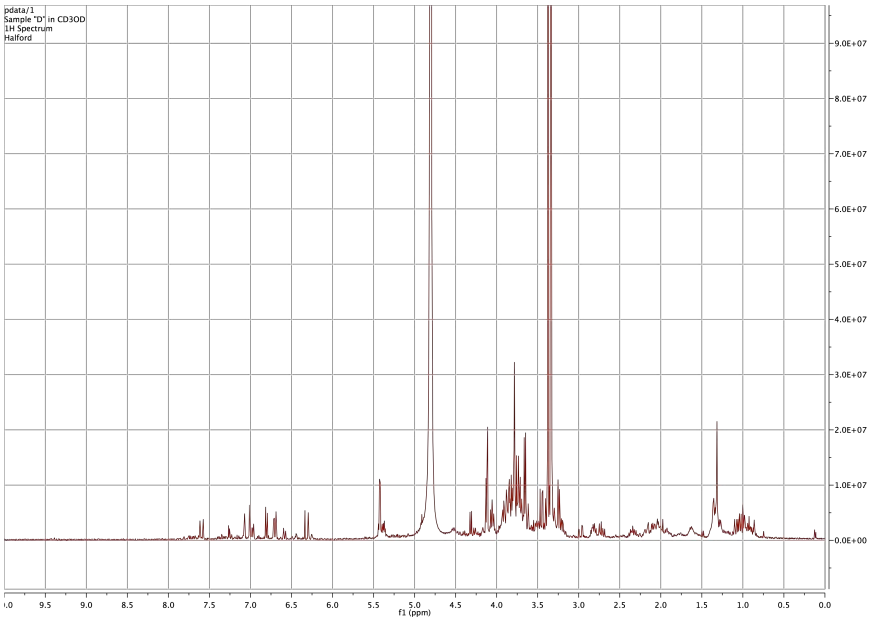
The following NMR spectra were used in confirming metabolite assignments. Spectrum images were generated in mNova 8.1.

1D-Proton Spectra

Hydrated Leaf

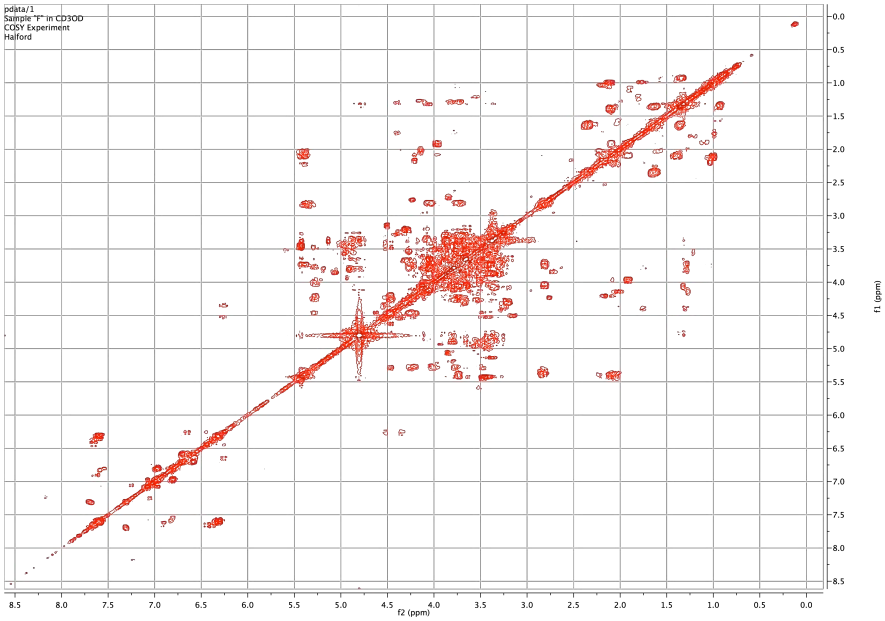


Desiccated Leaf

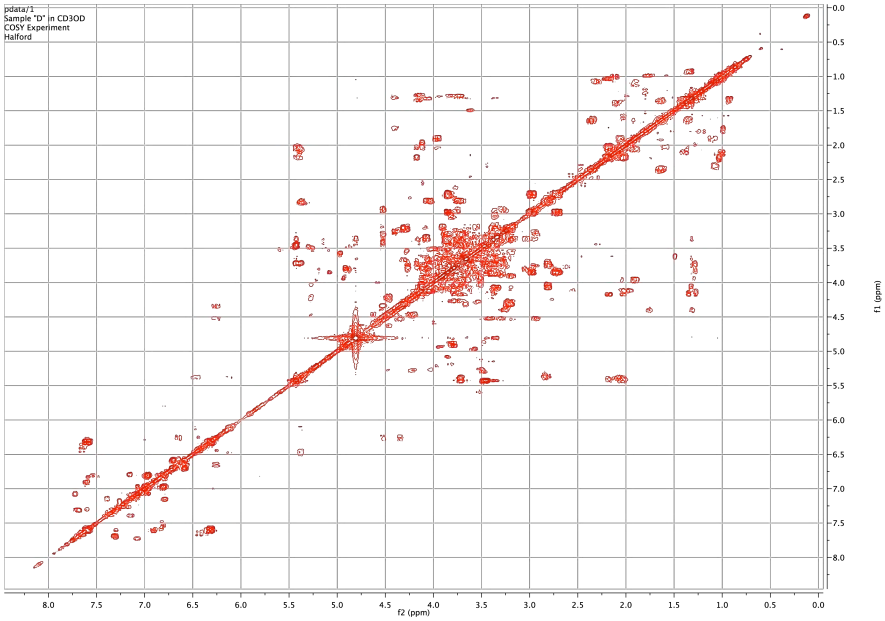


COSY Spectra

Hydrated Leaf

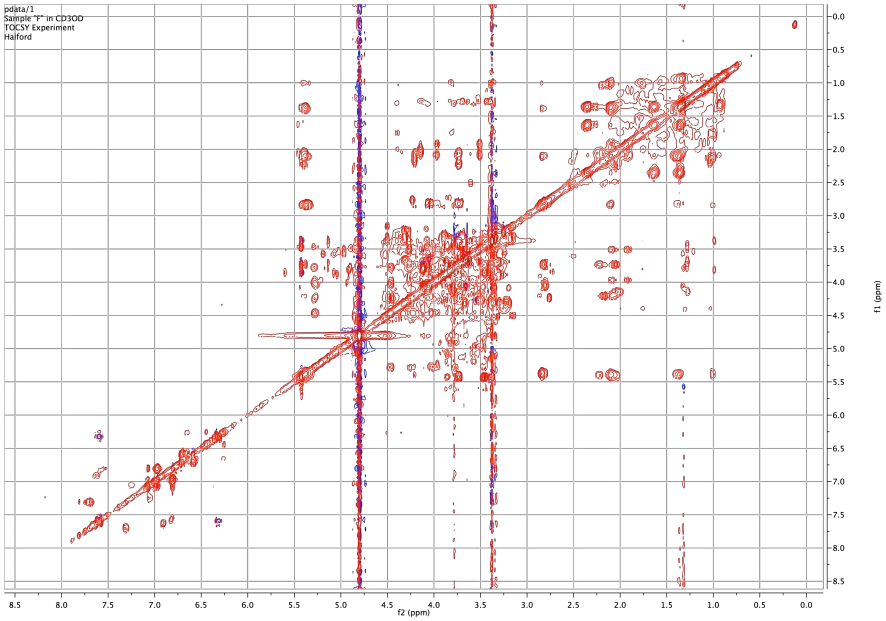


Desiccated Leaf

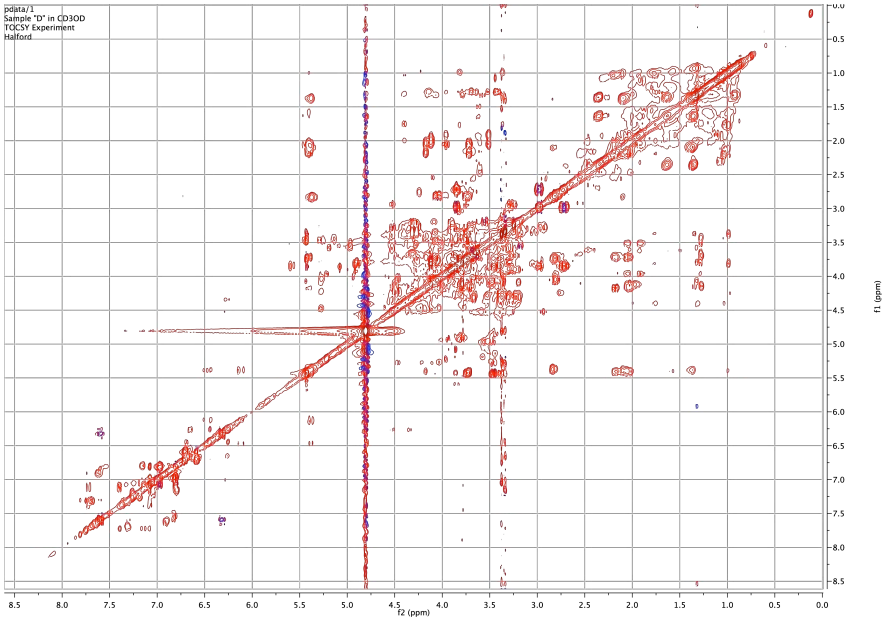


TOCSY Spectra

Hydrated Leaf

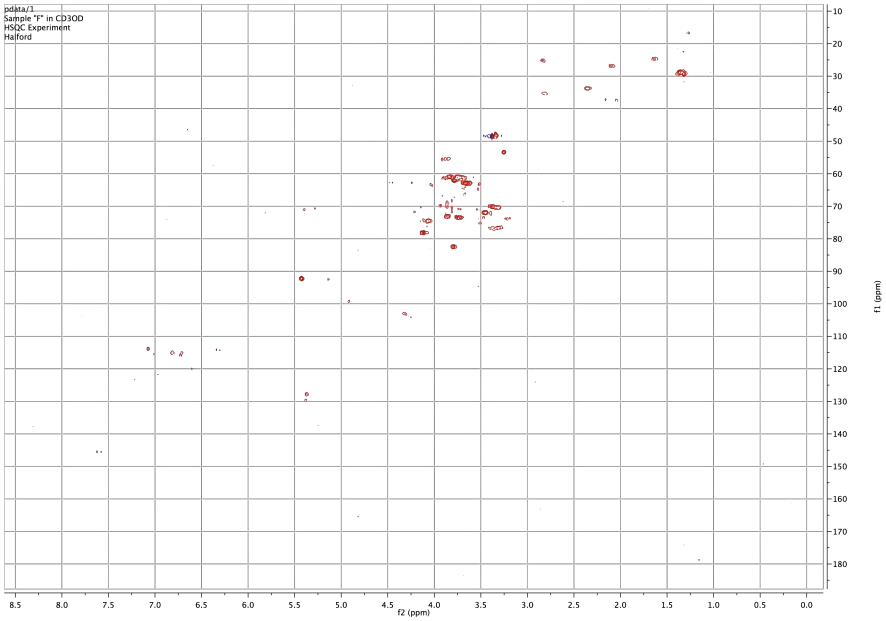


Desiccated Leaf

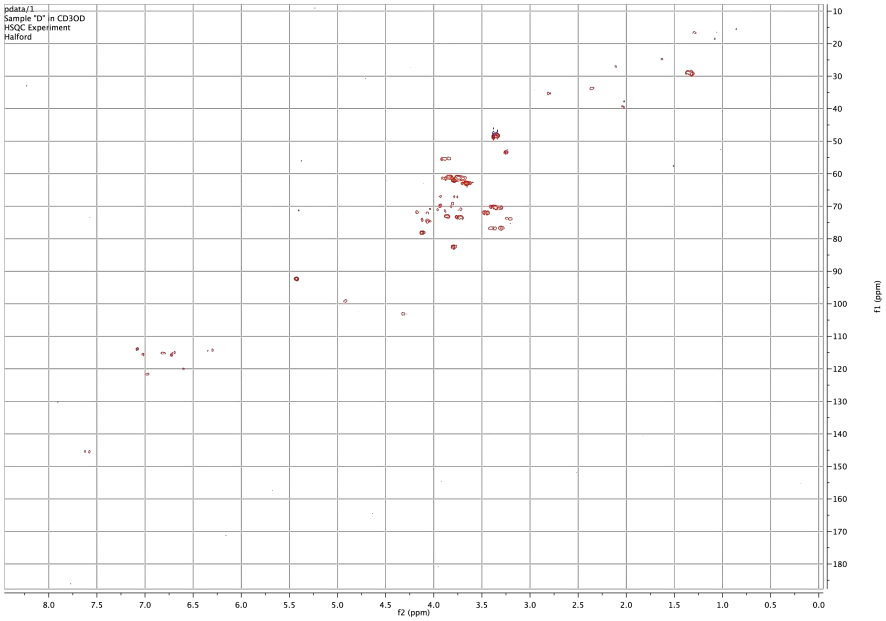


HSQC Spectra

Hydrated Leaf



Desiccated



Appendix F: Perl Code for Unit Mass Consolidation

```
#!/usr/bin/perl

my %spectrum = ();

while (<>) {
     @myitems = split /\t/;
     $myitems[0] =~ s/^\s+//;
     $myitems[0] =~ s/\s+$//;
     $myitems[1] =~ s/^\s+//;
     $myitems[1] =~ s/\s+$//;
     $mz = $myitems[0];
     $count = $myitems[1];
     if (defined $spectrum{$mz}) {
          $spectrum{$mz} = $spectrum{$mz} +  $count;
     } else {
          $spectrum{$mz} = $count;
     }
}
foreach $key (sort {$a <=> $b}  keys %spectrum) {
     print "$key ";
     print $spectrum{$key}."\n";
     #print "$key $spectrum{$key} ";
}
```

Appendix G: R Code for PCA

```
library("ChemometricsWithR")
setwd("~/Dropbox/MSc_Data/MS_Stats")
msdata <- read.table('lc_root.txt', header = TRUE, row.names =
1, sep = "\t")
mslevels_lc_root <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 3, 3, 3)
mslevels_lc_leaf <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2,
2, 2, 2, 2, 2, 3, 3, 3)
mslevels_gc_root <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 2)
mslevels_gc_leaf <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 2)
mslevels <- mslevels_lc_root
msdata[is.na(msdata)] <- 100

# Mean center dataset
msdata.mc <- scale(msdata, center=TRUE, scale=FALSE)
# Pareto scale
msdata.sc <- scale(msdata.mc, center=FALSE, scale =
sqrt(apply(msdata.mc, 2, sd)))

msdata.PCA <- PCA(msdata.sc)

scoreplot(msdata.PCA, pc=c(1, 4), pch =
c(16,17,18)[(mslevels)], col=c("red", "green",
"blue")[(mslevels)], )
biplot(msdata.PCA,pc=c(1,2), score.col=c("red", "green",
"blue")[(mslevels)], show.names=c("none"))
loadingplot(msdata.PCA, pc=c(1,4), show.names=TRUE, min.length
= 0.009)
screeplot(msdata.PCA)
screeplot(msdata.PCA, type = "percentage", main = "Cumulative
Variance", cex.names = 0.65)

vars <- variances(msdata.PCA)
```

```
relvars <- (vars / sum(vars))
barplot(100 * relvars[1:10], cex.names = 0.75,
        main = "Relative PC Variance",
        ylab = "Relative Variance %",
        xlab = "Principal Component",
        ylim = c(0,30))

msdata.mc2 <- sweep(msdata, 2, colMeans(msdata))
pcobj_pareto <- prcomp(msdata.mc2, center=FALSE,
scale=sqrt(apply(msdata.mc2, 2, sd)))
pcobj_noscale = prcomp(msdata.mc2, scale=FALSE)
pcobj_autoscale = prcomp(msdata.mc2, scale=TRUE)
print(pcobj_pareto)
plot(pcobj_pareto) # eigenvalues
biplot(pcobj_pareto, choices=c(1,2), arrow.len=0, cex=0.5,
col=c("black", "steelblue"))
biplot(pcobj_noscale, choices=c(1,2), arrow.len=0, cex=0.5,
col=c("black", "steelblue"))
biplot(pcobj_autoscale, choices=c(1,2), arrow.len=0, cex=0.5,
col=c("black", "steelblue"))
```

Appendix H: R Code for PLS-DA

```
library("mixOmics")
setwd("~/Dropbox/MSc_Data/MS_Stats")

# Select experiment and tissue before running
experiment <- "gc" # "gc" or "lc"
tissue <- "root" # "leaf" or "root"

filename <- paste0(experiment, "_", tissue, ".txt")
# Metadata file contains RT and BP mass values for assembling
final presentation table
filename_metadata <- paste0(experiment, "_", tissue,
"_metadata.txt")

msdata <- read.table(filename, header = TRUE, row.names = 1,
sep = "\t")
metadata <- read.table(filename_metadata, header = TRUE,
row.names = 1, sep = "\t")
row.names(metadata) <- sapply(row.names(metadata),
function(x){paste0("X",x)})

# Set up parameters for each experiment - the treatments per
sample
# and how many variables to include in the model for each
component
components <- 3
varsPerComp <- 50
if (experiment == "gc") {
  mslevels <- c("DRY", "DRY", "DRY", "DRY", "DRY", "DRY",
"DRY", "DRY", "DRY", "DRY", "HYDRATED", "HYDRATED",
"HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED",
"HYDRATED", "HYDRATED", "HYDRATED")
  if (tissue == "leaf") {
  } else if (tissue == "root") {
    varsPerComp <- 10
  }
} else if (experiment == "lc") {
  if (tissue == "leaf") {
```

```
    msdata <- msdata[1:18,]
    mslevels <- c("DRY", "DRY", "DRY", "DRY", "DRY", "DRY",
"DRY", "DRY", "DRY", "HYDRATED", "HYDRATED", "HYDRATED",
"HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED",
"HYDRATED")
    varsPerComp <- 200
  } else if (tissue == "root") {
    msdata <- msdata[1:20,]
    mslevels <- c("DRY", "DRY", "DRY", "DRY", "DRY", "DRY",
"DRY", "DRY", "DRY", "DRY", "HYDRATED", "HYDRATED",
"HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED",
"HYDRATED", "HYDRATED", "HYDRATED", "HYDRATED")
    varsPerComp <- 50
  }
}

mslevels <- as.factor(mslevels)

#Plot colours

col.mslevels <- as.numeric(mslevels)
col.mslevels[col.mslevels == 1] = 'red'
col.mslevels[col.mslevels == 2] = 'blue'
col.mslevels[col.mslevels == 3] = 'green'

msdata[is.na(msdata)] <- 100

# Mean centre
msdata.mc <- scale(msdata, center=TRUE, scale=FALSE)
# Pareto scale
msdata.sc <- scale(msdata.mc, center=FALSE, scale =
sqrt(apply(msdata.mc, 2, sd)))

# Construct a vector of the number of variables to select per
component

selectedVars <- rep(varsPerComp, components)

# Main PLS-DA calculation
```

```
result <- splsda(msdata.sc, mslevels, ncomp = components,
keepX = selectedVars)

# Pause between plots
par(ask = TRUE)

# Component Plot
plotIndiv(result, comp=1:2, ind.names = FALSE, col =
col.mslevels, pch = 16, X.label="Comp 1", Y.label="Comp 2")



# Legend
col.legend <- unique(col.mslevels)
legend(-0.23, 0.50, c("Desiccated", "Hydrated"), col =
col.legend, pt.cex = 1, pch = 16, title = "Plant Status")

#plot3dIndiv(result, comp=1:3, col=col.mslevels, cex=0.25,
axes.box="both")

varCoords <- plotVar(result, comp=1:2,
var.label=names(msdata), cex=0.5, X.label="Comp 1",
Y.label="Comp 2")

# Some really bad R because I need this fast and can't figure
out the vectorised approach

dry_peaks <-
sort(names(varCoords$coord.X[(varCoords$coord.X[,1] >
0.5),1]))
dry_pval <- numeric(length(dry_peaks))
dry_folds <- numeric(length(dry_peaks))
dry_rt <- numeric(length(dry_peaks))
dry_mz <- numeric(length(dry_peaks))
dry_peakcount <- numeric(length(dry_peaks))
wet_peaks <-
sort(names(varCoords$coord.X[(varCoords$coord.X[,1] < -
0.5),1]))
wet_pval <- numeric(length(wet_peaks))
```

```
wet_folds <- numeric(length(wet_peaks))
wet_rt <- numeric(length(wet_peaks))
wet_mz <- numeric(length(wet_peaks))
wet_peakcount <- numeric(length(wet_peaks))

for (i in 1:length(dry_peaks)) {
  v <- dry_peaks[i]
  dry_pval[i] <- t.test(msdata[[v]][mslevels ==
"DRY"],msdata[[v]][mslevels == "HYDRATED"])$p.value
  dry_folds[i] <- mean(msdata[[v]][mslevels ==
"DRY"])/mean(msdata[[v]][mslevels == "HYDRATED"])
  dry_rt[i]
  dry_mz[i]
  dry_peakcount[i]
}

for (i in 1:length(wet_peaks)) {
  v <- wet_peaks[i]
  wet_pval[i] <- t.test(msdata[[v]][mslevels ==
"DRY"],msdata[[v]][mslevels == "HYDRATED"])$p.value
  wet_folds[i] <- mean(msdata[[v]][mslevels ==
"HYDRATED"])/mean(msdata[[v]][mslevels == "DRY"])
  wet_rt[i]
  wet_mz[i]
  wet_peakcount[i]
}

dry_vals <- data.frame(peaks = dry_peaks, pval = dry_pval,
fold = dry_folds)
row.names(dry_vals) <- dry_peaks
dry_sig <- dry_vals[dry_vals$pval < 0.05,]
dry_sig$rt <-
metadata[row.names(dry_sig),"row.retention.time"]
dry_sig$mz <- metadata[row.names(dry_sig), "row.m.z"]
dry_sig$foundby <- "PLS-DA"

wet_vals <- data.frame(peaks = wet_peaks, pval = wet_pval,
fold = wet_folds)
row.names(wet_vals) <- wet_peaks
```

```
wet_sig <- wet_vals[wet_vals$pval < 0.05,]
wet_sig$rt <-
metadata[row.names(wet_sig),"row.retention.time"]
wet_sig$mz <- metadata[row.names(wet_sig), "row.m.z"]
wet_sig$foundby <- "PLS-DA"
```

Appendix I: R Code for Pathway Mapping

```
library(PAPi)
library(svDialogs)

metabolites <- read.table("metabolites.txt", header = TRUE,
sep = "\t", colClasses = "character")
metabolites[metabolites == ""] <- NA
keggcodes <- read.table("keggcodes.txt", header=TRUE,
sep="\t", colClasses = "character")

coded <- addKeggCodes(
  metabolites,
  keggcodes,
  save = FALSE,
  addCodes = TRUE
)

pathways <- papi(coded, save = FALSE, offline = TRUE,
localDatabase = "default")

irrelevancies <- c("Bacterial chemotaxis", "Bile secretion",
                   "Butirosin and neomycin biosynthesis",
                   "Carbohydrate digestion and absorption",
                   "Chlorocyclohexane and chlorobenzene
degradation",
                   "African trypanosomiasis", "GABAergic
synapse",
                   "Insulin secretion", "Insulin signaling
pathway",
                   "Metabolic pathways", "Methane metabolism",
"Mineral absorption",
                   "Microbial metabolism in diverse
environments",
                   "Primary bile acid biosynthesis",
"Streptomycin biosynthesis",
                   "Sphingolipid metabolism", "Serotonergic
synapse", "Taurine and hypotaurine metabolism",
```

```
                    "Retrograde endocannabinoid signaling",
"Proximal tubule bicarbonate reclamation",
                    "Synaptic vesicle cycle", "Taste
transduction", "Toluene degradation",
                    "Two-component system", "Type II diabetes
mellitus", "Circadian entrainment",
                    "Gap junction", "Alcoholism", "Amphetamine
addiction", "Amyotrophic lateral sclerosis (ALS)",
                    "Long-term depression", "Cocaine
addiction", "Glutamatergic synapse",
                    "Long-term potentiation", "Nicotine
addiction", "Meiosis - yeast",
                    "Huntington's disease", "Penicillin and
cephalosporin biosynthesis",
                    "Degradation of aromatic compounds",
"Protein digestion and absorption")

filteredPathways1 <- pathways[!(pathways$pathwayname %in%
irrelevancies),]

excesses <- c("Purine metabolism", "Lysine degradation",
"Thiamine metabolism")

filteredPathways <-
filteredPathways1[!(filteredPathways1$pathwayname %in%
excesses),]

pvals <- papiHtest(
  filteredPathways1,
  save = FALSE,
  StatTest = "T"
)

anovas <- papiHtest(
  filteredPathways1,
  save = FALSE,
  StatTest = "ANOVA"
)
```

```
papiLine(
  filteredPathways,
  relative = TRUE,
  setRef.cond = TRUE,
  Ref.cond = "HYDRATED",
  save = FALSE,
  legend.position="bottomleft"
)
```

Bibliography

Aggio, R.B.M., Ruggiero, K. & Villas-Boas, S.G., 2010. Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. *Bioinformatics*, 26(23), pp.2969–2976.

Allwood, J.W. & Goodacre, R., 2010. An introduction to liquid chromatographyâ "mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis*, 21(1), pp.33–47.

Allwood, J.W. et al., 2009. Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics*, 5(4), pp.479–496.

Alpert, P., 2006. Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *Journal of Experimental Biology*, 209(9), pp.1575–1584.

Anderson, J.W. & Beardall, J., 1991. *Molecular Activities of Plant Cells*, Blackwell Publishing.

Ankeny, R.A., 2001. Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project. *Philosophy of Science*, 68(3) pp.S251-S261.

Ankeny, R.A. & Leonelli, S., 2011. What's so special about model organisms? *Studies in History and Philosophy of Science*, 42, pp.313-323

Arabidopsis Interactome Mapping Consortium et al., 2011. Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science*, 333(6042), pp.601–607.

Bartels, D. & Hussain, S.S., 2011. Resurrection Plants: Physiology and Molecular Biology. In U. Luttge et al., eds. *Plant Desiccation Tolerance*. Springer.

Bartels, D., Luttge, U. & Beck, E., 2011. Synopsis: Drying Without Dying. In U. Luttge et al., eds. *Plant Desiccation Tolerance*. Springer.

Beckett, M., 2010. Investigation into the role of volatile organic compounds, and abscisic acid in stomatal regulation, in the resurrection plant Xerophyta humilis. *MSc thesis*, University of Cape Town

Beckett, M. et al., 2012. Photosynthetic limitations and volatile and non-volatile isoprenoids in the poikilochlorophyllous resurrection plant Xerophyta humilisduring dehydration and rehydration. *Plant, Cell and Environment*, 35(12), pp.2061–2074.

Berjak, P. & Pammenter, N.W., 2007. From Avicennia to Zizania: Seed Recalcitrance in Perspective. *Annals of Botany*, 101(2), pp.213–228.

Bicchi, C. & Maffei, M., 2012. The Plant Volatilome: Methods of Analysis. In *High-Throughput Phenotyping in Plants*. Methods in Molecular Biology. Totowa, NJ: Humana Press, pp. 289–310.

Bohnert, H.J., Nelson, D.E. & Jensen, R.G., 1995. Adaptations to Environmental Stresses. *The Plant Cell*, 7, pp.1099–1111.

Bray, E.A., 1993. Molecular Responses to Water Deficit. *Plant Physiology*, 103, pp.1035-1040.

Buitink, J. & Leprince, O., 2004. Glass formation in plant anhydrobiotes: survival in the dry state. *Cryobiology*, 48, pp.215-228.

Chaves, M.M., 2002. How Plants Cope with Water Stress in the Field? Photosynthesis and Growth. *Annals of Botany*, 89(7), pp.907–916.

Choi, Y.H. et al., 2011. Are Natural Deep Eutectic Solvents the Missing Link in Understanding Cellular Metabolism and Physiology? *Plant Physiology*, 156(4), pp.1701–1705.

Collett, H., 2003. Photosynthetic genes are differentially transcribed during the dehydration-rehydration cycle in the resurrection plant, Xerophyta humilis. *Journal of Experimental Botany*, 54(392), pp.2593–2595.

Collett, H. et al., 2004. Towards transcript profiling of desiccation tolerance in Xerophyta humilis: Construction of a normalized 11 k X. humilis cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration. *Physiologia Plantarum*, 122, pp.39-53.

Cushman, J.C. & Oliver, M.J., 2011. Understanding Vegetative Desiccation Tolerance Using Integrated Functional Genomics Approaches Within A Comparative Evolutionary Framework. In U. Luttge et al., eds. *Plant Desiccation Tolerance*. Springer.

Dace, H. et al., 1998. Use of metabolic inhibitors to elucidate mechanisms of recovery from desiccation stress in the resurrection plant Xerophyta humilis. *Plant Growth Regulation*, 24(3), pp.171–177.

Farrant, J.M., 2000. A comparison of mechanisms of desiccation tolerance among three angiosperm resurrection plant species - Springer. *Plant Ecology*, 151(1), pp.29–39.

Farrant, J.M. & Moore, J.P., 2011. Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Current opinion in plant biology*, 14(3), pp.340–345.

Farrant, J.M. et al., 2009. Desiccation tolerance in the vegetative tissues of the fern Mohria caffrorum is seasonally regulated. *The Plant Journal*, 57(1), pp.65–79.

Farrant, J.M., Brandt, W. & Lindsey, G., 2007. An overview of mechanisms of desiccation tolerance in selected angiosperm resurrection plants. *Plant Stress*, 1(1), pp.72-84.

Farrant, J.M., Cooper, K. & Nell, H., 2012. Desiccation Tolerance. In S. Shabala, ed. *Plant Stress Physiology*. CABI Publishing, Wallingford.

Fiehn, O., 2002. Metabolomics–the link between genotypes and phenotypes. *Plant Molecular Biology*, 48, pp.155-171.

Fiehn, O. et al., 2000. Identification of uncommon plant metabolites based on calculation of

elemental compositions using gas chromatography and quadrupole mass spectrometry. *Analytical chemistry*, 72(15), pp.3573–3580.

Fox, K. et al., 1998. Complementarity of GC-MS and LC-MS analyses for determination of carbohydrate profiles of vegetative cells and spores of bacilli. *Journal of Microbiological Methods*, 33(1), pp.1–11.

Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), pp.453–467.

Gaff, D.F. & Oliver, M., 2013. The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. *Functional Plant Biology*, 40, pp.315-328.

Guy, C. et al., 2008. Metabolomics of temperature stress. *Physiologia Plantarum*, 132, pp.220-235.

Halliwell, B., 1987. Oxidative damage, lipid peroxidation and antioxidant protection in chloroplasts. *Chemistry and Physics of Lipids*, 44(2-4), pp.327–340.

Han, X., 2003. Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *The Journal of Lipid Research*, 44(6), pp.1071–1079.

Hanson, J.R. ed., 2003. The classes of natural product and their isolation. In *Natural Products- The Secondary Metabolites*. Royal Society of Chemistry, pp.1–34.

Hermansson, M. et al., 2005. Automated Quantitative Analysis of Complex Lipidomes by Liquid Chromatography/Mass Spectrometry. *Analytical chemistry*, 77(7), pp.2166–2175.

Hoekstra, F.A., 2005. Differential Longevities in Desiccated Anhydrobiotic Plant Systems. *Integrative and Comparative Biology*, 45, pp.725-733.

Hoekstra, F.A., Golovina, E.A. & Buitink, J., 2001. Mechanisms of plant desiccation tolerance. *Trends in plant science*, 6(9), pp.431–438.

Hummel, J. et al., 2010. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics*, 6, pp.322-333.

Illing, N. et al., 2005. The signature of seeds in resurrection plants: a molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. *Integrative and Comparative Biology*, 45(5), p.771.

Ingle, R.A. et al., 2007. Proteomic analysis of leaf proteins during dehydration of the resurrection plant Xerophyta viscosa. *Plant, Cell and Environment*, 30(4), p.435.

Ingram, J. & Bartels, D., 1996. The molecular basis of dehydration tolerance in plants. *Annual review of plant biology*, 47, pp.377–403.

Iturriaga, G., Cushman, M. & Cushman, J.C., 2006. An EST catalogue from the resurrection plant Selaginella lepidophylla reveals abiotic stress-adaptive genes. *Plant Science*, 170, pp.1173-1184.

Jansen, J. et al., 2009. Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics*, 5(1), pp.150–161.

Jansen, J.J. et al., 2010. The photographer and the greenhouse: how to analyse plant metabolomics data. *Phytochemical Analysis*, 21(1), pp.48–60.

Jiang, G. et al., 2006. Proteome analysis of leaves from the resurrection plant Boea hygrometrica in response to dehydration and rehydration. *Planta*, 225(6), pp.1405–1420.

Kanani, H., Chrysanthopoulos, P. & Klapa, M., 2008. Standardizing GC-MS metabolomics. *Journal of Chromatography B*, 871(2), pp.191–201.

Kanehisa, M. & Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, 28(1), pp.27-30.

Kaplan, F. et al., 2004. Exploring the temperature-stress metabolome of Arabidopsis. *Plant Physiology*, 136, pp.4159-4168.

Kim, H.K. & Verpoorte, R., 2010. Sample preparation for plant metabolomics. *Phytochemical Analysis*, 21(1), pp.4–13.

Kim, H.K., Choi, Y.H. & Verpoorte, R., 2011. NMR-based plant metabolomics: where do we stand, where do we go? *Trends in biotechnology*, 29(6), pp.267–275.

Kopka, J., 2006. Current challenges and developments in GC–MS based metabolite profiling technology. *Journal of biotechnology*, 124, pp.312-322.

Kopka, J. et al., 2005. GMD@ CSB. DB: the Golm metabolome database. *Bioinformatics*, 21(8), pp.1635-1638.

Kranner, I. & Birtić, S., 2005. A Modulating Role for Antioxidants in Desiccation Tolerance. *Integrative and Comparative Biology*, 45, pp.737-740.

Kranner, I. et al., 2002. Revival of a resurrection plant correlates with its antioxidant status. *The Plant Journal*, 31(1), pp.13–24.

Kranner, I., Birtić, S., Anderson, K.M. and Pritchard, H.W. (2006) Glutathione half cell reduction potential: a universal stress marker and modulator of programmed cell death? *Free Radical Biology and Medicine* 40, 2155-2165

Lehner, A. et al., 2008. Changes in soluble carbohydrates, lipid peroxidation and antioxidant enzyme activities in the embryo during ageing in wheat grains. *Journal of Cereal Science*, 47, pp.555-565.

Leprince, O. & Buitink, J., 2010. Desiccation tolerance: From genomics to the field. *Plant Science*, 179, pp.554-564.

Majee, M. et al., 2012. Molecular Cloning, Bacterial Overexpression and Characterization of L-myo- inositol 1- Phosphate Synthase from a Monocotyledonous Resurrection Plant, Xerophyta viscosa Baker. *Journal of Plant Biochemistry and Biotechnology*, 14(2), pp.95–99.

Moore, J. et al., 2008. Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiologia Plantarum*, 134(2), pp.237–245.

Moore, J. et al., 2009. Towards a systems-based understanding of plant desiccation tolerance. *Trends in plant science*, 14(2), pp.110–117.

Moore, J. P. and Farrant, J. M. (2012) A Systems-Based Molecular Biology Analysis of Resurrection Plants for Crop and Forage Improvement in Arid Environments, in *Improving Crop Resistance to Abiotic Stress, Volume 1 & Volume 2* (eds N. Tuteja, S. S. Gill, A. F. Tiburcio and R. Tuteja), Wiley, Weinheim, Germany.

Moore, J.P. et al., 2006. Response of the Leaf Cell Wall to Desiccation in the Resurrection Plant Myrothamnus flabellifolius. *Plant Physiology*, 141, pp.651-662.

Moore, J.P. et al., 2005. The predominant polyphenol in the leaves of the resurrection plant Myrothamnus flabellifolius, 3,4,5 tri-O-galloylquinic acid, protects membranes against desiccation and free radical-induced oxidation. *The Biochemical journal*, 385(Pt 1), pp.301–308.

Moore, J.P., Nguema-Ona, E.E. & Vicré-Gibouin, M., 2013. Arabinose-rich polymers as an evolutionary strategy to plasticize resurrection plant cell walls against desiccation. *Planta*, 237, pp.739-754.

Mowla, S. et al., 2002. A novel stress-inducible antioxidant enzyme identified from the resurrection plant Xerophyta viscosa Baker. *Planta*, 215(5), pp.716–726.

Munns, R., 2002. Comparative physiology of salt and water stress. *Plant, Cell and Environment*, 25, pp.239–250.

Neale, A.D., Blomstedt, C.K. & Bronson, P., 2000. The isolation of genes from the resurrection grass Sporobolus stapfianus which are induced during severe drought stress. *Plant, Cell and Environment*, 23(3), pp.265-277.

Noble, E., 2006. Multilevel Modeling in Systems Biology: From Cells to Whole Organs. In Z. Szallasi, J. Stelling, & V. Periwal, eds. *Systems Modeling in Cellular Biology: From Concepts To Nuts And Bolts*. Cambridge, MA: MIT Press.

Oliver, M.J. et al., 2011. A Sister Group Contrast Using Untargeted Global Metabolomic Analysis Delineates the Biochemical Regulation Underlying Desiccation Tolerance in Sporobolus stapfianus. *The Plant Cell*, 23, pp.1231-1248.

Oliver, M.J. et al., 2004. The rehydration transcriptome of the desiccation-tolerant bryophyte Tortula ruralis: transcript classification and analysis. *BMC Genomics*, 5(89).

Oliver, M.J., Hudgeons, J. & Dowd, S.E., 2009. A combined subtractive suppression hybridization and expression profiling strategy to identify novel desiccation response transcripts from Tortula ruralis gametophytes. *Physiologia Plantarum*, 136(4), pp437-460.

Oliver, M.J., O'Mahony, P. & Wood, A.J., 1998. "To dryness and beyond" – Preparation for the dried state and rehydration in vegetative desiccation-tolerant plants - Springer. *Plant Growth Regulation*, 24(3), pp.193–201.

Oliver, M.J., Tuba, Z. & Mishler, B.D., 2000. The evolution of vegetative desiccation tolerance in land plants - Springer. *Plant Ecology*, 151, pp.85-100.

Oracz, K. et al., 2007. ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. *The Plant Journal*, 50, pp.452-465.

Pauli, G. & Jaki, B., 2005. Quantitative ^{1}H NMR: Development and Potential of a Method for Natural Products Analysis. *Journal of Natural Products*, 68, pp.133-149.

Peters, S. et al., 2007. Protection mechanisms in the resurrection plant Xerophyta viscosa (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. *Journal of Experimental Botany*, 58(8), pp.1947–1956.

Ramsay, J.O., Hooker, G. & Graves, S., 2012. *Functional Data Analysis with R and MATLAB*, Springer.

Roessner, U. et al., 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *The Plant Journal*, 23(1), pp.131–142.

Röhrig, H. et al., 2006. Desiccation of the resurrection plant Craterostigma plantagineum induces dynamic changes in protein phosphorylation. *Plant, Cell and Environment*, 29(8), pp.1606–1617.

Sanford, K. et al., 2002. Genomics to fluxomics and physiomics — pathway engineering. *Current Opinion in Microbiology*, 5, pp.318–322.

Schauer, N. & Fernie, A., 2006. Plant metabolomics: towards biological function and mechanism. *Trends in plant science*, 11(10), pp.508-516.

Schripsema, J., 2010. Application of NMR in plant metabolomics: techniques, problems and prospects. *Phytochemical Analysis*, 21(1), pp.14–21.

Sherwin, H.W. & Farrant, J.M., 1996. Differences in Rehydration of Three Desiccation-tolerant Angiosperm Species. *Annals of Botany*, 78(6), pp.703-710.

Sherwin, H.W. & Farrant, J.M., 1998. Protection mechanisms against excess light in the resurrection plants Craterostigma wilmsii and Xerophyta viscosa. *Plant Growth Regulation*, 24(3), pp.203–210.

Sherwin, H.W. et al., 1998. Xylem Hydraulic Characteristics, Water Relations and Wood Anatomy of the Resurrection Plant *Myrothamnus flabellifolius* Welw. *Annals of Botany*, 81(4), pp.567–575.

Smirnoff, N., 1998. Plant resistance to environmental stress. *Current Opinion in Biotechnology*, 9, pp.214-219.

Smirnoff, N., 1993. The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytologist*, 125, pp.27–58.

t'Kindt, R. et al., 2009. Joint GC-MS and LC-MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences*, 877(29), pp.3572–3580.

Tapp, H.S. & Kemsley, E.K., 2009. Notes on the practical utility of OPLS. *Trends in Analytical Chemistry*, 28(11), pp.1322-1327.

Tuba, Z. & Lichtenthaler, H.K., 2011. Ecophysiology of Homoiochlorophyllous and Poikilochlorophyllous Desiccation-Tolerant Plants and Vegetations. In U. Luttge et al., eds. *Plant Desiccation Tolerance*. Springer.

Tuba, Z. et al., 1993. Resynthesis of Thylakoids and Functional Chloroplasts in the Desiccated Leaves of the Poikilochlorophyllous Plant Xerophyta scabrida upon Rehydration. *Journal of Plant Physiology*, 142(6), pp.742-748.

Urano, K. et al., 2010. "Omics" analyses of regulatory networks in plant abiotic stress responses. *Current opinion in plant biology*, 13, pp.132-138.

van den Berg, R.A. et al., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(1), p.142.

Varmuza, K. & Filzmoser, P., 2009. *Introduction to Multivariate Statistical Analysis in Chemometrics*, Boca Raton, FL: CRC Press.

Vertucci, C.W. & Farrant, J.M., 1995. Acquisition and Loss of Desiccation Tolerance. In J. Kigel, ed. *Seed Development and Germination*. Springer.

Vertucci, C.W. & Leopold, A.C., 1987. The Relationship between Water Binding and Desiccation Tolerance in Tissues. *Plant Physiology*, 85, pp.232-238.

Vicré, M., Farrant, J.M. & Driouich, A., 2004. Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. *Plant, Cell and Environment* , 27, pp.1329–1340.

Walters, C. et al., 2002. Desiccation Stress and Damage. In M. Black & H. W. Pritchard, eds. *Desiccation and Survival in Plants: Drying Without Dying*. CABI Publishing, Wallingford.

Wang, W., Vinocur, B. & Altman, A., 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, 218(1), pp.1–14.

Wehrens, R., 2012. *Chemometrics with R*, Springer.

Whittaker, A. et al., 2001. Changes in leaf hexokinase activity and metabolite levels in response to drying in the desiccation-tolerant species Sporobolus stapfianus and Xerophyta viscosa. *Journal of Experimental Botany*, 52(358), pp.961-969.

Wiechert, W., Schweissgut, O. & Takanaga, H., 2007. Fluxomics: mass spectrometry versus quantitative imaging. *Current Opinion in Plant Biology*, 10, pp.323-330.

Williams, T.D. et al., 2011. Towards a System Level Understanding of Non-Model Organisms Sampled from the Environment: A Network Biology Approach. *PLoS Computational Biology*, 7(8), pp.1-20.

Yobi, A. et al., 2013. Metabolomic Profiling in Selaginella lepidophylla at Various Hydration States Provides New Insights into the Mechanistic Basis of Desiccation Tolerance.

Molecular Plant, 6(2), pp.369–385.

Zhang, W., Li, F. & Nie, L., 2010. Integrating multiple “omics” analysis for microbial biology: application and methodologies. *Microbiology*, 156(2), pp.287–301.